A World Without Reasons

From Moral to Metanormative Error Theory

Conrad Bakka

Thesis presented for the degree of
MASTER OF PHILOSOPHY

Supervised by
Associate Professor Sebastian Watzl

Department of Philosophy, Classics, History of Art and Ideas
Faculty of Humanities

UNIVERSITY OF OSLO
Autumn 2015
That [the concept of a normative reason] should have increasingly attracted the attention of philosophers is unsurprising; the only surprise is that it has taken so long. It is a concept fundamental to all thought. It is pervasive—actions, beliefs, and sentiments all fall within its range; primitive—all other normative concepts are reducible to it.

Skorupski (2012: 1)

It is an increasingly appreciated lesson of the recent decades’ metaethical literature that many of the concerns and arguments traditionally though to be about morality are really easily and naturally generalizable to metanormative concerns and arguments.

Enoch (2011: 2)

[T]he mysterious normativity of reasons is like the proverbial bump in the rug. You can push it around, but you can’t get rid of it.

Bedke (2010: 55)
Conrad Bakka

A World Without Reasons

From Moral to Metanormative Error Theory
Abstract

Moral error theory, the view that moral discourse is factually defective, is an influential view in contemporary metaethics. Although it has long been recognized that the arguments for this view might generalize and apply to other domains than the moral, it has traditionally been understood, defended, and evaluated as confined to the moral domain. This has begun to change.

In order to see whether the arguments for moral error theory generalize beyond the moral domain, I first formulate an argument for the view, which targets what error theorists usually attempt to capture – the authority of moral reasons. I argue that the resulting metaethical view is plausible, especially when compared to its main competitors. I then go on to show that most arguments for moral error theory, including the one I have presented, generalize to target all of irreducible normativity. Most moral error theorists should therefore become metanormative error theorists.

To explore the consequences of this generalization I show that at least two non-moral normative discourses are also committed to irreducible normativity, namely those concerning practical and epistemic reasons. I conclude by evaluating the plausibility of a metanormative error theory that holds not only moral but also practical and epistemic discourses to be factually defective.

© Conrad Bakka 2015

A World Without Reasons: From Moral to Metanormative Error Theory

https://www.duo.uio.no/
Acknowledgements

I was fortunate to begin working on this thesis when and where I did: It gave me the opportunity of having Sebastian Watzl as my supervisor. I will save the rhetorical flourishes I feel like heaping upon him, and simply say that he has been phenomenally supportive throughout the process and has provided an equal measure of insightful comments, objections and clearheaded suggestions.

The timing was also fortunate in that it allowed me to participate in a metaethics reading group along with Ainar Petersen Miyata, Martin Ravneberg, Marcus Bøhn, Øyvind Sætre Strøm, Sebastian Watzl and, sometimes, Caj Strandberg. I learned a tremendous amount during the group’s over a year and a half long run, and I owe a lot to all of you, not least for taking the time to read and discuss early drafts and providing stimulating discussions.

Another great debt is owed to all those who have provided insightful comments on drafts at different points in the writing process: Jonas Olson, Caj Strandberg, Knut Olav Skarsaune, Olav Gjelsvik, Hans Robin Solberg, Sivert Thomas Ellingsen, Maria Seim, Daniel Parmeggiani Gitlesen, Hallvard Sandven, Ludvig Fæhn Fuglestvedt, Bendik Hellem Aaby, Solveig Nygaard and Sara Vikesdal. I want to extend my deepest gratitude to those who have given me sustained comments on the whole thesis: Sebastian Watzl, Kim Phillips Pedersen, Mariona Eiren Sturm and Ainar Petersen Miyata.

The thesis you have before you is considerably better than it would have been without the valuable input from these commentators. I sincerely want to thank all of you for your time and efforts – and so should the reader!

I also want to register my appreciation for receiving a stipend from the Center for the Study of Mind in Nature (CSMN), and to IFIKK for subsidising travel costs involved in participating in the Edinburgh Masterclass in Practical Philosophy, and the Edinburgh Foundations of Normativity Workshop.

Lastly, I want to thank Loella Bakka for valuable practical assistance, and especially Ainar Petersen Miyata for assisting me with the unflattering practicalities of finishing a master's thesis.

C. B.

Oslo,

December, 2015
Table of contents

ABSTRACT ........................................................................................................................................ IV
ACKNOWLEDGEMENTS .................................................................................................................. V
TABLE OF CONTENTS ................................................................................................................... VI

1. INTRODUCTION: METHODOLOGICAL AND THEORETICAL PRELIMINARIES: .................... 1
   1.1 INTRODUCTION ................................................................................................................... 1
   1.2 METHODOLOGICAL PRELIMINARIES: THE STUDY OF MORAL DISCOURSE ..................... 4
   1.3 THEORETICAL PRELIMINARIES: JUDGEMENTS AND REASONS ........................................ 7
      1.3.1 Cognitivism vs. non-cognitivism .................................................................................... 7
      1.3.2 Motivational and normative reasons .............................................................................. 10
   1.4 ERROR THEORY: MOTIVATION, STRUCTURE AND FORMULATION .................................... 11
      1.4.1 Why would anyone want to be a moral error theorist? .............................................. 11
      1.4.2 The structure and formulation of error theories ......................................................... 13

2. MORAL ERROR THEORY I: THE COMMITMENTS OF MORAL DISCOURSE .................... 17
   2.1 FILLING IN THE CONCEPTUAL CLAIM .............................................................................. 17
      2.1.1 Locating core commitments of a discourse .................................................................. 17
      2.1.2 Are moral reasons categorical? .................................................................................... 19
      2.1.3 Foot’s challenge to moral categoricity and Joyce’s solution ..................................... 22
      2.1.4 Features of normative reasons: Prescriptivity, normative force and inescapability .... 26
      2.1.5 Reasons and normativity: The favouring relations framework ............................... 28
      2.1.6 Applying the framework: Reducible and irreducible favouring relations ............... 31
   2.2 CAN WE REJECT THE CONCEPTUAL CLAIM? .................................................................... 37

3. MORAL ERROR THEORY II: AN ARGUMENT FROM QUEERNESS ..................................... 38
   3.1 FILLING IN THE SUBSTANTIAL CLAIM ............................................................................ 38
      3.1.1 An argument from queerness ....................................................................................... 39
         3.1.1.1 Queerness claims .................................................................................................. 41
         3.1.1.2 Debunking explanations ....................................................................................... 49
         3.1.1.3 The plausibility claim .......................................................................................... 53

4. GENERALIZATION I: PRACTICAL REASONS ........................................................................ 57
   4.1 DO THE ARGUMENTS FOR MORAL ERROR THEORY GENERALIZE? .......................... 57
   4.2 THE REASONS OF PRACTICAL RATIONALITY .................................................................. 60
      4.2.1 Do hypothetical reasons imply categorical reasons? ............................................... 65
         4.2.1.1 Joyce’s answer ...................................................................................................... 65
   4.3 APPLYING THE FRAMEWORK OF FAVOURING RELATIONS TO HYPOTHETICAL REASONS .... 67
4.4 THE PROSPECTS FOR HYPOTHETICAL REASONS WITHOUT IRREDUCIBLE NORMATIVITY ........................................... 69
  4.4.1 The objection from normativity .......................................................................................................................... 71
  4.4.2 Competing reductive accounts and normative disagreement ............................................................................. 72
4.5 WHAT ABOUT OTHER ACCOUNTS OF PRACTICAL REASONS? .................................................................................... 75
4.6 SUMMARY ......................................................................................................................................................................... 77

5 GENERALIZATION II: EPISTEMIC REASONS ..................................................................................................................... 78
  5.1 THE REASONS OF EPISTEMIC RATIONALITY ........................................................................................................ 78
    5.1.1 Epistemic reasons and ordinary discourse ........................................................................................................ 78
    5.1.2 Epistemic justification and ordinary discourse ................................................................................................. 82
  5.2 THE STRUCTURE OF EPISTEMIC REASONS ................................................................................................................ 84
  5.3 THE POSSIBILITY OF REDUCTION .......................................................................................................................... 89
    5.3.1 Evidential support relations ................................................................................................................................ 90
    5.3.2 Non-normative accounts of support relations ...................................................................................................... 91
    5.3.3 Epistemic reasons as hypothetical reasons .......................................................................................................... 92
    5.3.4 Epistemic reasons as institutional reasons ............................................................................................................ 93
  5.4 SUMMARY ......................................................................................................................................................................... 94

6. CHALLENGES AND PROSPECTS FOR METANORMATIVE ERROR THEORY .............................................................. 94
  6.1 FROM MORAL TO METANORMATIVE ERROR THEORY .............................................................................................. 94
    6.1.1 Practical error theory: consequences and objections ............................................................................................. 95
      6.1.1.1 The self-undermining objection and the inconsistent practitioner objection ..................................................... 96
      6.1.1.2 Debunking explanations and non-moral reasons ............................................................................................... 98
    6.1.2 Epistemic error theory: consequences and objections .......................................................................................... 99
      6.1.2.1 Cuneo’s three undesirable consequences of epistemic error theory ................................................................. 100
      6.1.2.2 Epistemic error theory and the normativity of belief .......................................................................................... 104
    6.1.3 Metanormative error theory: consequences and objections .................................................................................. 106
  6.2 COUNTING THE COSTS, WEIGHING THE BENEFITS ...................................................................................................... 109
    6.2.1 Metanormative error theory and metanormative non-naturalist realism ................................................................. 109
    6.2.2 Other views ............................................................................................................................................................... 111
    6.2.3 Conclusion ................................................................................................................................................................. 111

APPENDIX: MORAL NATURALISM AND THE CONCEPTUAL CLAIM ......................................................................................... 112
  OPTION 1: NATURALIZING STRONG NORMATIVE FORCE AND INESCAPABILITY .......................................................... 113
  OPTION 2: GIVING UP ON STRONG NORMATIVE FORCE AND INESCAPABILITY ............................................................ 115

REFERENCES ........................................................................................................................................................................... 119
1 Introduction: Methodological and theoretical preliminaries

As contemporary metaethics moves ahead and positions gain in sophistication and complexity, victories, or even clear advantages, may become harder to achieve or consolidate. That is a kind of progress, but only if a clearer articulation of the surviving issues emerges as a result.

Darwall et al. (1992: 188)

1.1 Introduction

One goal of any full-scale metaethical theory is to explain the nature of moral discourse – that is, to explain what we do when we think, talk, argue and disagree about matters of morality. Here is one thing we usually take ourselves to be doing when engaging in moral discourse: discuss a moral reality where some actions are morally right and some are wrong, and where we might be either mistaken or correct in our judgements as to which is which. Many, but not all, metaethical theories accept these claims about ordinary moral discourse. One theory that tries to account for this seemingly obvious aspect of our moral practice is moral error theory.

Moral error theory nonetheless differs from the many other metaethical theories that accept these claims about our moral thought and talk in that it only accepts that we take ourselves – that we attempt – to discuss such a moral reality. According to moral error theory, our attempts are unsuccessful since there is, in fact, no such moral reality. We are, therefore, deeply in error; ordinary moral discourse is systematically and overwhelmingly flawed.

Despite its radical conclusion, moral error theory has experienced a surge of interest in the past few decades, and, while still being a minority view, has gained a number of defenders who have tried to argue for its plausibility (e.g. Garner 1994; Joyce 2001, 2006; Streumer forthcoming; Marks 2014; Olson 2011, 2014). Arguments for the view can usually be understood as consisting of two steps. First, the moral error theorist locates a problematic feature, for instance a belief, which is central to and deeply ingrained in our thoughts and talk about morality. The problematic feature that is supposed to support a moral error theory cannot be peripheral, but must be so central that without it moral discourse would no longer be what it is – it would become something else.

The next step is to argue that the feature in question suffers what we might call a factual defect, such as being false, non-existent, uninstantiated, incoherent or equally wrongheaded in some other way.1 The upshot of this two-part argument is that our ordinary moral discourse is centrally committed to something false, untrue, non-existent, incoherent, or the like, and that the discourse is therefore systematically and overwhelmingly factually defective, for instance by failing to say anything true.2

1 Field (1994) uses “factually defective discourse” in a related, though different sense. Note that holding a feature (e.g. a belief, concept etc.) to be factually defective is not a normative claim as I use it – it simply means that the feature in question is one of the following: false, incoherent, uninstantiated etc.

2 I will explain the claims made by an error theory in more detail towards the end of this chapter (section 1.4.2).
To take a simple example, it could be argued that much, if not all, of our thought and talk about morality is somehow committed to the belief that we are free agents who are able to choose our actions (cf. Haji 2002). The claim is therefore that this belief is essential to moral discourse, and that it would not continue to be moral discourse if it was removed. Suppose that this is correct, and that it turned out to be the case that we are not, in the required sense, free agents, say, because determinism is true. In this scenario, it seems that the truth of our moral judgements would require us to be free agents, and, since we are not, our moral discourse would be factually defective – it would contain no, or at least very few, truths.

In the event that moral error theory turns out to be true, or at least the best metaethical theory we have, and we come to accept that moral discourse is riddled with error, the following practical question is likely to arise: “What do we do now?” Do we eliminate moral discourse (Garner 1994), do we continue it, although in a non-assertoric fashion (Joyce 2001) or do we do something else? This is an interesting question, but one that I do not intend to say much about in this thesis. Nonetheless, it is important to recognize that the question of whether moral error theory is true is logically independent of the question of what one chooses to do if it is true (cf. Joyce 2015: §4).

There is also another, theoretical, consequence of moral error theory that has long been recognized: The arguments used to support moral error theory might also apply to certain non-moral domains. For instance, in his early and seminal formulation, J. L. Mackie claims that his arguments for an error theory targeting moral values can easily be generalized to also target “non-moral values, notably aesthetic ones, beauty and various kinds of artistic merit” (1977: 15). Mackie does not himself directly apply his arguments to any of these other domains, but leaves it as an exercise for the reader. Despite such early exploratory comments on the theoretical ramifications of moral error theory, it has been common to understand, defend, and evaluate moral error theory as if it could be confined to the moral domain (cf. Cuneo 2007: 1–5). Richard Joyce has this general tendency in mind when he states that “typically the moral error theorist thinks that there is something especially problematic about morality, and does not harbour the same doubts about normativity in general” (2015: §4).

Recently there has been an increasing awareness of the deep similarities between the moral and other normative domains, much in the way Mackie envisioned. For instance, it seems like important normative domains, such as practical rationality, which concerns what we should (or have reasons to) do, and epistemic rationality, which concerns what we should (or have reason to) believe, share important features with morality (Cuneo 2007; Bedke 2010; Streumer 2013b; Olson 2014; Enoch 2013: 2).³

³ Of course, there is a relatively popular position, compatibilism, that claims that freedom (of the will) is compatible with determinism. This is only meant to be a toy example, so I will not get into these issues here.


⁵ In references to The Stanford Encyclopedia of Philosophy, I will mark section numbers with the section sign, “§”.

⁶ I will discuss both of these domains in detail in Chapters 4 and 5, respectively.
Terence Cuneo (2007: 3) is perhaps overly bold in claiming that “on the face of it, epistemic values and reasons don’t appear to be very different types of creatures from moral values and reasons,” since these similarities have not been spotted by theorists who have taken more than a cursory glance at these issues. No matter how transparent or hidden these similarities are, a central aim of this thesis is to show that Cuneo is exactly right in that there are profound similarities between different normative domains, and, furthermore, that the theoretical importance of these similarities have been badly underappreciated by all parties in the metaethical and the metanormative debate.\footnote{A \textit{metanormative} view is a view concerning the nature of normativity (e.g. the nature of normative properties), in the same sense in which \textit{metaethical} theories concern the nature of morality (e.g. the nature of moral properties).}

As a result of the trend to notice substantial similarities between different normative domains there has been a growing interest in exploring whether the features that are supposedly problematic in the moral domain also have counterparts in other domains. If they do, this would mean that the arguments used in support of moral error theory could very well generalize to these other, non-moral domains. Detractors of moral error theory have often focused their attention on just such an argumentative strategy (Cuneo 2007; Stratton-Lake 2002: Introduction; cf. Lillehammer 2007). Such arguments attempt to show that the features that are supposedly both highly problematic and central to moral discourse are, in fact, also found within non-moral normative domains, where they are taken to be much less worrying.

Such detractors further point out that since the features are not problematic enough to support error-theoretic arguments within non-moral domains, it should remove our confidence in their success in the moral domain as well. This is because the error-theoretic arguments hold that the features are problematic in themselves, and they should therefore be no less problematic simply in virtue of appearing within a different domain than the moral – either it is a problem in all domains where it occurs, or in none.

Questions about to what extent the arguments supporting moral error theory generalize, and with what degree of success are therefore central in evaluating the plausibility of moral error theory itself. This thesis is an attempt to provide the beginning of an answer to the following question: Does moral error theory have theoretical implications beyond the moral domain, and, if it does, how do those implications influence the theory’s plausibility?

However, in order to say anything about whether the arguments for moral error theory can or cannot be contained to the moral domain, it will be necessary to say something about what the arguments for moral error theory are. I will therefore present, and tentatively defend, an argument for moral error theory in Chapter 2 and 3, and then go on to explore the theoretical ramifications of this and similar arguments in Chapters 4, 5 and 6. But first, in the rest of this chapter, I will provide the methodological and theoretical backdrop for later discussions.
1.2 Methodological preliminaries: The study of moral discourse

One aspect of human morality might usefully be labelled *moral discourse*. This aspect expresses itself in a myriad of ways, but it centres on the act of judging – more specifically, on the making and communication of *moral judgements*, which are judgements that represent acts, motives, persons or the like as having some moral property or other.\(^8\)

Why think *moral* discourse is especially significant, in contrast with, say, our discourse about whether books are good or bad? That there is a theoretically valuable distinction between our judgements concerning moral issues and other normative judgements can be argued for in a number of different ways. One popular way of doing so is by pointing to research within developmental psychology that argues for a clear distinction between conventional normative judgements, such as those concerning etiquette, and moral judgements (e.g. Turiel 1977, 1983; cf. Kumar 2015). This research seems to show that the judgements of children down to three years old show a reliable and substantial difference in their formal features depending on whether they concern moral or conventional issues. While conventional rules are usually conceptualized by children as, for instance, dependent on the decrees of some authority and not very serious,

\[\text{[p]sychological research on the moral/conventional distinction suggests that in moral judgments right and wrong, good and bad, praiseworthiness and blameworthiness, etc. are conceptualized as (1) serious, (2) general, (3) authority-independent, and (4) objective. (Kumar 2015: 2887)}\]

Appealing to the claim that we conceptualize moral judgements differently from other normative judgements, and that they possess different formal features, is one way of defending the claim that moral judgements have a distinctive nature.\(^9\) Another way to argue for this is by pointing to the arguments on behalf of a universal moral grammar, which claims that humans possess a specialized faculty for processing moral judgements, analogous to the postulated specialized faculty for processing language (e.g. Mikhail 2007; Hauser, Young, and Cushman 2008). This would clearly carve out a separate and significant role for moral judgements vis-à-vis other normative judgements. I will not defend any of these ways of carving up the space of judgements, but I will continue on the assumption that it is meaningful to discuss moral judgements, and by extension, moral discourse, as a separate case from other normative domains.

What I have said so far is compatible with the moral domain sharing many features with other non-moral normative domains – after all, we are assuming that the similarities hold across different domains. As we will see later, morality is *more* similar to some non-moral normative domains than to others, and in the next chapter, I will also provide a framework for comparing different normative domains.

Moral judgements can take the form of private thoughts or they can be uttered, for instance by speech or in writing. Someone who watches the news alone and thinks, but does not utter, “Robbery is bad”,

---

\(^8\) The nature of moral judgements is highly contested and I will say more about the view of them which will inform my discussion in section 1.3.1.

\(^9\) The there is substantial controversy as to the strength and validity of the moral/conventional distinction (e.g. Kelly et al. 2007).
has made a moral judgement, just like someone who sincerely utters those words in a conversation does. There is, however, room for terminological stipulation here. Some authors restrict their investigation of moral judgements to the proper interpretation of moral sentences or a similarly narrow part of moral language (cf. Hills 2015: 1 n. 1). That works well for practical or semantic purposes, but, strictly speaking, it is surely persons who judge something as right or wrong, not sentences. Of course, a person can employ utterances of moral sentences to make moral judgements. With this in mind, I will, for reasons of simplicity, not always distinguish between moral judgements as mental acts and as utterances. Simply put, I will take moral judgements to consist in the judging of some action, motive or the like as right or wrong, or good or bad, either in thought or in language. The making, voicing and discussion of moral judgements, as well as the collective reflection on moral issues, mark out what I will call moral discourse. The study of such moral discourse takes many forms. In order to make my methodological commitments clear, I will categorize them into two distinct approaches, one normative and one descriptive.

Given the prevalence and importance of moral issues in human society, one obvious question that seems to immediately spring from our moral practice is “Which moral judgements are correct?” In an attempt to answer this question, there are ethicists who seek to develop normative ethical theories that attempt to determine the criteria for morally right or good action. These criteria might then in turn be applied to more specific issues, such as whether abortion is morally permissible. The development of such criteria and their application is the province of normative and applied ethics respectively, and these fields therefore concern themselves with, in essence, normative questions about what makes something morally right or wrong. This enterprise is therefore aptly named the normative approach to moral discourse.

Another way to approach the study of moral discourse steps back from trying to arbitrate how we should decide which judgements are right and which are wrong, and instead tries to descriptively detail the various aspects of human moral practice, and, in addition, try to understand the nature of moral reality itself. This approach to moral discourse can therefore be helpfully termed descriptive. The descriptive approach in turn has two distinct methodological branches: an empirical and a conceptual (or philosophical).

Now, one might think that since making a moral judgement is a psychological (or linguistic) act it would naturally fall under the purview of psychologists (or linguists) to study them, and, consequently, that many, if not all, issues about moral judgements fall squarely within the province of an empirical approach. As a historic reality, however, philosophers have discussed the nature of moral judgements since long before

---

10 Depending on the theory of moral judgement in question, the two forms of moral judgements are likely to have certain differentiating properties. For instance, moral judgements understood as utterances could conceivably have pragmatic effects that moral judgements taken as mental states does not.

11 This is a grossly simplified view of the relation between normative and applied ethics. For discussion, see Beauchamp (2005).

12 As we will see in the coming chapters, moral error theorists thinks that because of the content of moral concepts and meaning of moral terms, we can be reasonably sure that there is no such moral reality. Metaethicists of certain other stripes, such as moral realists, do take themselves to be investigating such a reality.
psychology existed as a separate discipline (e.g. Nadelhoffer, Nahmias and Nichols 2010). Questions concerning moral judgements, and the attempts at answers to these, therefore have a long history within the discipline of philosophy. These questions have traditionally been the province of metaethics, which studies foundational, descriptive questions about various aspects of morality and our thought and talk about it.\(^\text{13}\) Metaethicists often point out that not all the relevant questions are empirical, but that many are conceptual in nature, and are best solved by methods typical of philosophical analysis.\(^\text{14}\) Unlike the problems tackled by the empirical approach, the questions handled by metaethics therefore often straddle the boundary between conceptual and empirical, and they quite often have deep connections to traditional philosophical issues within metaphysics or epistemology, as well as philosophy of language. As such, metaethics embraces both empirical and conceptual questions and methodologies.

A by-product of the fact that philosophers have traditionally handled questions about moral discourse and judgements is that metaethicists have had a habit of making armchair assumptions about what ordinary moral discourse is actually like, and building their theories on those assumptions (Darwall et al. 1992: 188–189). Metaethicists, until rather recently, have therefore employed two main methodologies – they have either tackled conceptual and (non-empirical) philosophical questions, or they have made armchair assumptions about empirical questions.

It could be argued that such an armchair approach to empirical questions is an acceptable way of proceeding when there are no relevant empirical data and one is unable to gather them oneself. Nonetheless, there has been a distinct lack of an actual empirical approach to metaethical questions to complement the conceptual approach within metaethics.

Admittedly, there have been strains of empirical psychology, especially within developmental psychology, which have attempted to investigate issues with a strong relevance to metaethical questions.\(^\text{15}\) Furthermore, all stripes of anthropologists, psychologists and neuroscientists, among others, have contributed to charting different facets of our moral thought and talk, for instance by documenting the moral convictions of a particular group, or the change in moral beliefs over the course of the lifespan, or by investigating connections between which ethical theories one subscribes to and particular configurations of brain wiring.\(^\text{16}\)

---

\(^{13}\) Note that typical metaethical investigations of questions such as “Do moral properties exist?” are descriptive even though they concern normative entities. For a more thorough description of metaethics, see Miller (2013: 1–3). Note also that the study of such descriptive questions can have normative implications. After all, if we, come to believe in, say, moral error theory because of a descriptive investigation along the lines of that which I will carry out in Chapter 2 and 3, that would mean that we would take all, or almost all, moral judgements to be untrue. Surprisingly, Mackie (1977: 16) disagrees, and holds that second-order moral inquiries (i.e. metaethics) are consequentially isolated from first-order moral inquiries (i.e. normative ethics).

\(^{14}\) For one statement of such methods, see Finlay (2014: Introduction).

\(^{15}\) For instance, the work of Lawrence Kohlberg and Elliot Turiel.

\(^{16}\) The project of mapping and documenting the moral systems of various cultures has been central to many disciplines such as social anthropology. For some discussion of the changes in moral views through the lifespan, see
However, these empirical strains have often been carried out somewhat in isolation from the traditional metaethical debates, and have therefore often answered different questions. It is only in the past few decades that there has been a drastic increase in the popularity of what has been called “moral psychology,” which has been carried out both by philosophers and psychologists, sometimes more or less in isolation from each other and sometimes through interdisciplinary collaboration.\(^17\) There has also been a growing trend of philosophers doing experimental work themselves (see Alfano and Loeb 2014 for an overview). In reviewing the changes the field has experienced, Joshua Greene goes as far as saying that “moral psychology has undergone a renaissance … the scientific study of morality has become a broad, interdisciplinary enterprise, drawing on insights and methods from philosophy, neuroscience, economics, anthropology, biology, and all quarters of psychology” (Greene 2011: 221).

Out of this collaborative effort there has appeared relevant empirical work on the nature of moral judgements – how they arise and what they are.\(^18\) In attempting to answer metaethical questions concerning the nature of moral judgements and moral discourse more generally, it is therefore necessary to at the very least consider the relevant empirical findings as a constraint on philosophical theorizing.

In the coming pages, I will make a number of assumptions – many of which are argued for, and some that are not – about moral judgements and about various discourses. Many of these assumptions bottom out in what is, ultimately, empirical claims. Further empirical study of moral judgements and discourse may therefore help support – or disprove – some of my arguments and unargued claims.

In what remains of this chapter, I will introduce some theoretical background necessary for later discussions; first, a more precise framework for understanding moral judgements, second, an important distinction between different forms of reasons, and, lastly, I will motivate and formulate moral error theory.

### 1.3 Theoretical preliminaries: Judgements and reasons

#### 1.3.1 Cognitivism vs. non-cognitivism

Consider the fact that in some countries and regions it is still customary to use stoning as a form of punishment for various offences (Batha 2013). Imagine that someone, upon reading a news article about the use of stoning as form of punishment, makes the following moral judgement: “Stoning is wrong”. Such judgements appear to have certain distinct features: They appear to say something about the world, and what they say seems best captured by the indicative mood and, by extension, by declarative sentences. It also seems that the judgements are either true or false, or at least that they are capable of being so – they are truth-

---

\(^{17}\) For an overview of the changes that moral psychology have undergone in the past few decades, see Greene (2015).

\(^{18}\) See for instance Jonathan Haidt’s work on moral judgements (e.g. 2001; with Joseph 2008), and the flurry of responses it has generated.
apt. They seem to most naturally be captured by a subject-predicate form, and the moral predicates can be transformed into abstract singular terms (e.g. “x is wrong” → “wrongness”), which is most naturally taken to suggest that they ascribe a property (i.e. the property wrongness) to the subject of the sentence (i.e. the act of stoning) (Joyce 2001: 13). Furthermore, moral language operates very much like ordinary descriptive language in that moral statements can be embedded in various contexts (Geach 1960, 1965).

If you came upon loganberries for the first time, you might make the judgement “Loganberries are dark red”. This type of descriptive judgement seems to share all the features of moral judgements mentioned above. Descriptive judgements are often taken to be a particular type of mental state, namely a belief (Hills 2015: 1; cf. Kumar 2015). Likewise, the utterance of such a judgement is taken to express a belief. The sense in which an utterance of a moral judgement expresses something is determined by what expectations the utterance produces, which is again determined by linguistic conventions. A liar’s utterance can therefore express a belief she does not have, and by introducing a statement with “Once upon a time …” one can cancel the expression, since the expectation of a corresponding belief is removed (Joyce 2001: 12).

When you utter the judgement, “Loganberries are dark red” to a companion, you thereby express your belief that loganberries are dark red. Since moral judgements appear very similar to descriptive judgements in the relevant respects, it seems natural to assume that moral judgements too express (or are) beliefs. Overall, then, moral judgements seem very much like ordinary descriptive judgements in these respects (cf. Hills 2015; Joyce 2001: 12–16).

There are cases where moral judgements do not seem similar to beliefs or descriptive judgements. One example is the how moral judgements seem strongly linked to motivation in ways belief is not usually thought to be (Smith 1994). I will not discuss such cases, but only note that they would be problematic only if they turned out to be incompatible with the account of moral judgements presented below.

The view that moral discourse is normally in the business of aiming at truth, in the sense of trying to veridically represent the world, and that ordinary moral judgements express (or are) beliefs in the aforementioned way, is known as cognitivism. Cognitivism might be more precisely defined as having two dimensions, one which concerns moral language and one which concerns moral thought. With respect to moral thought, cognitivism is committed to the thesis of psychological cognitivism, which says that moral judgements are beliefs. When it comes to moral language, cognitivism is committed to semantical factualism, which states that moral judgements have robust truth conditions, and that such judgements are therefore

---

19 The mental state expressed by moral judgements (or the mental state that moral judgements are) might not be belief per se, but could be some similar cognitive state. See Hills (2015) for an argument to this effect. For simplicity, I will only speak of beliefs being expressed, but one might just as well read “expresses a belief or some similar cognitive mental state” instead.

20 For one attempt at reconciling cognitivism with such features, see Strandberg (2012).

21 This particular way of spelling out the claims involved in cognitivism is based on van Roojen (2013), though my presentation differs in important ways.
evaluable in terms of robust notions of truth and falsity.\textsuperscript{22} A natural way of spelling this out, as we saw above, is by assuming that moral discourse is semantically similar to descriptive language in that moral predicates usually pick out properties, and that predicative moral sentences ascribe properties to the subject of the sentence (cf. van Roojen 2013: §1.1).

To illustrate these two claims, imagine again that someone, upon reading a news article about the use of stoning as form of punishment, sincerely utters, and thereby judges, “Stoning is wrong”. According to psychological cognitivism, moral judgements such as this express (or are) beliefs, in this case the belief that stoning is wrong. More generally, then, to judge that \( p \) expresses (or is) the belief that \( p \). According to semantic factualism, moral judgements are usually capable of being true or false, and in most moral uses the predicate (i.e. “is wrong”) picks out a moral property (i.e. \textit{being wrong}). Furthermore, the judgement ascribes this property to the subject of the sentence (i.e. the act of stoning).

Given this cognitivist account of moral judgements, what would it take for such judgements to not only \textit{aim} at truth, but to actually secure it? A plausible suggestion is that a statement such as “Stoning is wrong” has truth conditions such that the statement is true if and only if the act of stoning really does possess some property that makes it morally wrong.

This account, then, claims that moral discourse is ordinarily truth-apt; expresses (or consists in) beliefs; and purports to represent actions, motives, persons etc. as having certain moral properties.

That someone who under everyday circumstances sincerely utters “Stoning is wrong” asserts, and thereby expresses their belief in, the wrongness of stoning, and that their assertion can be evaluated in terms of truth and falsity might seem, at the very least, eminently plausible. There are, however, rival views of what we do when we engage in moral discourse. I will not dwell on these alternatives, but only give a brief overview of such \textit{non-cognitivist} accounts of moral discourse and judgements.

Non-cognitivist theories of moral judgement usually deny both psychological cognitivism and semantic factualism. A simple form of non-cognitivism, \textit{emotivism}, holds that moral judgements such as “Torture is wrong”, although similar in grammatical form and other surface features to descriptive judgements, actually \textit{do not} express (or consist in) beliefs or state facts (e.g. Ayer 1946). Rather, according to emotivism, moral judgements express certain non-cognitive states, such as (dis)liking or some other form of emotion, and make no claim that could be considered true or false. On this account of moral discourse, it does not aim at truth or attempt to represent some kind of moral reality, but is rather an expression of a state not evaluable in terms of truth or falsity.

Since the early emotivist theories, non-cognitivist views have tried to accommodate more and more of the features we usually take moral discourse to have, such as the possibility of moral disagreement, the possibility of embedding moral terms, and even a notion of moral truth, although only in a minimalist

\textsuperscript{22} The reasons for the qualifications “robust” and “substantial” stems from the rise of a breed of minimalist theories of truth. For more on this, see note 23.
sense.\(^{23}\) There have also been attempts to incorporate certain elements of cognitivism into forms of expressivism, resulting in so-called hybrid-expressivist views (e.g. Ridge 2006, 2007).

Whether we accept cognitivism or non-cognitivism about moral judgements will obviously be very significant for our understanding of moral discourse. As the debate between the two camps has been raging for decades, I will not argue for or against either view at length here. Rather, in order to discuss certain standard metaethical positions that assume the truth of cognitivism, I too will assume its truth.\(^{24}\)

### 1.3.2 Motivational and normative reasons

In the next chapter, I will introduce a framework that attempts to describe the structure of reasons, but it will be helpful to say a few things about reasons already at this point.

We often talk about reasons for why someone did what they did. Similarly, we often give a reason for why we did what we did. There are, however, distinctively different ways in which we talk of “the reason for someone’s action”. First, there are reasons in the sense of psychological states of an agent that (causally) explain their behaviour (cf. Smith 1994: 95–96), which are called *motivational reasons*.\(^{25}\) When explaining why someone did what they did, we often appeal to this kind of reasons. For instance, when children choose to receive one marshmallow instantaneously rather than accepting an offer of two marshmallows a short time later, we might say that their reason for doing so was that they *really* wanted the marshmallow then and there. In such cases, we are discussing the actual mental or psychological states that motivated the action. This type of reason talk can, on a very crude account, be understood as combination of (at least) a desire and a belief, which together motivates an agent to pursue a course of action in order to satisfy the desire in question (Davidson 1963).

On such an account, the motivating reason specifies what *in fact* motivated, say, the 39 members of the Heaven’s Gate cult to ingest food containing arsenic and cyanide. Their motivating reason, we can speculate, was that they *desired* to enter a new form of being on board a space ship travelling behind the Hale-Bopp comet, and *believed* that this could be achieved through collective suicide. Now, even though we use the word “reason” here, it is important to take note of the fact that we are only talking about what, in some sense, *explains* their action; it is *because* they had a certain belief–desire pair that they acted as they did.

---

\(^{23}\) Minimalism about truth roughly says that all there is to a statement or proposition *p* being true is that it satisfies this schema: ‘*p*’ is true if and only if *p*. If the non-cognitivist has some analysis of moral wrongness, which she usually does, she can claim that it is true that *p* is morally wrong, without claiming that this entails that the act in question has some moral property. This is in contrast to the view mentioned earlier, where the truth of such sentences or propositions requires the act to possess a moral property. For some of the leading proponents of variants of non-cognitivism, see Gibbard (e.g. 1990) and Blackburn (e.g. 1998).

\(^{24}\) Popularity is hard to measure, and many would claim, a poor guide to truth. A rough feel of the relative popularity of cognitivism and non-cognitivism among practicing metaethicists may nonetheless be found in the 2009 PhilPapers Survey (Bourget & Chalmers 2014). Of the respondents who reported metaethics as an area of specialization, 74.5% responded that they accept or lean toward cognitivism, while 13.7% answered that they accept or lean toward non-cognitivism.

\(^{25}\) There has been much debate over whether reasons are causally efficacious or not, with the former being the dominant position (Wilson and Shpall 2012). I do not intend to enter that debate here.
This is only stating descriptive, psychological facts, as this type of reasons is only invoked for causal and/or explanatory purposes. Using only the motivational sense of “reason”, it makes no sense to say that someone, such as the members of the Heaven’s Gate cult, “had reason to act differently”.

As should be apparent, this is not always the way we use the term “reason”. It is just as common to intend to say something about whether an action was justified, rational, good etc. or not. We might be inclined to say that the cult members really had no reason at all to commit collective suicide. In such cases, it is not the notion of a motivational reason that is in play, but that of a normative reason (Raz 2011: 16–20; Broome 2013: Chapter 4). This type of reasons says something about what an agent should or should not do, what an agent has or does not have reason to do, or whether an action was justified or not. Furthermore, in the case of normative reasons, it is perfectly possible to claim that, while the agent took herself to have good reasons to perform a certain action, there was, in fact, something else that she had more reason to do.26

In one sense, then, one could specify the (motivational) reasons the individuals had for participating in a collective suicide, while at the same time claiming that their (motivational) reasons were not particularly good, or even that there was really no (normative) reason at all to do what they did. It is normative reasons that will be our main focus, and unless I specify otherwise, I will from here on use “reason” in the sense of “normative reason”.

1.4 Error theory: Motivation, structure and formulation

1.4.1 Why would anyone want to be a moral error theorist?
One question that is sure to arise for many who encounter moral error theory is the following: Given its radical nature, how could one come to find such a view attractive? On the one hand, moral error theory is a sceptical theory, in that it questions our common beliefs about morality.27 Such sceptical theories have a tendency to be provocative and important, despite seldom having many adherents. It is revealing in this respect to see how even metaethicists often see moral error theory as a theoretical bogeyman that looms in the background and threatens to force itself on them, unless a plausible alternative is produced (e.g. Smith 1994). On this picture, even those who have no inclination towards moral error theory, or even personally dislike its consequences, might come to feel its pull, just as with other forms of scepticism. Nonetheless, it is usually seen as something that must be overcome.

To see why moral error theory often appears to loom in the background, consider that any student, or scholar for that matter, of metaethics is likely to have been struck by how, as three of field’s preeminent
practitioners put it, “remarkably diverse” the metaethical landscape is (Darwall et al. 1992: 124). Diversity might have its benefits, but a less cheerful way of describing the situation is forcefully expressed by Michael Smith:

[I]t seems to me best to begin by acknowledging, and attempting to diagnose, the difficulties involved in giving any convincing answers at all to meta-ethical questions. For if one thing becomes clear by reading what philosophers writing in meta-ethics today have to say, it is surely that enormous gulfs exist between them, gulfs so wide that we must wonder whether they are talking about a common subject matter. … Nor should it be thought that though there are disagreements, there are certain dominant views. The situation is quite otherwise. There are no dominant views. (Smith 1994: 3)

Smith goes on to list a veritable smorgasbord of wildly different and often mutually incompatible views about moral properties, facts, judgements, and about moral discourse itself (1994: 3–4). Smith’s diagnosis of the situation is that no current metaethical theory is capable of capturing all of the central platitudes about morality that ordinary moral discourse contains. For this reason – and herein lies the problem – different metaethical theories have to jettison some central platitude or other. That is, a metaethical theory has to give up on something that is ordinarily taken to be central to our understanding of morality. As a result,

no matter which proposition these philosophers choose to reject, they are bound to end up denying something that seems more certain than the theories they themselves go on to offer. Moral nihilism rightly looms. (Smith 1994: 13, citations removed)

The upshot of all this is what Smith has famously termed “the moral problem” – though, as has been pointed out, it would be more precise to call it “the metaethical problem” (Sayre-McCord 1997). The problem is this: We seem unable to come up with a theory that gives a unified picture of our discourse about morality, and which accounts for all of its central features. Despite this, there has been a proliferation of metaethical theories, and, as Smith points out, what characterizes them is that they discard one or more common platitude about morality, or that they require troublesome metaphysical postulations, or in some cases both. The general idea, then, is that if we are incapable of coming up with a metaethical theory that sufficiently manages to capture our ordinary understanding of morality, this can be taken to show that our ordinary conception of morality is deeply flawed in some way, and that we might very well be stuck with moral nihilism in the form of moral error theory.28

The idea in play here is that moral error theory might turn out to be the least bad metaethical theory. After all, moral error theory can capture more or less all of our ordinary platitudes about morality, since it has no problem with accepting an inconsistent set of such platitudes.29 In other words, an error theory could

28 Smith (1994) criticises a wide range of views in his book, although he himself, of course, goes on to present his own attempt at constructing a metaethical theory that respects the platitudes about morality while remaining an attractive theoretical position.
29 One platitude an error theory might be thought unable to capture is the platitude that there are moral truths. However, it can fully account for our belief that there are moral truths, which might be all that we should ask. In Chapter 3, I discuss how it can account for our beliefs about such truths, without in any way being committed to the actual existence those truths.
certainly help explain why the project of metaethics has had a very difficult time coming up with acceptable full-scale metaethical theories, let alone consensus on smaller questions. If we simply list the features we intuitively take morality and moral discourse to have, moral error theory might very well be the only theory that is capable of encompassing all of them. While it might therefore allow for a natural and adequate interpretation of ordinary moral discourse, it also comes with a serious cost, namely that our moral beliefs are systematically and perhaps entirely false.

This, then, is one, negative, way to motivate moral error theory; as a last resort when nothing else works. However, such an argumentative strategy has a significant drawback. First of all, compare the state of metaethics with that of modern physics: For almost a century we have been aware of the lack of compatibility between some of our best physical theories of different domains (relativity and quantum mechanics). As it stands, physics looks to be more or less internally inconsistent. Nonetheless, it would certainly seem premature to adopt an error theory about physics.\footnote{Thanks to Sebastian Watzl for this analogy.} Maybe it will turn out to be impossible to unify different levels of physical theory, maybe not. What is clear in any case is that this argumentative strategy on behalf of an error theory essentially requires waiting for inquiry to come to an end before we can conclude that nothing else works. As such, it is suited only for those of a very patient disposition.

Fortunately, there is a direct, and positive way to argue for moral error theory. This strategy consists in, first, providing a truthful and adequate description of actual moral discourse, and, second, claiming that of the theories that are able to account for this description, moral error theory is the most plausible of the lot. As we will see, there are multiple candidates in the running, but to give away the ending, my claim will be that moral error theory scores the most points in the end – although not by much. This is the argumentative strategy that I will pursue in the next two chapters.

1.4.2 The structure and formulation of error theories

It is now time to go into more detail about what it means to be an error theorist about a given discourse, and how one could go about arguing for such a view. As with varieties of philosophical realism, there are many different types of error theories, and their specific content depends on the particular domain to which they apply. Likewise, the specific motivations for adopting an error theory about a given discourse will vary from case to case. Different error theories do nonetheless share a common feature in that they hold the discourse in question to be systematically and overwhelmingly flawed, often by holding all (or almost all) statements within that discourse to be false (or as I will put it, untrue\footnote{The use of “untrue” rather than “false” is meant to allow for error theories that claim that moral judgements suffer presupposition failure, and therefore might be held to be neither true nor false (cf. Olson 2014: 8–9; Joyce 2001: 6–9).}). We can therefore give a general definition of an error theory about some discourse, $D$, as follows:
There has been influential error theories proposed about many different discourses, including colour (e.g. Boghossian and Velleman 1989), arithmetic (e.g. Field 1980), folk psychology (e.g. Churchland 1981) and, of course, ethics (e.g. Mackie 1977). However, according to Chris Daly and Stephen Liggins, “[m]any contemporary philosophers rate error theories poorly”, and such theories are “commonly dismissed out of hand” (2010: 210). One goal of the next two chapters is to show that moral error theory is a plausible theory of moral discourse and morality, especially when compared to its strongest competitors. I will therefore mostly be discussing moral error theory in the next two chapters, but later chapters will also explore error theories about other discourses, among them epistemic and practical error theory. Let us therefore begin by defining an error theory about moral discourse:

**Moral Error Theory**

All moral judgements, or some significant subset thereof, are untrue.

These are standard ways of formulating moral error theory (cf. Daly and Liggins 2010). However, such definitions give rise to a difficulty concerning how exactly one is to understand the exact formulation of moral error. The difficulty is brought out by the following line of reasoning: Moral error theory holds that most, if not all, moral judgements are untrue. For instance, it will likely hold that the following judgment is untrue: “Stealing is morally wrong”. But, remember, the law of excluded middle says that for any statement, either that statement or its negation must be true. The judgement “Stealing is not morally wrong” must therefore be true. But is not this latter judgement a moral judgement, and a true one at that? If it is, it would seem to be a counterexample to moral error theory, one that follows straightforwardly from the formulation of the theory itself.

Furthermore, that “Stealing is not morally wrong” would mean that stealing is morally permissible, since permissibility can be understood as equivalent with “not morally forbidden”. However, according to moral error theory, the judgement “Stealing is morally permissible” is most naturally taken to be untrue. We therefore seem to be able to derive multiple contradictions (e.g. that stealing is both morally permissible and not morally permissible) from the definition of moral error theory given above. Generating contradictions is, of course, a serious problem for any theory.

There are multiple solutions available to this problem, and nothing of importance hangs on exactly how one solves it, as long as one does solve satisfactorily. In order to avoid spending an inordinate amount

---

32 For a slightly different formulation, but one that I have modelled mine on, see Olson (2014: 8). In the service of elegance, I will sometimes drop the qualification “or some significant subset thereof” and simply claim that moral error theories hold all moral judgements to be untrue, though the qualification should still be kept in mind.

33 There are cases where some argue that this does not hold (e.g. Strawson 1957). Joyce (2001: 6–9) applies this line of thinking to his formulation of error theory. I do not follow him on this score.
of space on this problem, I will simply sketch one solution to show that the problem is solvable.\textsuperscript{34} This solution consists in defining moral error theories as concerned only with \textit{positive moral judgements}.

By “positive moral judgement” it is meant that the judgement must entail a claim about what 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). That is, the error theorist must \textit{list} the moral predicates that are taken to be positive moral predicates. Importantly, the error theorist \textit{does not} include “is morally permissible” in its list of such predicates, and is therefore silent with respect to permissibility judgements.

On this formulation, it would no longer be possible to generate a contradiction since the theory does not claim that the judgement “stealing is not morally wrong” is false, and therefore does not allow the inference that stealing is not morally permissible. This is because “not morally wrong” is not a \textit{positive} moral predicate. A consequence of this formulation is that moral error theory is silent about such \textit{negative facts}.

The drawbacks of this position are, first, that it says nothing about supposed (negative) moral facts such as “It is not the case that stealing is wrong”. However, the status of such negative facts is very much up in the air, and remaining neutral about them might not seem like a large cost (Sinnott-Armstrong 2006: 34). Next, this solution has the consequence that \textit{all} acts are morally permissible. While perhaps inelegant, if all that is meant by “permissible” is “not morally forbidden”, then this does not seem like a very serious problem either. After all, “Stealing is morally forbidden” \textit{is} false according to this formulation of moral error theory.\textsuperscript{35} Note that the solution under consideration can say that both “It is morally wrong to perform abortions” and “It is morally wrong not to perform abortions” are false, as the negation in the second sentence takes narrow scope. However, in a sentence like “It is not morally wrong to perform abortions”, the negation takes wide scope, and it is therefore not a \textit{positive} moral fact (Sinnott-Armstrong 2006: 35).

By restricting the scope of moral error theory to positive moral judgements, one removes the generation of contradictions. With these changes in place, we get the following refined definition of moral error theory:

\begin{center}
\textit{Moral Error Theory}\textsuperscript{*}: All positive moral judgements are untrue.
\end{center}

However, as the choice of solution to this problem is of little consequence, and for reasons of simplicity, I will use the previous, unrefined definition in this thesis. The attentive reader might therefore find that I at some point say something not compatible with the specific solution I outlined above, but again, nothing of importance will hang on such minor issues, so I will prefer simplicity to consistency on this point.

\textsuperscript{34}I consider Olson’s (2014: 11–15) solution to the problem preferable to the one I develop here, as he avoids the drawbacks of the position I outline here. But as this is of relatively little importance for my purposes in this thesis, I prefer to use the simpler workaround provided by Sinnott-Armstrong (2006: 34–36). For other suggested solutions, see Joyce (2001: 6–9) and Pigden (2007).

\textsuperscript{35}Again, for a solution that avoids these two minor drawbacks, see Olson (2014: 11–15).
Having now defined moral error theory, it is time to look at how one can argue for it. On this note, it is important to recognize that there are many ways in which one can be a moral error theorist, and, consequently, many roads to the conclusion that all moral judgements are untrue. I will not, however, dwell on what has been dubbed “non-standard error theories” (Olson: 8–11), or spend much time looking at other roads than the one I am myself pursuing.36

Arguments for error theories can generally be understood as having a bipartite structure, which ground their claim that all judgements in a given discourse, D, are untrue. The two components at work are, first, the identification of core commitments of the relevant discourse, and secondly, the claim that at least one of these core commitments is factually defective in some way. Based on this general argumentative strategy, we can, generally speaking, say that error theories can be understood as consisting of two claims, a conceptual claim and a substantial claim.37 Let us call this the master argument for error theory. It can be presented as follows:

**THE MASTER ARGUMENT FOR ERROR THEORY**

<table>
<thead>
<tr>
<th>Conceptual claim</th>
<th>Substantial claim</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discourse D has a core commitment to x</td>
<td>x is factually defective (e.g. untrue/nonexistent/uninstantiated/incoherent etc.)</td>
<td>Discourse D is systematically and overwhelmingly factually defective</td>
</tr>
</tbody>
</table>

The conceptual claim is meant to secure the error theorists’ aim of showing that the discourse in question is centrally committed to some (problematic) feature, where what “centrally committed” entails will be explained in the next chapter. The substantive claim then goes on to state that this central feature is in fact deeply flawed, either by being untrue, incoherent, or equally wrongheaded in some other way. Since it is a core commitment that is flawed in this way, it makes the whole discourse, or at least most of it, factually defective (for instance, by being entirely untrue).

Since I seek to investigate the extent and degree of success with which the arguments for moral error theory generalize to other domains, it is first necessary to determine what the arguments for moral error theory are. I will not, however, investigate all the ways in which the master argument can be filled in on behalf of moral error theory, but will restrict myself to what I take to be the strongest and most plausible way of doing so.

---

36 See Joyce (2015: §4), Olson (2014: 15–117) and Kirchin (2012: 83–84) for other ways to be a moral error theorist. 37 Many authors take what I have called “the substantial claim” to be concerned with only metaphysical (or ontological) issues, and therefore call it “the metaphysical (or ontological) claim” (e.g. Miller 2013). However, error theories can target other features than these more narrow labels imply, and I therefore prefer the term “substantial” (cf. Kirchin 2012: 82; Daly and Liggins 2010: 213–215). In Chapter 3, I will, in fact, fill in the substantial claim in a way that relies on metaphysical and epistemological issues, but it is important to be aware that there are other ways for error theorists to argue for their view.
There are various ways of supporting both the conceptual and the substantial claim. In the next chapter, I will provide what I take to be the most plausible way of spelling out the conceptual claim. Then, in Chapter 3, I will provide what I take to be the strongest way of supporting the corresponding substantial claim. Together this will give us a suitably detailed argument for moral error theory, which in turn will enable us to investigate whether the argument for moral error theory generalizes to other, non-moral domains, and if so, to what extent such generalizations are successful.

2 Moral error theory I: The commitments of moral discourse

Error theorists, as is widely noted, at least acknowledge the full strength of the commitments of normative discourse.

Enoch (2011: 81)

2.1 Filling in the conceptual claim

Having motivated and given a general outline of error theories and the argumentative strategy they employ in the previous chapter, it is now time to begin filling in the details of the argument for moral error theory. This will primarily consist in filling in the two claims of the master argument for error theory, namely the conceptual and the substantial claim. After we have formulated a viable argument in favour of moral error theory in this chapter and the next, we will proceed to investigate whether such arguments generalize outside to the moral domain.

In this chapter, I will fill-in and defend the version of the conceptual claim that I take to be most suited to construct a moral error theory. Let us look again at the general form of the conceptual claim:

\[\text{CONCEPTUAL CLAIM} \quad \text{Discourse } D \text{ has a core commitment to } x.\]

In order to construct a suitably strong and simultaneously plausible version of the conceptual claim for moral discourse, it is necessary to give an adequate and truthful account of that discourse. In filling in the conceptual claim, then, the error theorist is after the same thing as any other theorist of moral discourse – to account for it in the most theoretically adequate way possible. That is what we will attempt below.

2.1.1 Locating core commitments of a discourse

The conceptual claim requires that we locate a core commitment of moral discourse, and, in order to serve the purpose of moral error theory, the commitment should be deeply problematic. Note that this does not mean that the error theorist should distort her account of ordinary moral discourse in order to secure her claim, but only that if the core commitment is not problematic, an error theory would not get off the ground.
I will deal with what makes a commitment problematic in the next chapter, so for now we will begin with the following question: What does it mean to identify central commitments of a discourse? In order to answer this, it is helpful to begin by working through a relatively simple example. Elaborating on an analogy from Simon Kirchin (2012: 79), we can use an error theory about traditional Christian discourse to make the process more perspicuous.

It seems plausible to claim that a central commitment of traditional Christianity, and therefore of what we might term Christian discourse, is a belief in the existence of God. There are also many other, more peripheral, commitments that might be characteristic of Christian discourse, such as believing in the power of prayer or believing that Jesus turned water into wine. It nevertheless seems like some beliefs are more central to a discourse than others, and we might call the maximally central features of a discourse its core commitments, features without which the discourse would not be the one it currently is.

Imagine someone who claims to be professing the truth of Christianity, but additionally claims to have no belief in the existence of any deity. In such a case, it would seem that the person is not really championing the truth of Christianity, as traditionally understood, but of something else. On the other hand, it seems entirely possible to be sceptical of the more peripheral commitments of Christianity, such as the belief in prayer, while still championing something that is recognizable as traditional Christianity.

Joyce (2001: 3–5) has characterized such core commitments as being non-negotiable, in the sense that without them, a given discourse would no longer be the discourse it currently is. Joyce further attempts to explicate this notion of non-negotiability in terms of a “translation test”, which roughly goes like this (2001: 3): Suppose we encounter a foreign culture with a foreign language. This culture has a term, “Christianism”, which we discover has a very similar content to our term, “Christianity”, except in one respect: it fails to capture one (or more) important commitments of our term, such as, say, a belief in the existence of God. That is, it turns out that Christianists do not believe in God, but otherwise share all relevant features with traditional Christians in our culture. If this fact would stop us from translating the foreign term into our similar but importantly different term, then the commitment in question is non-negotiable. As Joyce is well aware, this does not give us an unproblematic decision procedure, involving as it does a rather complex counterfactual, but it does nonetheless go some way towards making sense of the distinction between core and peripheral commitments.

Having now seen an example of a core commitment of a discourse, it seems straightforward how such a commitment could turn out to be deeply flawed by being factually defective. If, for instance, it turns out either that God does not exist (or that the concept of God is incoherent), it would make the core

38 Is adopting, say, a non-cognitivist interpretation of religious discourse in order to avoid a commitment to belief in God’s existence is not ruled out (cf. Schweiker and Lovin 2005: 25)? In general, when certain central commitments of a discourse comes under enough pressure, theoreticians have a tendency to come up with creative ways of attempting to avoid such commitments. Such theoretical maneuvering often comes at a steep cost in terms of plausibility. Arguably, it was this kind of maneuvering that originally motivated moral non-cognitivism (Baldwin 1990: 66; Schroeder 2010: chapter 1 and 2).
commitment untrue. In either case, we would have the material needed to construct an error theory about traditional Christian discourse as per the master argument for error theory outlined in the previous chapter. \(^{39}\)

While identifying a core commitment that is also potentially deeply flawed seems relatively straightforward in the case of Christianity, it has turned out to be more challenging with respect to morality. One reason for this is that the commitments of moral discourse are often much more implicit. It is therefore important to note that it is not required that participants in the discourse themselves are able to explicitly voice such commitments, but only that the commitments are in some sense implicit in their actions and thoughts (cf. Joyce 2006: 192).

2.1.2 Are moral reasons categorical?
Now that we have a grip on how to go about locating core commitments of a discourse, it is time to introduce one candidate that has been the target of many formulations of moral error theory: categorical reasons, in one form or another (e.g. Mackie 1977; Garner 1990; Joyce 2001; Olson 2011). Before I go on to introduce this notion, we need to get clear on how reasons relate to what we ought to do. One commonsensical answer to this question is what Joyce has termed Mackie’s platitude. \(^{40}\)

\[ \text{Mackie’s platitude} \quad \text{It is necessary and a priori that, for any agent } A, \text{ if } A \text{ ought to } \Phi, \text{ then } A \text{ has a reason to } \Phi. \] \(^{41}\)

According to Mackie’s platitude, it is a fact about our ordinary concept “ought”, that if we ought to do something, then it follows that we have a reason to do so. Furthermore, it is therefore part of what we mean by “ought” (or “should”), that we have a reason to comply with the ought-statement (Mackie 1977: 64–66, 72–77; cf. Gibbard 2012: Chapter 1). This seems plausible enough, as the alternative would be the possibility of circumstances where you ought to do something, but were you have no reason to do it. This would certainly be rather odd, and Joyce points to John McDowell’s characterization of this view as an “insane thesis” (McDowell 1978: 14, cited in Joyce 2001: 39). Until we have reason to think otherwise, we should therefore accept that it is part of our concept of “ought” (or “should”), as applied to actions, that if we subscribe to a (atomic) statement containing it, we take there to be an accompanying reason to comply with

---

\(^{39}\) In the case of incoherence, it might be necessary to give a slightly different account of the error the discourse suffers from. The judgments of a discourse that has core commitments that involve incoherence might be meaningless and therefore not even truth-apt (cf. Kirchin 2012: 83–84; Cappelen 2013).

\(^{40}\) See Joyce Joyce (2001: 38) for the relation between Mackie’s platitude and predicative uses of “ought” as a possible counterexample. I will also discuss another possible counterexample, “logical oughts”, in Chapter 5. I will take the same to hold for most uses of “should”, as it would be equally strange to say, “You should not steal, but there’s no reason not to”.

\(^{41}\) In order to keep symbol and variable use consistent, I will sometimes modify quotes in line with my use, and not it as done here.
it. This licenses us to take a claim of the form “You ought to save drowning children” to entail something of the form “You have a reason to save drowning children”.

There are, however, metaethicists who hold that moral ought-claims do not necessarily entail reason-claims (e.g. Railton 1986; see also the Appendix). Given the plausibility of Mackie’s platitude, it seems unproblematic to hold that such theorists, who try to introduce reforming definitions of moral concepts, simply fail to capture the ordinary moral concept they set out to reform when they leave out this reason-implying feature. I discuss this issue further in the Appendix, and we will see a plausible way to defend Mackie’s platitude later in this chapter.

In addition to being reason-implying, ought-claims can be either qualified or unqualified (Broome 2013: 25–26). One might say, “You ought not to break into someone else’s cabin, but now you ought to”. Or, in a phrase from Joyce (2001: 35), “You ought to do what you ought not to do”. The qualified “oughts” in these phrases are domain-restricted, say to what you should morally do, while the unrestricted ought-claims should be understood as

overall or all-things-considered judgments about what one should do, and if there is something one should do, all things considered, it is because there are reasons for performing that action that ‘win out’ among all other competing reasons. (Bedke 2010: 43)

So we could say that the reason an ought-claim entails a reason-claim is because you ought to do that which you have reasons to do. Moreover, that which you have most reason to do is what you ought to do in an unqualified sense. Similarly, an ought-claim that contains such an unqualified ought, such as “You ought not to torture cats for fun”, entail a commitment to moral reasons that possess this all-things-considered quality. When we utter such imperatives, we are therefore committed to that the intended listener has such an all-things-considered reason to comply.

This short discussion of the role that reasons play in moral discourse shows just how central they are. Moral claims, by and large, entail claims about moral reasons. Furthermore, if moral reasons are held to be categorical, then most, if not all, moral claims can be shown to be (conceptually) committed to categorical reasons. Though we have now looked at how moral reasons permeate moral discourse, the question of what categorical reasons are, and whether moral reasons really are categorical, remains.

In order to see why error theorists claim that categorical reasons are a good way of conceptualising moral reasons, it is instructive to look closer at the relation between reasons for action and the ends of agents. Consider first non-moral reasons. A natural, descriptive fact such as that there lies a park outside my window does not, in and of itself, seem to have much bearing on what I should do. If I am intent on getting some writing done or on doing some other indoor activity and do not otherwise desire or require fresh air, I do not

---

42 One explanation for these entailments is that one might define various forms of “ought” in terms of reasons (cf. Broome 2013).
seem to have a reason to go outside for a walk in the park. If, however, I do desire to take in the fresh autumn air, then I seem to have a reason to go outside. If my desire for fresh air was to somehow wane dramatically, for instance after I get a call informing me that I must have a piece of writing finished before tomorrow morning, my reason to go outside disappears, or, as we might say, it evaporates (Joyce 2001: 35).43

This example illustrates a type of reasons called hypothetical reasons, which depend constitutively on an agent’s ends. Hypothetical reasons are prevalent in many non-moral contexts, such as in practical rationality, but they appear to be a bad fit for moral reasons.44 First of all, it does not seem that moral reasons tend to evaporate in the relevant sense given a change in an agent’s ends. Suppose it is a fact that it is morally wrong to torture innocent persons, and that Sally desires to employ such torture in some context (i.e. takes it as an end), and let us also stipulate that doing so would have no (outweighing) adverse effect on Sally. In such a case, Sally’s ends seem to favour employing torture – that seems to be what her ends give her a reason to do. However, this would not make it any less morally forbidden for Sally to torture an innocent person – if it is morally wrong at all, it will continue to be so independently of the ends of this particular agent. To see that this is so, consider whether we would be likely to retract our condemnation of Sally upon learning that she had certain ends which were fulfilled by engaging in torture, and none which were substantially frustrated. We would not, I strongly suspect, retract our condemnation in such a situation. Moral reasons simply do not appear and evaporate in the way one would predict if they were of a hypothetical character, and where dependent on the ends of an agent (Joyce 2001: 30–7; Olson 2014: 126–130). Rather than evaporate in this way, moral reasons seem to be inescapable.

If moral reasons are not hypothetical, what are they? A possible answer can be gleaned from reflecting on the way in that moral facts seem to provide us with reasons. The supposed fact that torture of innocents is morally forbidden carries with it some implication that any agent should refrain from engaging in this act irrespective of their ends. By Mackie’s platitude, then, this would provide any agent with a moral reason to abstain from such torture, and not because of anything to do with the ends of the perpetrators. That is, no matter which motivational states the agent possesses, the agent nonetheless has a reason to abstain from such acts. Such reasons, which obtain independently of their relation to an agent’s ends, are categorical

43 Joyce uses “evaporate” with respect to hypothetical imperatives (which I will introduce below), but the term can also be applied to hypothetical reasons (cf. Joyce 2001: 121).
44 I will discuss the practical rationality and nature of hypothetical reasons in more detail in Chapter 4.
In other words, they “do not depend for their existence on their being instrumental to the achievement of any of an agent’s desires, goals or cares” (cf. Shafer-Landau 2009).  

Categorical reasons get their force and legitimacy without reference to the attitudes or ends of agents, and as such they appear “objective”, at least in one sense of the term. If moral reasons are understood as categorical, this could explain why all suitably situated people, no matter what their ends are, have a reason (or ought) to $\Phi$. One error theorist who sets his targets on such reasons is Mackie:

[[It] can plausibly be maintained at least that many moral judgments contain a categorically imperative element. So far as ethics is concerned, my thesis that there are no objective values is specifically the denial that any such categorically imperative element is objectively valid. (Mackie 1977: 29)]

Based on this quick sketch of the hypothetical-categorical distinction, moral reasons might appear to be of the latter sort. And since we have seen that moral reasons are ubiquitous in moral discourse, categorical reasons might seem like a good candidate for a core commitment of that discourse.

Despite the apparent appeal of this picture, I will argue in the next sections that the hypothetical-categorical distinction is not the best way of capturing either the core commitments of moral discourse or the nature of reasons or imperatives, moral or otherwise. One subsidiary goal of this chapter will be to show why error theorists, and others who attempt to analyse moral discourse in terms of the hypothetical-categorical distinction should take up a more fine-grained analysis to bring out the metaphysically interesting aspects moral and other forms of reasons.

2.1.3 *Foot’s challenge to moral categoricity and Joyce’s solution*

Exactly how widespread the tendency to align moral requirements, and by extension moral reasons, with categorical ones is underlined by Philippa Foot at the beginning of her article “Morality as a System of Hypothetical Imperatives”:

[[It] is generally supposed, even by those who would not dream of calling themselves his followers, that Kant established one thing beyond doubt – namely, the necessity of distinguishing moral judgements from hypothetical imperatives. (Foot 1972: 305)
Before we move on to look closer at Foot’s challenge to moral categoricity let me introduce the notion relevant notion of an imperative. *Hypothetical imperatives*, such as “You ought to go outside and enjoy the fresh air!”, like hypothetical reasons, derive their force from the ends of the agent to whom it is directed.⁴⁸ If the utterer of the imperative just mentioned were informed that I had an important piece of writing that must be done before tomorrow morning, then she would be likely to retract it, and instead encourage me to keep on writing. Thus, hypothetical imperatives are also evaporating.

Categorical imperatives, on the other hand, command “an action as objectively necessary in itself apart from its relation to a further end.” (Kant 2009 [1785]: 75). A categorical imperative such as “You should not torture the innocent” possesses its force and legitimacy independently of the ends of the person it is directed at, and it does not evaporate given a suitable change in the agent’s ends. As such, someone would be unlikely to retract a categorical imperative of this sort upon discovering that agent had, for instance, a deep desire to perform such an act and would not suffer any negative consequences by doing so.

Based on the distinction between the two types of imperative, Foot argues that moral imperatives are not the only type of requirements that does not seem to be hypothetical:

For instance, we find [a] non-hypothetical use of “should” in sentences enunciating rules of etiquette, as, for example that an invitation in the third person should be answered in the third person, where the rule does not *fail to apply* to someone who has his own good reasons for ignoring this piece of nonsense, or who simply does not care about what, from the point of view of etiquette, he should do. (Foot 1972: 308)⁴⁹

The requirements of etiquette do seem to qualify as non-hypothetical imperatives; they are, as Foot claims, inescapable (i.e. non-evaporating) requirements that obtain independently of their relation to an agent’s ends. But while we might agree that the rules of etiquette are non-hypothetical, are they *categorical*? If it is true that, conceptually, we take categorical imperatives (i.e. ought-statements) to entail there being categorical reasons, then we should expect the same from the imperatives of etiquette. However, the reasons the imperatives of etiquette provides us with, if any, seem *very* different from moral reasons. As we saw in the moral case, “oughts” or “shoulds”, such as “You should not torture the innocent”, are often thought of as having an “all-things considered” force, where it tells you what you, when all is said and done, should do. The claims of etiquette, on the other hand, can seemingly be legitimately disregarded, at least in many contexts. Foot therefore thinks that the proponent of categorical requirements within morality would *deny* that the requirements of etiquette are categorical, and, furthermore deny that such imperatives have any

---

⁴⁸ For Kant’s discussion of imperatives, see *Groundwork to the Metaphysics ofMorals*(2009). Kant uses “imperative” in an idiosyncratic way: “In writing about imperatives Kant seems to be thinking at least as much of statements about what ought to be or should be done, as of injunctions expressed in the imperative mood” (Foot 1972: 305).

⁴⁹ Foot (1972: 308) also points to the rules of games and clubs as being similarly non-hypothetical. For instance, a carrousel at an amusement park might have a rule that “You must be this tall to ride”, a requirement that does not evaporate simply because you want to ride the carrousel, despite being too short. I will return to such rules later.
reason-giving force at all. In short, then, she thinks that the “oughts” of etiquette do not adhere to Mackie’s platitude. On this picture, we therefore have a distinction between three kinds of imperatives:

i) Hypothetical imperatives

ii) Non-hypothetical imperatives
   a. Categorical imperatives (e.g. moral rules)
   b. Non-reason-giving imperatives (e.g. rules of etiquette, games, clubs etc.)

Foot’s challenge therefore puts pressure on the proponent of categorical (i.e. reason-giving) imperatives; in order to justify the claim that moral requirements give rise to categorical reasons while not bestowing that power upon the rules of etiquette, one must come up with some feature that distinguishes them in terms of their reason-giving powers (Foot 1972: 309–310).

Foot goes on to present some candidates for what this special feature – variously called “inescapability”, “binding force” or “necessity” – of categorical requirements could be, but finds none of them satisfactory (1972: 312). She famously terms this missing ingredient “the fugitive thought” and concludes that the most likely answer to why we think of moral requirements in this special way is because we feel so strongly about them – we feel that we must do as morality commands, even if this is only because of contingent psychological factors (1972: 311–312).

Foot thinks that there is no way of legitimately bestowing any form of “automatic reason-giving powers” on moral requirements. She therefore concludes that there are, in fact, no categorical requirements in the reason-giving sense. Since hypothetical reasons are reason-giving, and since they are metaphysically much less strange than moral requirements, Foot takes moral imperatives to really be hypothetical imperatives.50 Furthermore, according to Foot, it is mostly philosophers, not laymen, who have been tricked into thinking that there is something more – “a magic force” – that undergirds moral reasons (1972: 315; cf. Joyce 2001:37).51

Foot’s dilemma to the proponent of categorical moral reasons is therefore this: Either you find something that differentiates requirements like those of etiquette from those of morality, and thereby explain the difference in their reason-giving powers, or you must accept that neither of them are categorical in the reason-giving sense. Foot herself, as we saw, chose the latter horn of the dilemma. Joyce (2001: chapter 2) has attempted to take on the former horn, by showing that there is a feature that distinguishes the reason-giving powers of imperatives of etiquette and those of morality. He does this by arguing that etiquette

50 She later changed her view, and even wrote a “Recantation” to go along with a reprint of her 1972 article in Darwall et al. (1997) where she says: “The idea that morality is a system of hypothetical imperatives is so alien to my present views that I have long thought the positive part of it misconceived” (Foot 1997: 22).

51 Note that if Foot had not held this, but had claimed that ordinary people do indeed think that morality requires categorical moral imperatives and reasons, and, as she does, that no defensible notion of such things exist, she could quickly end up providing support to a form of moral error theory.
consists of weak categorical imperatives, which are only reason-giving in a weak sense, while the imperatives of morality are strong categorical imperatives in that they provide us with “real reasons” (Joyce 2001: 35–44).

Joyce cashes out the weak-strong distinction in terms of what kind of reasons different imperatives give rise to. First of all, note that Foot, by denying the requirements of etiquette reason-giving power, has broken with Mackie’s platitude. She holds that even though the rules of etiquette tell us what we ought to do, such requirements do not seem to entail that we have a reason to comply. But it is open to us to question this claim. It is quite natural to think that there is a sense in which a rule of etiquette such as “Do not put your feet on tables” licences us to speak of a reason. In order to capture this sense, and to avoid conflict with Mackie’s platitude, Joyce employs the notion of institutional reasons, reasons that are grounded in conventional or institutional rules. Joyce claims that such reasons are reason-giving, but only in a minimal, or weak sense, in that they make it linguistically legitimate to speak of “having a reason” (Joyce 2001: 34–37).

It would therefore be legitimate to say that “You have a reason not to put your feet up on the table”, in the sense that this would be correct according to the rules of etiquette. That does not imply anything over and beyond there being a certain set of rules, and that this set of rules says that you should do so-and-so. Such rules do not, then, give us “real” reasons, in an all-things-considered sense, to follow their commands.

Let us grant, as seems plausible, that when speaking from some within some set of rules or some normative framework, saying something of the form “You ought to Φ” entails “You have a reason to Φ” and that this is furthermore a legitimate and linguistically acceptable use of the word “reason” (cf. Joyce 2001: 43). For instance, in many sports it is considered bad sportsmanship to let an opponent win. Assume that some athlete wanted to let his opponent win, but was told that he ought not to do such a thing. When asking “But why?” a likely reply to the athlete would be “Because it’s bad sportsmanship!” That is, it is wrong according to the norms of the institution, in this case, some according to the rules or norms of some sport. This seems like it can legitimately be spoken of as a reason, if only in a minimal or weak sense.

On Joyce’s view, then, there are the following types of imperatives:

iii) Hypothetical imperatives
iv) Non-hypothetical imperatives
   a. Strongly reason-giving categorical imperatives (e.g. moral rules)
   b. Weakly reason-giving categorical imperatives (e.g. rules of etiquette, games, clubs etc.)

Unlike weak categorical imperatives, which engender institutional reasons, “real reasons” are, according to Joyce, only to be found in the dictates of practical rationality (Joyce 2001: 49–52). And practical rationality, Joyce thinks, is exactly where the rules of morality implicitly claim to be grounded. As such, the requirements of morality are supposed to be entirely different from the dictates of institutions and conventions like etiquette. On this basis, then, Joyce thinks that there is a theoretically informative notion of
a categorical imperative and of a categorical reason that singles out what is unique about morality, and which excludes the institutional rules of etiquette and the like, namely strong categorical imperatives.

As Joyce seeks to build an error theory on our commitment to such strong categorical reasons, he does not think that there are, in fact, any such reasons. Roughly, the explanation for this is that practical rationality does not give rise to categorical reasons, so moral reasons, which are both supposed to be categorical and grounded in practical rationality do not exist. Here and throughout, it is important to separate our the question of what our concept of a moral reason is from the question of whether there in fact is any reasons which answer to that concept. Joyce agrees that moral reasons exist in the sense that we have the concept of a moral reason, but holds that no such reason claim is true.

One way to fill in the conceptual claim is therefore to use Joyce’s distinction between strong and weak categorical imperatives, and claim that moral discourse has such strong categorical imperatives (and therefore a corresponding form of reasons) as a core commitment.

2.1.4 Features of normative reasons: Prescriptivity, normative force and inescapability

I think Joyce’s way of answering Foot’s challenge is on to something, but I also think we can do better. To this end, I will introduce three important features of normative reasons that I believe will help bring out the interesting differences between different kinds of such reasons, and which opens the way for a more theoretically attractive way to answer Foot’s challenge.

The first has to do with the feature of normative reasons that makes them normative. For instance, when we have certain ends, and certain ways of satisfying those ends, this is taken to constitute or provide us with a hypothetical reason to do that which would satisfy those ends. We can therefore say that there is something – in this case ends and ways of satisfying them – that implies, indicates, or is a reason for an agent. This feature of telling an agent what she should do (and thereby what she has a reason to do) we might call prescriptivity (cf. Cuneo 2007: 58; Garner 1990: 139).

Since we have allowed that institutional rules, such as etiquette or the rules of games, to give rise to reasons we have set the bar for being prescriptive fairly low. This is because many things in the world, such as children, tell us what to do, and only certain actions would, say, be correct according to the dictates of your toddler. Hypothetical reasons, too, tell us what we should do, but they do so in an especially strong sense, since acting on one’s hypothetical reasons, at least most of the time, constitutes what it is to be practically rational. On the other hand, refusing to do what your child tells you to might very well be perfectly rational. We might call this second feature, the weightiness with which something tell us what we should do their normative force. This notion is meant to capture Joyce’s distinction between the two types of categorical imperatives.

---

52 Cuneo uses the term in this sense (2007: 96–97), as does Bedke (2010: 44n9). Joyce also uses similar terms: “Let us say (as a first approximation) that an imperative has authority if the subject would be irrational in ignoring it, or at
Of course, different reasons of the same type may have different strengths. A putative moral reason not to torture innocents seems, in a vague sense, stronger than a moral reason to give to charity. However, if it is true that you ought, morally speaking, to give to charity, this will still be a weightier reason than any reason provided by etiquette. The point about strong normative force, then, holds independently of whether the moral reason is weak or strong in this other sense, which I will leave vague. I will say more about how we should understand the weak-strong normative force distinction notion later, but it seems clear it seems clear that moral reasons in general possess strong normative force, as does hypothetical imperatives, which spring from practical rationality.

Lastly, the third feature of normative reasons I will discuss concerns whether they are inescapable or if they have the ability to evaporate – that is, whether or not they can disappear given a suitable change in the ends of an agent. As we have seen, hypothetical reasons can indeed evaporate.

We can also classify institutional and moral reasons (and imperatives) according to the above three features; both are prescriptive, in that they tell us what to do, and while institutional reasons have weak normative force, moral reasons have strong such force. Furthermore, they are both inescapable (i.e. non-evaporating). This gives us the following taxonomy of the types of normative reasons we have looked at so far.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Institutional</th>
<th>Moral</th>
<th>Hypothetical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prescriptive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Inescapable</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Normative force</td>
<td>Weak</td>
<td>Strong</td>
<td>Strong</td>
</tr>
</tbody>
</table>

Table 2.1

So far, it might be thought that all I have done is to relabel Joyce’s terminology, by exchanging inescapable-evaporating for categorical-hypothetical and using “normative force” as a label for the distinction between strong and weak categorical imperatives. However, as I hope to show in the next section and in later

least the subject has a reason of genuine deliberative weight to comply” (2006: 62, my emphasis). Some authors wish to introduce a separate terms for those things which only tell us what to do in the institutional or conventional sense. Parfit (2011b: 308-10) has suggested calling such things normative in a “rule-implying” sense, while reserving the term “reason-implying” for the reasons with more deliberative weight. Similarly, Hattiangadi (2007) has, in discussions of semantic normativity, suggested separating “norm-relative” from “normative”. These and similar distinctions should line up fairly close with the distinction between weak and strong normative force which I introduce below.
chapters, it is not the prescriptivity of normative reasons, nor the inescapable-evaporating distinction (i.e. the categorical-hypothetical distinction), that is of most interest if one wants to look at the metaphysical peculiarities of reasons. Rather, I will argue that it is in the distinction between weak and strong normative force that the metaphysically interesting issues reside. And, as I hope to already have made clear, the categorical-hypothetical distinction seems orthogonal to this distinction.

I have now provided an account of normative reasons that is fairly neutral. In the following, I will provide a specific way of cashing out the notion of a normative reason, one that provides us with a better understanding of the weak-strong distinction with respect to normative force. This framework for understanding normative reasons and their force will both provide us with a clearer picture of what the peculiar features of moral reasons are, and they will also open the way for a unification of our understanding of reasons, and in addition provide some highly novel theoretical consequences.

2.1.5 Reasons and normativity: The favouring relations framework

In light of the failure of the hypothetical-categorical distinction to capture the metaphysically interesting aspects of reasons as they occur in moral discourse, we should look for another framework for conceptualizing reasons. Fortunately, an account that seems to capture the nature of our reason talk better than the hypothetical-categorical distinction does has recently been developed. The framework in question has found popularity both among moral non-naturalists and error theorists.53

The account of reasons I have in mind is usually embedded within a larger philosophical programme concerning the nature of normativity, and often goes under the name of reasons primitivism.54 Reasons primitivism can be understood in two separate ways, which it is important to keep apart. First, there is the project that attempts to show that reasons are fundamental within normativity and that all other features of normativity can be reduced to reasons. This project does not say anything about the metaphysical status of normativity itself, such as whether it can, in some sense, be reduced. We might call this claim the reasons first view. The following quote from Joseph Raz summarizes the idea behind this project succinctly:

The normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons. (Raz, 2000: 67)

Secondly, reasons primitivism is also used to label the project of attempting to show that normativity, in the form of irreducibly normative reasons, is primitive. This claim, following Broome (2015), we might call normative fundamentalism. Scanlon (e.g. 2014: Chapter 1) and Parfit (e.g. 2011a, 2011b) have been among the spearheads of this project:

53 Glossing over details, moral non-naturalists can be said to “think that normative properties including moral properties are different enough from the paradigmatic natural properties that it is a mistake to think of them as a kind of natural property” (Roojen 2015: 253). I will discuss such non-naturalism at length in chapter 3 and 6.

54 For a critique of this general program, see Väyrynen (2011).
I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favour of it. ‘Counts in favour how?’ one might ask. ‘By providing a reason for it’ seems the only answer. (Scanlon 1998: 1)

[T]he concept of a reason is indefinable in the sense that it cannot be helpfully explained merely by using words. (Parfit 2011a: 31)

Understood as the combination of the projects of reasons first and normative fundamentalism, reason primitivism attempts to show that all normative features can be reduced to (normative) reasons, and furthermore, that such normative reasons are fundamental and cannot themselves be further explained. What we then get, is the idea of irreducibly normative reasons. And as we saw, such reasons are taken to be primitive.

Whether or not one buys into the larger programme of reducing all of normativity and value to reasons, or whether one holds normativity to be fundamental, the framework that has sprung out of this project can help illuminate the notion of a moral reason (cf. Cuneo 2007: 65). Before I specify more exactly how this framework can help clarify reason talk, let me first introduce two ways of understanding what it is to be a reason.

According to this framework, and in line with the quote from Scanlon above, the idea of a normative reason consists in there being some consideration (e.g. a fact), that counts in favour of (or counts against) something (cf. Broome 2013: 54). The term “reason” is therefore used about a consideration that favours (or disfavours) something. Being a reason, therefore, applies to entities such as facts. This is the sense of “reason” that Cuneo (2007: 65) has in mind: “Let it be noted that I have claimed that it is the grounds [i.e. facts] themselves that are reasons. I have not said that reasons are the favouring relation or that they are the grounds in addition to the favouring relation.”

To take an example: Imagine that you see a co-worker in great pain. It is natural to think that there would be a reason to help your co-worker. What would this reason be? The fact that your co-worker is in pain seems like a plausible answer. In other words, the fact that your co-worker is in pain counts in favour of helping him (hereafter, just “favours”), and it is the fact itself – that your co-worker is in pain – that is a reason.

I will call this use of “reason” the reason property sense. In such cases, “reason” is used as a monadic predicate.55

\[ \text{REASON PROPERTY} \quad R(x) \]

---

55 A monadic predicate expresses a property of an entity – e.g. predicates the property being a reason of some fact. The use of “reason” in the property sense (i.e. as monadic predicate) is found in Olson (2014) and Cuneo (2007).
This can be read as “x is a reason”. Given this definition of a reason, it becomes clear that it is not necessarily reasons themselves – as facts – that are either normative or irreducible. After all, a natural fact can be a reason in this sense, and such facts need not be irreducible or normative. Although it is correct that this use says of a fact that it is a reason, there is clearly also another element that is entailed by such uses. Such accounts specify that in order to be a reason, the fact must count in favour of something. In order to see what is normative about normative reasons, therefore, we have to look closer at this dimension of reasons.

We can begin by noting that it seems likely that reasons are always had (or owned) by someone, and that they are always a reason to respond in some way. This gives rise to another sense of the term “reason”, a sense in which a reason is a relation between multiple entities. A reason on this construal is therefore a relation between at least three things: a fact, an agent and a behaviour. As we saw earlier, the framework we are now looking at suggests that the relation that relates these three entities is a favouring relation. For instance, a fact such as your co-worker being in pain can be said to favour you helping your co-worker.

I will call this the reason relation sense of “reason”. “Reason” is here used as a polyadic predicate. More specifically, I will take this use of the term to capture a four-place predicate, as I will also include a particular context, in addition to a fact, an agent and a behaviour. This relation can then be symbolized as follows, where \( F \) is some fact, \( A \) is an agent, \( \Phi \) is an action and \( C \) is a context:

\[
\text{REASON RELATION} \quad R(F, A, \Phi, C)
\]

This can be read as “\( F \) counts in favour of (or disfavours) \( A \)’s \( \Phi \)ing in context \( C \)” (Bedke 2010: 48). As we saw previously, reason primitivists like Scanlon and Parfit sometimes seem to claim that being a reason in the property sense cannot be defined or explained further, and therefore seem to imply that a reason in the sense of being a fact (which favours something) must be taken as primitive and undefinable. However, if we assume that all reasons are had by someone, and are reasons to do something, we can define what it is to be a reason (in the property sense) in terms of something being a reason for someone to do something – i.e. in terms of the reason relation. This can be done as follows:

---

56 Compare the term “gift”. It can be applied to individual entities (“x is a gift”), and, at the same time, such an ascription entails a relation: In order for it to be true that x is a gift, there must be some entities, A and B such that A gifts x to B. Thanks to Hans Robin Solberg and Ainar Petersen Miyata for discussion on this point.

57 This claim is not central to my thesis, but simplifies exposition.

58 A polyadic predicate expresses a relation between two or more entities. Bedke (2010) use “reason” in this sense.

59 Exactly how many places one understands the reason relation to have varies from author to author. For instance, Bedke (2010) suggests, but is not committed to, there being three, while Scanlon (2014: 31) and Cuneo (2007: 65) claim that there are four places. I am not committed to the exact number of places suggested in the main text, only that it is a relation with a structure similar to this.

60 This is not intended on their part, as becomes clear from reading the passages in context.
For it to be the case that F is a reason is for it to be the case that, for some A, Φ, and C, F is a reason for A to Φ in C.\textsuperscript{61}

This is helpful, since it is the reason relation sense of the word “reason” that I will focus on, and I will therefore use the word “reason” in the reason relation sense, unless I specify otherwise.\textsuperscript{62} Similarly, I will use the term “favouring relation” or “reason relation” to capture this same concept.\textsuperscript{63} Given this, when I call something “a reason”, “a favouring relation” or “a reason relation” I will have in mind the (favouring) relation that relates a fact, an agent, a behaviour and a context. A last clarification is that my use of “favouring relation” is also meant to include disfavouring relations.

2.1.6 Applying the framework: Reducible and irreducible favouring relations

Having now laid out the general structure of this framework for understanding reasons, we now want to see if we can use it to conceptualize moral reasons in a profitable fashion. To this end, let us look at a situation where a moral reason seemingly is in play.

\textit{FACTORY} While working with mechanical equipment in a factory, you hear a scream, and discover that a co-worker has gotten his arm stuck in a machine and is writhing in pain. You know that the machine has a button that will cause it to retract its mechanical plate and free your co-worker. You know that doing so will be beneficial for you and your co-worker and you are in a situation to press the button without putting yourself in danger.

In this scenario, most people, and most normative ethicists, will, I believe, hold that you have a moral reason to press the button and release your co-worker. For instance, this could be because one is committed to an ethical principle such as “if it is in our power to prevent something bad [e.g. pain] from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it” (Singer 1972: 231). Let us therefore assume there is a moral reason in play in this situation. How should we understand its conceptual structure?

\textsuperscript{61} The idea of defining the notion of a reason in terms of the reason relation crystallized in my thinking during a talk from John Broome held at CSMN, November 2, 2015, “Reasons fundamentalism and what is wrong with it”, where he emphasize just this point. The specifics of the formulation are taken from him, though I have introduced some changes.

\textsuperscript{62} Cuneo (2007:65) cautions against this use; he claims that it would not be in line with ordinary use of the word “reason” to respond to the question “What reason did you have to Φ” by answering “The fact that F favoured my Φing given the circumstances is my reason for Φing”. I must admit that I see no particular problem with this use in a technical context such as this. See also e.g. Bedke (2010) for this use of the term.

\textsuperscript{63} For our purposes here I will take these two as synonymous, though nothing of importance hangs on this terminological choice. If one is unhappy with it, read “reason relation” for all relevant instances of “favouring relation”.

31
First of all, there is the fact that your co-worker is in considerable pain (let us keep the presumed favouring out of the picture for now). There is also an agent involved that is capable of acting in a way that would reduce the pain of the co-worker, and there is also a particular course of behaviour involved in doing so. Additionally, there is the specific context, which includes the agent knowing that following the pain-reducing course of behaviour is practically risk-free. So far, then, we have a fact (your co-worker being in pain), an agent (you), a course of behaviour (pushing the button) and a set of circumstance (e.g. no risk involved in pushing the button).

These things in and of themselves do not seem enough, conceptually speaking, to constitute a reason to do anything. They are simply a natural fact, a person, a course of action and a context. In other words, these things in and of themselves do not seem to tell us what to do, since there is nothing here resembling prescriptivity, normativity, action-guidance or any of the other things we associate with moral reasons. We might feel as if the reason to act simply springs from these ingredients, but that is most likely because there is another element that we implicitly add to the conceptual mix. This additional element has to do with how all of the aforementioned ingredients are related or structured.

What is missing, then, is the favouring relation between the fact, the agent and the action in the context. Since the circumstances you find yourself in are such that you can do something about your co-worker’s pain without any real cost to yourself, the fact that your co-worker is in considerable pain seems to require or demand that you, the agent, act in way that helps him, i.e. by pushing the button. In other words, we take the fact to favour acting in a certain way, and the resulting relation in turn constitutes the reason. If we did not take the fact to favour acting in this way, it would be unclear why there would be a reason involved. Such favouring thus seems to capture what we mean by there being a moral reason in this scenario.

However, as is often pointed out, such talk of a fact or a situation “requiring”, “demanding” or “favouring” is rather vague and metaphorical (cf. Olson 2014: 105). We should therefore strive to make it at least a little clearer, though perhaps we cannot expect much. One way of attempting to clarify this notion is by structuring the elements of FACTORY in the following way.

**The conceptual structure of a moral reason**

<table>
<thead>
<tr>
<th>Fact</th>
<th>Circumstances</th>
<th>Favours</th>
<th>Agent</th>
<th>Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-worker in pain</td>
<td>Pushing alleviates pain</td>
<td>You</td>
<td>Pushing the button</td>
<td></td>
</tr>
</tbody>
</table>

A moral reason

There is a moral reason for you to push the button.
It is worth noting that the fact entering into the moral reasons relation does not itself have to be inherently normative in any way. They can be naturalistically and metaphysically unproblematic, such as the fact that someone is in pain (cf. Bedke 2010: 51–53).

Moral favouring relations seem to hold independently of an agent’s ends. That is, independently of whether you desire to help your co-worker or not, it seems that the moral favouring relation obtains, and it favours you acting in a specific way. Moral favouring relations are therefore inescapable for the agent involved, in the sense that the agent cannot release herself from taking part in it simply by a change in her ends. Furthermore, such favouring relations are prescriptive: they tell us what we should do. Lastly, they seem to have strong normative force.

If the account presented above is a good fit for moral features, it should be able to tell us something about why moral reasons so often have been taken to be peculiar, to the point of causing many philosophers to be sceptical of their existence (cf. Gibbard 2012: 10). We should also expect it to give us a more natural taxonomy of the kinds of normative reasons there are than the hypothetical-categorical distinction did.

Let us therefore take another look at the putative moral reason in FACTORY and see if we can discern any clues as to what is held to be peculiar about such reasons.64 Is it the fact that your co-worker is in pain that has raised suspicion? It is clear that it is not. One must of course have an ontology that includes pain, but there does not seem to be anything particularly metaphysically troubling about this fact. Possible courses of behaviour do not seem to license much metaphysical suspicion either, at least not for the moral domain in particular. The existence of agents and behaviours likewise does not appear to bring anything particularly suspicious with it, and the same can be said for a set circumstances. Again, there might be theoretical difficulties in spelling out these notions, but nothing that affects moral reasons in particular.

What remains, as I have said, is a way to relate these things – a relation where a fact favours (or disfavours) an agent performing an action in a context. The question therefore arises as to exactly what kind of relation this is. To begin, we should note that while the notion of a reason in the property sense was found to not be primitive, the notion of “favouring” plausibly is. We should nonetheless aspire, I believe, to illuminate it a bit further.

The favouring relation is not a causal relation. It does not require that an agent in fact perform the course of behaviour in the context, given that the fact obtains; the relation would still hold even if the agent did not act in this way. Nor is it a psychological connection that claims that an agent would be motivated to perform the action in the context, given that the fact obtained. And it is not a relation that merely specifies what would be correct according to some institution or set of rules. The putative moral reason is not like the institutional reasons that spring from etiquette or the rules of games – it is not simply a specification of what would be correct according to some institution or convention, but which carries little or no deliberative

64 This pedagogical tactic is employed in Bedke (2010).
weight. Of course, certain things are correct according to the requirements of morality, but it is something over and above this correct-according-to relation that grants moral reasons their strong normative force. In order to try to say something about what this relation is, rather than what it is not, let us first contrast the moral reason relation with another type of reason relation.

Consider the following rule of chess:

\[
\text{CHESS} \quad \text{The bishop may move to any square along a diagonal on which it stands.}
\]

As we have seen, the rules of games give rise to an institutional reason to comply with them; the fact that the rules of chess dictate that bishops can only be moved diagonally provides, indicates or is a reason for moving the bishop diagonally and only diagonally (cf. Olson 2014: 120–122; Joyce 2001: 39–41). On the favouring relations account, the rule can therefore be said to favour agents only moving the bishop diagonally. 65

Now, you may have a desire to win the game, and therefore not to be disqualified on the grounds of performing an illegal move, and you might therefore have a hypothetical reason not to move the bishop any other way than diagonally. But let us ignore that for the moment, and simply focus on the supposed reason which arises from the rules of chess in and of themselves. Let us therefore try to represent the institutional reason involved in CHESS.

The conceptual structure of an institutional reason

<table>
<thead>
<tr>
<th>Fact</th>
<th>Circumstances</th>
<th>Favours</th>
<th>Agent</th>
<th>Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules of chess</td>
<td>-</td>
<td></td>
<td>You</td>
<td>Bishops moving diagonally</td>
</tr>
</tbody>
</table>

An institutional reason

There is a reason to move bishops diagonally from the standpoint of the rules of chess.

The fact that the rules of chess state that a bishop may legally only move diagonally gives anyone, no matter their interests, a reason to only move the bishop diagonally. It provides a reason because of the fact that it is legitimate to say “According to the rules of chess you ought only to move the bishop diagonally”.

To see why such rules are different from, say, hypothetical imperatives, consider what we would say to someone who did not want to move the bishop diagonally, but rather horizontally. Would we say, “Well,

65 In a different sense, it might be true that in order to play chess at all one must only move the bishop diagonally. Such constitutive rules might give rise to institutional reasons in the same way as regulative rules. For discussion of such different types of rules, see Searle (2010: 9–10).
then, in that case, the rules of chess no longer tell you to move the bishop diagonally”? No, we would not, and that is why such requirements and reasons are inescapable.

The rules of chess therefore provide an institutional reason to move bishops diagonally, and this in turn can be understood as the claim that the fact favours a certain course of behaviour of agents in some, or all, contexts. Of course, such institutional reasons do not carry much, if any, deliberative weight (i.e. they have weak normative force), and they only supply an agent with a reason in the sense that they give you a recipe for acting correctly according to a given set of rules or norms. We can therefore see that the favouring relation involved in such institutional reasons can be reduced to something else – i.e. to a correct-according-to relation.

While institutional reasons and rules are clearly non-hypothetical, in that they are inescapable, they do not really give us a reason to do any particular thing; they just specify what one must do in order to fulfil certain roles or participate in certain rule-governed activities (Olson 2014: 121). Furthermore, the reasons which spring from institutional favouring relations are clearly prescriptive, in that they tell us what to do, and inescapable, in that they obtain independently of their relation to an agent’s ends. And we have seen that their weak normative force stems from the favouring relation involved being reducible to a correct-according-to relation. There does not seem to be any particular mystery how there could come to be rulebooks or institutions which have their own set of rules (see Searle 2010). It therefore seems relatively easy to explain both why this is so, and also what it consists in – it consists only in there being certain rules according to which certain things are correct while others are not.

The important difference between moral and institutional favouring relations does not hinge on their being inescapable or not, but rather on the way we cash out the favouring relation involved in these types of reasons. In the case of institutional reasons, the favouring relation can be understood in terms of something else; it can be reduced to something familiar and seemingly unproblematic, metaphysically speaking.

With moral reasons, however, things are different. There, the favouring relation is not a correct-according-to relation, where there is no further reason to comply with the moral rulebook, so to speak. Rather, in the case of moral reasons, there is a primitive, unexplained favouring going on, of a unique sort that is sui generis. It does not seem possible to explain this type of favouring in terms of something else, as reason primitivists have pointed out (e.g. Scanlon 1998: Chapter 1, 2014: Chapter 1; Parfit 2011a: 31). It therefore seems that it must simply be a primitive fact about the universe that certain facts favour an agent performing certain behaviours, and, consequently, this type of normativity is irreducible.

The irreducibility of this form of normativity explains why moral reasons have strong normative force. While institutional reasons are only backed up by a correct-according-to relation, moral reasons are backed up by an irreducibly normative favouring relation. We might therefore begin to speculate that the normative force of a form of reasons is determined by whether they are irreducibly normative or not. This is
a claim I will defend and elaborate further in Chapters 4 and 5. If we now apply the earlier table to reason relations, and exchange normative force for irreducible normativity, we get the following matrix:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Type of reason</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Institutional</td>
</tr>
<tr>
<td><strong>Prescriptive</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Inescapable</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Irreducibly normative</strong></td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2.2

This alternative way of conceptualizing reasons opens up new theoretical questions, such as whether hypothetical reasons are irreducibly normative or not; if they are, that would be highly surprising, and if they are not, we must be able to find some way of reducing the favouring relation they involve. Either way, that promises to shed new light on the nature of reasons.

I hope to have already shown that the favouring relation framework is an improvement over the categorical-hypothetical distinction when it comes to conceptualizing how the notion of a moral reason is used in moral discourse. We have seen that by understanding moral reasons as instances of irreducibly normative favouring relations, we have also found a more elegant solution to Foot’s challenge to moral categoricity than Joyce’s separation of categorical reasons into weak and strong.

I will further support this claim in later chapters by showing that understanding normative reasons in the way outlined above can illuminate the nature of various forms of reasons, including practical and epistemic reasons (in Chapters 4 and 5, respectively), and that it does so in a way that is highly surprising and interesting for error theorists, but also for metaethicists and metanormative theorists of all stripes.

Furthermore, irreducible normativity, in the form of irreducibly normative favouring relations, indeed looks like it could be used as a target for moral error theorists, since it appears to be a core commitment of ordinary moral discourse and appears metaphysically unique, if not outright mysterious. Such relations therefore seem both to fall out of an adequate description of ordinary moral discourse, and they also help bring into sharp relief what has traditionally been seen as peculiar about moral reasons.
2.2 Can we reject the conceptual claim?
We now have what looks like the beginning of a powerful argument on behalf of moral error theory. In the foregoing, I have tried to defend the claim that ordinary moral claims, such as “You ought to Φ”, “You should not Φ” etc., entail a commitment to moral reasons, and, furthermore, that such moral reasons entail irreducibly normative favouring (and disfavouring) relations. More precisely, the claim is that our concept of a moral reason, as it figures in ordinary moral discourse, cannot be satisfactorily accounted for without invoking irreducibly normative favouring relations. We have therefore located a core commitment of ordinary moral discourse – irreducibly normative favouring relations – and it seems like such relations could very well be metaphysically troublesome. We therefore seem to have successfully filled in the conceptual claim:

\[\text{CONCEPTUAL CLAIM} \quad \text{Moral discourse has a core commitment to irreducibly normative favouring relations.}\]

At this point, many metaethicists would want to stop the error theorist, and claim that the theory, and in particular the conceptual claim can be rejected. One such group of metaethicists is moral naturalists, who hold that there are “objective moral facts and properties and [that] these moral facts and properties are natural facts and properties” (Lenman 2006).66 Such naturalists attempt to reject the conceptual claim by arguing that moral reasons do not necessarily have to be understood in terms of metaphysically unique properties, such as irreducibly normative favouring relations. Instead, they claim that we can perfectly well account for the commitments of moral discourse in a naturalistically acceptable manner.67

Against this, moral error theorists usually argue that there are no adequate naturalistic theories of morality (as ordinary moral discourse takes it to be) currently on the table. Until such an account is forthcoming such accounts are simply less satisfactory than those provided by, say, error theorists or non-naturalists for reasons outlined in the Appendix (Olson 2014: 126–135; Joyce 2006: 190–209; though contrast Enoch 2011: 109).

---

66 Note that “moral naturalism” (and “moral naturalist”) is sometimes used loosely, in the sense of a (person who holds a) theory of morality and moral discourse that is naturalistically acceptable. Under this definition, moral error theory is an instance of a moral naturalist theory, as it does not have any inherent commitments that are naturalistically unacceptable. In metaethics, however, a moral naturalist (theory) is more precisely used of someone who holds a theory of the kind described by the definition given in the main text, and would therefore constitute a form of moral realism, something which error theory is certainly not.

67 Of course, if the naturalist could account for irreducible normativity in a naturalistically acceptable manner, then if the naturalist could accept the conceptual claim while rejecting the substantial claim (Olson forthcoming). For some arguments that the naturalist cannot account for this, see the Appendix.
As my goal in this thesis is not primarily to defend the truth of error theory, and therefore defend it against all possible rivals, I will not say much about this debate here. I will however, present some reasons for thinking that no adequate moral naturalist account is forthcoming in the Appendix.

Assuming that the conceptual claim has been established, it is therefore time to move on to fill in and defend the substantial claim by looking at the ways in which irreducibly normative favouring relations might be metaphysically objectionable. This is the task of the next chapter.

3 Moral error theory II: An Argument from Queerness

The … lines of argument we’ve considered hinge on the claim that, were moral facts to exist, they would be objectionably ‘mysterious’ and, hence, we have excellent reason to believe they do not exist. While it is fairly commonplace to hear philosophers say such things, it is not always so obvious what is meant when they do so.

Cuneo (2007: 101)

3.1 Filling in the substantial claim

I have now filled in and tentatively defended the first, conceptual claim moral error theories make. Having done so makes it relatively straightforward to begin filling in the second, substantial claim. Together, these two claims provide an error theoretic conclusion.

AN ARGUMENT FOR MORAL ERROR THEORY

Conceptual claim  Moral discourse has a core commitment to irreducibly normative favouring relations.

Substantial claim  There are no irreducibly normative favouring relations.

Conclusion  Moral discourse is systematically untrue.

The substantial claim, then, is that there are no irreducibly normative favouring relations. As I argued in Chapter 2, such relations are a core commitment of moral discourse, in that they are entailed by the existence of moral reasons, which are themselves entailed by moral ought- and should-claims, and plausibly many others as well. If there are no such irreducibly normative favouring relations, all judgements concerning what should or should not be done, what we have or do not have moral reason to do, etc., are untrue. For instance, the judgement, “You should not steal” would entail that there are (moral) reasons not to steal, as per Mackie’s platitude discussed in Chapter 2. And such reasons themselves entail irreducibly normative
disavouring relations that involve stealing. If there are no such relations, moral discourse would therefore be systematically untrue.\textsuperscript{68}

Of course, it is not sufficient for the error theorist to just claim that there are no irreducibly normative favouring relations – further argument is needed to secure the substantial claim. That is what I will provide in this chapter in the form of an \textit{argument from queerness}. Such arguments have been among the most discussed arguments in metaethics, and employing them is a standard strategy when defending the substantial claim of different versions of moral error theory (e.g. Mackie 1977: 38–42; Joyce 2001: 30–31; Olson 2014: Chapters 5 and 6; Garner 1990). However, exactly how such arguments are supposed to work has been the source of much debate. In what follows I will attempt to make this clearer than is usually done, by taking some cues from Mackie’s (1977) original argument, and Olson’s (2014) attempt at improving upon it.

\subsection*{3.1.1 An argument from queerness}

Until now, we have operated with the vague suspicion that there is something odd, and perhaps objectionable, about irreducibly normative favouring relations. It is now time to flesh out this suspicion into something more tangible. One way of doing so is to take a page (or, actually, four) from Mackie and argue that the odd thing about such relations is that they are \textit{queer}, metaphysically speaking.\textsuperscript{69} Mackie introduces his discussion of the argument from queerness as follows:

\begin{quote}
    If 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. (1977: 38)
\end{quote}

Based on the conclusion of the previous chapter, it certainly seems that Mackie was right; there is undeniably something strange about the type of relations we have found to undergird moral reasons and moral discourse more generally.\textsuperscript{70} The main goal of focusing on the queerness of moral features lies in using this as leverage in order to establish a presumption against the existence of such features. The queerness should, in other words, make us ontologically suspicious, and the overall goal of the argument from queerness is to secure the claim that we should deny such features a place in our ontology.

\textsuperscript{68} Depending on the extent to which one can make good on the claim that moral claims involving terms such as “good”, “bad”, “right”, “wrong” etc. entail reason claims, one might secure either the claim that \textit{most} of the judgements of moral discourse are untrue, or that all such judgements are untrue (cf. Joyce 2001: 175–177). Remember that we in Chapter 1 defined error theory as compatible with both these options.

\textsuperscript{69} As far as I know, this use of “queer” stems from Mackie (e.g. 1946; 1977). Despite being somewhat outdated, I will keep the term in order to preserve terminological continuity with the literature on the topic. For Mackie’s discussion of queerness, see Mackie (1977: 38–42; cf. 1946: 78).

\textsuperscript{70} A terminological note: In this chapter, I will move swiftly between talk of moral features, facts, truths, relations and properties. By this, I do not mean to go beyond the discussion of previous chapters, and all such talk should be cashed out in terms of the obtaining of moral favouring relations.
I take the argument from queerness to consist of three parts: a queerness claim, a debunking explanation, and a plausibility claim. The queerness claim is supposed to establish that some feature is queer, in a metaphysically objectionable sense. The debunking explanation is then supposed to give an alternative explanation of how the feature in question came to be central to our moral practice, without in any way relying on the existence (or instantiation) of the feature in question. Together, the claims above support a plausibility claim, which says that, all things considered, it seems that we would do better to reject the existence of the metaphysically queer feature rather than to accept it in our ontology. From this, I will claim, we can conclude that the features in question are non-existent. The argument might therefore in general be construed like this:

**ARGUMENT FROM QUEERNESS**

**Queerness claim**
Fs are metaphysically queer.

**Debunking explanation**
An explanation of why we are likely to believe there are Fs, which in no way depends on there actually being any Fs.

**Plausibility claim**
Given all available evidence, it seems preferable to reject the existence of Fs than to accept them in our ontology.

**Conclusion (the substantial claim)**
There are no Fs.

Before I go on to discuss how we should fill in the details of this argument, it is worth noting the role the argument from queerness is meant to play in metaethical debates. As we saw in the last chapter, moral naturalists usually do not think that irreducibly normative favouring relations are centrally involved in moral discourse. For all they care, therefore, the argument from queerness may very well be entirely sound – it will not matter to them. The same goes for non-cognitivists. The dialectical target of the argument from queerness are those who accept the conceptual claim and who in addition are committed to the existence of the core commitments of moral discourse (cf. Olson 2014: 79–83). Typically, non-naturalist realists fulfil this description, and they are therefore the argument’s main target. In this chapter, I will mainly use Enoch’s (2011) recent and refreshingly clear and honest statement of non-naturalism as a foil, though most points will apply to the majority of other forms of non-naturalism as well.

---

71 Mackie takes the argument to consist of “two parts, one metaphysical, the other epistemological” (1977: 38), and so does Olson (2014). I add the plausibility claim in order to explicate the abductive nature of the argument and to show that all possible evidence should be included in the inference. Olson uses “queerness argument” for what I have termed a “queerness claim”.

72 It could be argued that this argument does not license the conclusion that such relations do not exist, but only the weaker conclusion that we should neither believe that they exist nor that they don’t (Morton and Sampson 2014). That is to say, the argument might show that we are not justified in believing them to exist, but neither are we justified in believing them not to exist. I will not pursue this line of argument here, but it seems to me that this weaker conclusion, too, could be used to support a non-standard form of moral error theory or, at the very least, a more traditional form of moral scepticism (cf. Joyce 2014a; Sinnott-Armstrong 2006).
In this Chapter, we are exploring whether an argument from queerness will make moral error theory more plausible than non-naturalist realism. It is therefore important to keep in mind a central feature of theory choice, which Enoch (2011: 165) makes vivid: Theory choice is a matter of counting the overall plausibility points of the different available views, and it is therefore necessary to take all relevant evidence into account. I will return to this point in the section on the plausibility claim.

3.1.1.1 Queerness claims

One of the central debates between the non-naturalist and the moral error theorist revolves around whether the core commitments of moral discourse (i.e. irreducibly normative favouring relations) are sufficiently queer to deserve to be excluded from our ontological inventory. Error theorists argue – through what I will call queerness claims – that certain moral features are metaphysically queer, in the sense that they are objectionable. Non-naturalists often oppose such claims. How are we to decide?

First of all, we must get clear on what the alleged queerness is supposed to amount to. This has proved difficult to spell out with any precision, and rebuttals of the argument often focus on the lack of clarity concerning what the charge of queerness is actually meant to consist in. In due time I hope to improve upon this situation by being more specific than is usual in the literature on queerness. However, in order to say something about how I think we should understand queerness, it will be illuminating to first look at how we should not.

Some critics of the argument from queerness have taken Mackie’s claim that moral features would be “of a very strange sort” to simply mean that they seem a bit unusual or strange. This way of spelling out the queerness claim, however, would leave the error theorist vulnerable to the following type of reply:

> Of course there are entities that meet these criteria. It’s true that they are queer sorts of entities and that knowing them isn’t like anything else. But that doesn’t mean that they don’t exist. ... For it is the most familiar fact of human life that the world contains entities that can tell us what to do and make us do it. (Korsgaard 1996: 166)

Similarly, but somewhat more to the point, is the attempt to come up with things other than moral features that also appear strange to us.

> The strangest thing about [the argument from queerness] is the claim that it is an argument. … The world is a queer place. I find neutrinos, aardvarks, infinite sequences of objects, and (most pertinently) impressionist paintings peculiar kinds of entities, but I do not expect nuclear physics, zoology, formal semantics or art history to pay much regard to that. (Platts 1980: 72)

As Mackie puts it, coming up with other things that seem equally queer “is an important counter to the argument from queerness” (1977: 39). However, it is not enough to simply point to things that appear a little

---

73 Attempts to spell out the notion of queerness can be found in e.g. Brink (1984), Garner (1990), Joyce (2001: 30–31), Cuneo 2007: 101–103), Shepski (2008), and Olson (2014: chapter 5 and 6).
strange to the casual observer. If someone pointed out to Mackie that neutrinos also seem like “oddball entities” (Horgan and Timmons 1991: 449), he would simply say that

[t]he only adequate reply to it would be to show how … we can construct an [adequate] account of the ideas and beliefs and knowledge that we have of all these matters. (1977: 39)

So, while we might all occasionally be struck by the strangeness of some of the things that we take to inhabit our universe, that fact alone – that something seems strange to us – does not, in and of itself, seem like a good reason to deny it a place in our ontology. As Mackie makes clear in the above quote, the queerness arises when we are not able to give an adequate account of the entities in question. However, many of the things that might appear strange to us – including neutrinos – can be shown to fit rather nicely into our best account of the world, for instance, because they fit into our explanatory frameworks or because they are postulated by our best science. Such things will therefore not help relieve the charge of queerness levelled at moral features.

This type of reply is nonetheless on the right track, since, as Mackie points out, “the best move for the moral objectivist is … to look for companions in guilt.” However, the companions the non-naturalist must look for are things that are not easily accounted for in this way; and if it turns out that they cannot be adequately accounted for, they must go:

If some supposed [entities] resist such treatment, then they too should be included, along with objective values, among the targets of the argument from queerness. (1977: 39)

An important consideration when deciding whether something is metaphysically queer or not is therefore how the supposedly queer features “fit into our picture of the world”. If the features fit nicely into, say, the picture of the world our best science provides us with, then there would probably not be much cause for metaphysical concern. From this, we can conclude that the relevant notion of metaphysical queerness should not be understood to simply consist in what David Enoch has termed sheer queerness (2011: 134–135), where this is taken to mean only that something evokes a feeling of strangeness in us.

Note that we might, for all this, very well agree that neutrinos are queer, but that such sheer queerness is not enough to qualify as metaphysical queerness. What we are after is a sense of queerness that will make us doubt the existence of the entities it applies to. One way to give a positive characterization of

74 Mackie is here writing from an empiricist standpoint, which takes it for granted that it must be capable of receiving an empiricist account. As I will explain below, that is not of the essence for a successful argument from queerness, and we should abstract from it.

75 It might be thought that, in particular, mathematical objects and our knowledge of them (e.g. infinite sequences of objects) will be particularly difficult to account for. That may be so, and in that case, see the next paragraph. As for Aardvarks and impressionist paintings, those are things we encounter face to face, which seems like good prima facia evidence of their existence.

76 Enoch (2011: 134) correctly notes that one could pack in more features in the queerness argument, which would make it stronger. This is what I do below.
what such queerness consists in is what we saw Mackie gesturing at earlier: Something is metaphysically queer if it does not adequately fit into our best, current understanding of things.

This, however, raises the question of what our best, current understanding of things is. Arguably, the most trusted and reliable guide to what there is, is our best current science. While this is not to say we must assume without evidence or argument that science is the be–all and end–all when it comes to what we count as evidence for taking up ontological commitments, it certainly has a better track record than most other epistemic enterprises. One way of answering the question of what we take there to be in the world, therefore, is to take the viewpoint of a form of naturalism, where this is understood as the claim that when it comes to deciding which things we should populate our ontology with, this question should be delegated to our best current science.

Sometimes, then, the charge of queerness simply consists of this, that the entities in question are not held to exist by our best current science. This is one line of thought in Mackie, who sometimes seems to think that clashing with empiricism is enough to qualify an entity as metaphysically queer (e.g. 1977:39). Many other error theorists are also motivated by similar concerns, though usually in terms of naturalism rather than empiricism (cf. Olson 2014: 86, n. 20; Cuneo 2007: 101–103).

Despite this tendency, it would be disappointing if the argument from queerness, which, as we saw, is meant to target the non-naturalist, presupposed a commitment to naturalism. Fortunately, there seems to be room to construct queerness arguments that do not rely on naturalism as a premise:

Even philosophers who are not naturalistically inclined and endorse ontologies inclusive of things like Cartesian souls, Leibnizian monads, irreducible qualia, or abstract entities existing outside the spatiotemporal realm may agree that moral properties and facts are queer ... In short, one need not endorse ontological naturalism in order to find Mackie’s argument from queerness compelling or attractive. (Olson 2014: 86)

In order not to beg the question against the non-naturalist, there must be some independent argument for why postulating the existence of queer features puts pressure on the non-naturalist. Let us therefore try to further home in on exactly what it is about moral features, such as irreducible normativity, that makes them queer, without relying on the truth of naturalism.

While not speaking about irreducibly normative favouring relations specifically, Richard Garner has pointed to one of the reasons why moral facts have seemed mysterious: “Moral facts are unusual in an unusual way – they demand” (Garner 1990: 143). Furthermore, as we saw, the favouring relations involved in moral reasons – the irreducibly normative kind – are taken to be primitive and unexplained. It does not seem possible to explain this type of favouring in terms of something else, as reasons primitivists have pointed out (e.g. Scanlon 1998: 1; Parfit 2011a: 31). Such relations and properties also look unique, i.e. *sui generis*. There does not seem to be anything else that we know of that has a similar property of being
irreducibly favouring, and consequently it must simply be a primitive fact about the universe that certain facts favour an agent performing certain behaviours in certain contexts.

Let us see if we can pinpoint something metaphysically queer in all of this. First, there is the primitiveness (or inexplicableness or fundamentality). However, primitiveness in itself cannot be objectionable, since, for (almost) any given theoretical framework, something, of necessity, must be taken as primitive or fundamental. What about uniqueness? The relations in question, as Mackie says, fundamentally different from everything else in the universe. However, this might be thought to hold for any number of things – any fundamental property of, say, physics might be taken to be fundamentally different from any other such property. The fundamental properties of physics – or neutrinos for that matter – are perhaps strange, but they are strange things that are introduced because they fit into an explanatory scheme. They fit comfortably, or more comfortably than any alternative, within our best (scientific) theories.

This is arguably not the case with moral features (on a non-naturalist account). Irreducible normativity would seemingly have to be added to a naturalist’s ontological inventory without any possibility of explanatory integration (Harman 1977: Chapter 1, 1986). There would be even less integration with scientific theory, as non-naturalist theories make moral features escape scientific investigation by definition (cf. Ridge 2014: §3). Most non-naturalists, therefore, hold that moral features do not even figure in the best explanations of why we have the moral beliefs that we do (Enoch 2011: 175). The reason for this is primarily that non-naturalists hold moral features to be causally inert.

To see why the inability to appeal to causal relations creates a problem for the non-naturalist, it is helpful to explicitly spell out an epistemological challenge that arises with respect to moral features on most non-naturalist accounts. Here is a slightly modified version of Enoch’s (2011: 159) formulation of the problem:

---

77 Harman formulates a challenge for moral properties that focuses only on the explanatory dispensability of moral facts and properties. Queerness claims have a broader scope than this particular challenge, as it, for instance, also covers other forms of indispensability arguments than the explanatory sort (cf. Enoch 2011: 54–56).

78 While Enoch agrees that the non-naturalist is not able to provide the best explanation in this particular case, he does think that overall, non-naturalism is able to provide the best metaethical or metanormative theory, despite losing out with respect to some particular challenges.

79 Oddie (2005) claims to hold a non-naturalist position that nonetheless secures causal efficacy. Wedgwood (2007) seems in many ways to qualify as a non-naturalist who believes normativity – in his sense – to be irreducible, but who nonetheless thinks that normative properties can be causally efficacious. He also thinks his view is compatible with a form of naturalism: “I shall argue that this metaphysical conception of the normative is entirely consonant with a broader version of naturalism – specifically, with the idea that normative facts both supervene on, and are realized in, purely natural facts” (2007: 135). Importantly, he does not engage certain metaphysical issues that are central for our purposes, such as “whether there is any reasonable interpretation on which Hamlet was right to say that ‘there is nothing either good or bad but thinking makes it so’ (Hamlet II. ii. 247–8)” (2007: 135 n. 1). Obviously, this is not compatible with what Enoch (2011) calls “robust realism”.

80 Enoch presents the challenge for normative truths generally. I have here restricted it to moral truths. As Enoch points out, there have been many attempts at specifying what the particular epistemological challenge which faces moral non-naturalism is, not all of them successful (2011: Chapter 7).
Very often, when we accept a moral judgement \( m \), it is indeed true that \( m \); and very often, when we do not accept a moral judgement \( m \) (or at least when we reject it), it is indeed false that \( m \). So there is a correlation between (what the realist takes to be) moral truths and our moral judgements.

The challenge, in short, is to explain this striking correspondence between our beliefs about morality and the moral truths themselves.¹ Why is this a challenge? To help bring this out, Enoch provides the following analogy:

Suppose that Josh has many beliefs about a distant village in Nepal. And suppose that very often his beliefs about the village are true. Indeed, a very high proportion of his beliefs about the village are true, and he believes many of the truths about this village. In other words, there is a striking correlation between Josh’s beliefs about that village and the truths about that village. (2011: 158)

In order to see what the challenge of explaining the correlation between moral truths and moral beliefs amounts to, it is helpful to consider how one would go about explaining Josh’s knowledge of the Nepali village. What is the explanation for the correlation between the facts about the village and Josh’s beliefs about it? That is, of course, obvious. So obvious, in fact, that we would not even think of framing it as a challenge to answer it. Assuming that we are not in the grip of superstition, we would simply expect there to be some *causal chain* which connects facts about the village, such as where it is located and who lives there, and Josh’s beliefs about these things.

For instance, maybe Josh read a richly illustrated guidebook about the village. It would thereby be quite easy to explain how Josh, say, knows what material the buildings in the village are made of. The buildings exist in physical space, and through some form of reproduction (e.g. mechanical or digital), representations of those physical structures have been made available to Josh. Even though there are many fascinating questions that can be asked about the processes involved, there does not seem to be any epistemological *mystery* here. In other words, the strong correlation between facts about the village and John’s beliefs about it has a straightforward, causal, explanation.

Consider now a situation where we did not manage to find *any* such causal explanation of the correlation in question. *That* would be striking, almost miraculous. It seems eminently implausible that such a correlation could obtain simply in virtue of a coincidence. In the moral case, therefore, there is a strong pressure on the non-naturalist to come up with some sort of explanation of this correlation. But, remember, the non-naturalist has given up the right to *any* causal explanation. We now begin to see why the non-naturalist is often thought to be in an awkward position, epistemologically speaking. In order to get out of this bind, the non-naturalist must come up with some non-causal explanation of this correlation, or leave it unexplained.

¹The presumed moral truths in question can be understood as the obtaining of moral favouring relations. So, the fact that there is a consideration which favours some particular agent acting in a particular way in a particular context is a moral truth, and when our judgements (or beliefs) about the obtaining of such relations match the facts, our judgements are true.
We are now in a position to say something a little more informative about how the notion of queerness should be understood. Based on the previous discussion, and one feature, (v), that will be discussed later, we now have some candidates as to what makes a feature queer. As a first approximation, I therefore suggest that a feature, \( F \), is queer if it:

(i) Possesses sheer queerness  
(ii) Is primitive  
(iii) Is epistemologically mysterious (e.g. requires unexplained correlations)  
(iv) Is not explanatorily integrated (with science or some other independent domain)  
(v) Is theoretically or practically dispensable

While, as we have seen, (i) and (ii) might make a feature seem *prima facie* queer, they are not sufficient for something’s being *metaphysically* queer. Rather, it seems to me that (iii), (iv) and (v) are the properties central to making a feature metaphysically queer.\(^{82}\) If a feature possesses all three of these, I submit, it will be metaphysically queer. I will not here take a stance of whether a certain configuration of two of (iii)–(v) will be enough to make a feature queer. Rather, I will see which of (i)–(v) applies to the specific feature in question, namely irreducibly normative favouring relations.

As it stands, such relations are likely to satisfy all of (i)–(iv), and below I will argue that they also satisfy (v). As such, irreducibly normative favouring relations seem metaphysically queer in a way that should make us ontologically suspicious of them. Faced with this charge the non-naturalist seems to have at least three lines of reply available.

First, the non-naturalist can double down; he can claim that, metaphysically queer or not, it is simply a brute fact that there are moral truths and the irreducibly normative favouring relations they entail. And because of this, it is wrongheaded to expect any further explanation. There really is not all that much an error theorist can say to this, and at this point the discussion is likely to end in a stalemate unless independent arguments can be brought in on either side. Fortunately, there is such an argument available to the error theorist, and I will introduce considerations that provide the error theorist with just such independent evidence below.

Second, the non-naturalist can try to deny that irreducibly normative favouring relations are queer, by denying that one or more of (i)–(v) applies to them. Denying sheer queerness would be unnecessary, because as we have seen, it is not that serious a charge in itself. As for (ii), primitiveness, it seems that the non-naturalist will most likely be inclined to embrace the primitiveness of irreducible normativity, since it

---

\(^{82}\) Number (iii), the epistemological mystery, is likely to obtain in virtue of (iv), lack of explanatory integration, which again is likely to obtain because of causal inertness.
seems difficult to give a reductive explanation of it which nonetheless allows it to maintain its characteristic normative force (cf. the Appendix).

When it comes to attempting to solve the epistemological challenge (iii), it seems that the non-naturalist must be satisfied with trying to cope with the challenge, rather than providing an ideally adequate response to it. Enoch (2011), for instance, admits that non-naturalism is likely to lose plausibility because of this challenge, but thinks that the superiority of non-naturalism on other fronts will make up for this. In order to cope with the challenge, he suggests appealing to a “pre-established harmony” (2011: 165–177). First, we must simply assume that there is some normative fact (presumably a metaphysically queer one), such as that survival (or some similar notion) in some sense is good (2011: 168).

This normative fact would then explain why there are moral truths, because such facts can be taken to be downstream consequences of the normative assumption: If survival is good, then, plausibly, harm is bad etc. Furthermore, Enoch thinks that the non-naturalist should further claim that the “aim” of evolution, i.e. survival or some such, incidentally aligns with this postulated normative fact. That is, the “aim” of evolution and the prior assumed normative fact are in harmony.

The solution, then, would consist of claiming that on the one hand there are moral truths (this follows from the brute assumption of a normative truth), and on the other, the “aim” of evolution (e.g. survival) aligns with these truths. Evolution could then causally explain the origin and shape of our moral judgements and beliefs, while the alignment of the “aim” of evolution and what is good would secure the strong correlation between our beliefs and judgements about morality and the moral truths themselves. While Enoch admits that the “fortunate coincidence” – the alignment of evolutionary aims with a postulated normative truth – still seems like an unexplained miracle, he claims that the “remaining miracle does not place a particularly heavy burden on [non-naturalist realism]” (Enoch 2011: 173).

Enoch’s solution seems to only push the proverbial bump in the rug further back, from a correlation between our beliefs and moral truths to a correlation between the aims of evolution and a normative fact. Enoch nonetheless thinks that the unexplained correlation between the aims of evolution and what is good is less of a problem for the non-naturalist than the original correlation (2011: 172–173). However that may be, it should be clear that the non-naturalist cannot provide an equally attractive response to this particular

---

83 Such a brute assumption might seem unmotivated, but Skarsaune (2011) brings up the following point, which may make the brute normative assumption somewhat more palatable: Either the assumption is true, in which case the postulation is acceptable, or it is false, in which case we would not have any explanation of the correlation, and non-naturalism would likely be committed to some form of epistemological moral scepticism. However, if such an intuitive normative claim is false, then scepticism might seem like the right way to go in any case. In short, it might seem that either way, the moral realist gets things right by making such an assumption.

84 I here adopt Enoch’s loose talk of evolution’s “aim” and so on, together with the qualification that they are to be understood as “shorthand for the usual respectable, non-teleological, evolutionary way of putting things” (Enoch 2011: 168 n. 38).

85 As Enoch (2011: 166–67) notes, there does not need to be a perfect harmony here; it would be enough to have a rough correspondence.
challenge compared to theories that do not have to postulate such unexplained correlations, such as error theory. However, as the choice between error theory and non-naturalist realism requires holistic evaluation, this does not refute non-naturalism, but merely subtracts some plausibility points from the view’s total score.

When it comes to the claim that moral features fail to provide any explanatory value (iv), this seems to follow from the non-naturalist’s commitment to the causal inertness of moral features. There are, however, interesting argumentative strategies open to the non-naturalist, where one can question the claim that moral features are not explanatory (e.g. Sayre-McCord 1988). I will not enter this debate, other than to submit my belief that it is unlikely that the non-naturalist will come out ahead on this issue either.

As for claiming that moral features are practically or theoretically indispensable in some way (v), I am not sure how one would go about arguing for such a claim. Enoch (2011: Chapter 3) argues that normative properties might be indispensable for deliberation, but specifies that this does not apply to moral properties in particular (2011: 81). Consequently, if one claims that moral features are queer without implicating that other normative properties are queer as well, then his indispensability argument does not have bite against such a moral error theory (Enoch 2011: 81; cf. Joyce 2014: 845).

In providing these short comments on the possibilities of rejecting the charge of queerness, I do not mean to conclusively settle that the non-naturalist cannot rebut one or more of these charges, but I hope to have shown that, collectively, they put heavy pressure on the position. The most likely scenario, it seems to me, is that the non-naturalist might come up with slightly less implausible solutions than what has been argued here. So, while this argumentative strategy is open to the non-naturalist, until a good rebuttal of two or more of (iii)–(v) is in place, moral features still appear metaphysically queer.

A third line of reply open to the non-naturalist is to argue that if the features in question are queer in the ways specified, then so are the features of certain other phenomena outside the moral domain. As a consequence, if the metaphysical queerness of moral features makes them ontologically suspicious, then this would also apply to the metaphysically queer features in other, non-moral domains. But, the argument goes, that charge would be implausible since the same feature in these other domain is commonly taken to be unproblematic by most of the practitioners within it. This appears to me as by far the most viable way of defending non-naturalism against an argument from queerness. I will pursue this argumentative strategy further in section 3.2 and 3.3 below, and in the chapters that follow.

86 For a discussion of what theoretical dispensability might be over and above explanatory power, see Colyvan (2015: §2, 2001). For an argument involving practical indispensability, see Enoch (2011: Chapter 3).
87 However, if the moral error theory in question turns out to generalize to the normative domain more generally, then it too will be targeted by Enoch’s argument. We will return to how far the argument for moral error theory we are now considering generalizes in the next chapter, and we will return to Enoch’s indispensability argument for normative properties in Chapter 6.
Debunking explanations

Discussions about whether a given feature really is queer to an extent that warrants ontological suspicion can quickly end up at an impasse. This is because both sides have the option of simply digging in their heels. “Yes, we should be suspicious!” says the error theorist; “No, we shouldn’t!” replies the non-naturalist. What is needed at this point is therefore to bring in independent evidence in favour of either view. As many error theorists have held, this can be done by employing so-called debunking explanations (e.g. Mackie 1977: Chapter 5; Joyce 2006: Chapter 6; Olson 2014: Chapter 7). The general idea is to first put forth the charge that some feature is metaphysically queer, and then in addition show that we can explain the origin and preservation of our practices and beliefs involving the queer feature without in any way assuming that the feature in question exists (or is instantiated etc.). The debunking explanation is, in other words, meant to make postulating the existence of the relevant feature redundant. Olson makes this explicit:

Step two of the argument from queerness is to offer explanations of our moral practices and beliefs that make no use of the assumption that there are moral properties and facts. … [Such an account] gives us a phenomenological and psychological explanation of why we tend to think and speak as if there are moral facts and properties when in fact there are none. (Olson 2014: 85)

In other words, the second part of the argument from queerness shows that not only are the features in question metaphysically queer, but also that they are explanatorily (and otherwise) redundant. In order to make this notion of ontological redundancy a little clearer, let us first introduce a general metaphysical principle that should be acceptable to naturalists and non-naturalist alike. The principle goes under many names, and has been given many formulations, but I will adopt a slightly modified version of the one found in Enoch (2011: 53), which he calls the minimal parsimony requirement.

THE MINIMAL PARSIMONY REQUIREMENT

Do not multiply ontological commitments without sufficient evidence.

This principle tells us that whoever wishes to introduce ontological commitments should provide sufficient evidence in favour of doing so. For instance, if the entities introduced by the non-naturalist explicitly conflict with positive postulations of our best science, this would constitute a strong defeating reason, as science is arguably our most reliable indicator of what exists. Not surprisingly, non-naturalists tend to avoid claims that have this explicitly conflicting feature. What is more common is therefore to claim that it is allowed to introduce additions to a scientific ontology, where these additions do not clash with any positive claims of science.

---

88 The principle is often called Occam’s razor, but “minimal” is meant to bring out that the principle does not automatically entail that explanatory indispensability is required for ontological commitment. Note also that this principle is to be understood as an institutional norm that governs certain practices, such as ontology (Olson 2014: 147–48).
However, the principle does not allow for the addition of any arbitrary but non-conflicting ontological commitment; the introduced entity must do some work or other. As we saw earlier, what qualifies as doing work can consist in providing explanatory power or being practically or theoretically indispensable. In order to decide whether an addition to our ontology has sufficient evidence backing it, one must look at the reasons for introducing the new entity, and at whether there are defeating reasons against it. Debunking explanations, together with the charge of queerness, are intended to provide just such a defeating reason.

At this point, one might ask why we need to bother with whether the features in question are queer, if we in any case have such a debunking explanation available. The answer is that if we are debating whether we want to allow a given feature, $F$, into our ontology, and we have an explanation of $F$ that does not assume that $F$ actually exist, this alone would not be sufficient evidence to deny $F$ a place in our ontology, in line with the minimal parsimony requirement. For instance, a proponent of introducing $F$ might argue that $F$ is practically or theoretically indispensable. On such grounds, the proponent could claim that including the feature in our ontology, despite the availability of a debunking explanation, would not constitute a multiplying of ontological commitments without sufficient evidence, since the feature is, in some sense, indispensable to some theoretical or practical enterprise. Indispensability, the argument would go, is sufficient evidence for introducing ontological commitments (cf. Enoch 2011: Chapter 3). For this reason, both queerness and a debunking explanation are needed.

So, what might a debunking explanation look like? First of all, we are after a causal account of how we came to have the moral practice and the beliefs and judgements that we in fact have. Furthermore, in order to play the required role in an argument from queerness, the account must make no assumption about these things actually existing. The general idea behind such accounts is usually to link either the origin of our capacity for moral judgements and/or the particular contents of such judgements with their evolutionary and social benefits. The evolutionary pressures that have selected or otherwise given rise to our (capacity for) moral judgements are then in turn explained purely by the workings of biological natural selection, which again can also be specified in purely causal terms. Such accounts therefore have no need to be in touch with a separate non-causal realm of moral truths in order to explain why we have the beliefs we have.

I will not dwell much on the details of such accounts, because as I shall explain below, they are not really all that important for the argument from queerness to go through. I will nonetheless give a brief overview of one such proposed debunking explanation, which is tentatively defended by Joyce (2006). Joyce

---

89 The non-naturalist could then appeal to a solution of the kind Enoch provided for the epistemological challenge, and claim that, even though the non-naturalist cannot provide the best explanation of that particular challenge, overall such a non-naturalist picture scores more plausibility points than its competitors.

90 Note that such accounts are also employed in the service of metaethical positions other than error theory, such as various form of non-cognitivism (cf. Olson 2014: 142).
takes the human capacity for moral judgments to be the result of evolutionary pressures. In a précis of his book *The Evolution of Morality*, Joyce gives a brief description of the central claim:

[H]uman morality is a distinct adaptation wrought by biological natural selection. … In claiming that human morality is innate, I do not mean that humans are innately social, or innately nice and friendly, or even that we innately have emotions that favor social cohesion; rather, I mean that humans have an innate tendency to make moral judgments. (2008a: 213)

The capacity to make moral judgments includes the distinctive phenomenological aspect which comes with such judgements, the “blunt conviction that certain actions just ‘must be done’ – a sense of what is ‘fitting’ that is not tied to the subject’s ends and is not ultimately contingent on the decrees of any authority figure” (Joyce 2008b: 246). What we are interested in, then, is not morality in the sense of acting pro-socially, but in terms of making prescriptive judgments, judgments about what is required and prohibited, about what should and should not be done etc. Furthermore, the specific content of the judgements is not necessarily innate: “I accept that mechanisms of cultural transmission play an enormous and perhaps exhaustive role in determining the content of an individual’s moral convictions” (Joyce 2006: 216). Of course, in order for a postulation of such innate tendencies or faculties for moral judgements to be plausible, there must be benefits involved for the agent or a group or both. Joyce thinks that

> self-directed moral thinking can advance an individual’s welfare by acting as a kind of psychological bulwark against various kinds of motivational infirmity, such as weakness of will or the discounting of future profits. Thinking that a certain action “simply must be done” may, in some circumstances, engage motivational structures more resolutely than even an awareness that the action is to one’s own advantage. … Second, … the general conspicuous costliness of moral conformity makes it well suited to function as an interpersonal commitment device. … [M]oral thinking can benefit the individual operating in the social sphere by foreclosing certain practical possibilities, thus bringing about desirable ends via altering interactants’ choices. (2008a: 215–216)

As for how this is carried out, Joyce subscribes to projectivism, and this licenses the explanation that

> what gives moral phenomenology its quality of “out-there-ness” – as if our moral evaluations are responses to a normative realm that precedes them – is the fact that we project aspects of our emotional lives onto our experience of the world. Projectivism sits comfortably with the empirical evidence indicating moral thinking to have both emotional and cognitive elements. (2008a: 216)

While admittedly speculative, Joyce’s account seems to satisfy his goal in this enterprise, which is not to establish the truth of his account, but to fulfil “the more modest ambition of sketching out a clear, coherent, productive, plausible, and testable hypothesis” (Joyce 2008b: 245). Since this and similar accounts rely only on evolutionary, social and cultural factors, there is nothing in such an account that licenses the thought that

---

91 For accounts of how the content of moral judgements might be shaped, see e.g. Nichols (2004); Sripada and Stich (2005).

92 For further discussion of how projectivism can be a natural companion for moral error theory, see Olson (2014:4–8).
the formation of either the capacity itself or the particular contents of our judgements tracks some objective, mind-independent realm of truths.

There are many such accounts on the table, some of which also try to fill out how different cultural and psychological pressures might shape the particular contents of our moral judgements (e.g. Nichols 2004). However, in this dialectical context, it does not matter much which particular approach turns out to be empirically preferable. The reason is that there does not seem to be any principled objection to such accounts, and the problem they present to the non-naturalist remains regardless of what the specific empirical details are. The main point is that the non-naturalist will remain unlikely to claim that our moral judgements and beliefs are not shaped by, say, evolutionary and cultural pressures. As Enoch points out with respect to a different debunking explanation:

Replace any other (non-tracking) causal explanation of why we make the normative judgements that we do in fact make, and the [non-naturalist] will find herself up against the problem. (Enoch 2011: 164)

A non-tracking account of our moral beliefs is any account which explains our beliefs without reference to whether those beliefs are true or not. That is, the account does not take our beliefs to have been shaped by the fact that they track some independent set of truths. As long as an account is non-tracking and causal, it will present trouble for the non-naturalist. At this point, it might be wondered whether such a non-tracking account cannot be presented for any number of other phenomena, and if it cannot therefore be held to overgeneralize. In most cases, however, there seems to be a clear difference, namely that it seems very difficult to come up with a debunking account of, say, our belief in external objects, as such an account would have to make no reference to mind-independent truths.93

The reason a debunking explanation of moral beliefs seem reasonable in comparison, is that such as a belief might be just as useful even if they are false. The primary function of such beliefs can be understood as the effective regulation of individual and social behaviour (e.g. Mackie 1977: 107–105; Joyce 2006: 108–123). In coordinating social interaction, false beliefs – if shared by all or almost all – can be just as effective as true beliefs would be. One way to secure such intra- and interpersonal coordination of social action would be therefore by inculcating a belief in the practical authority of moral reasons (cf. Joyce 2006: 108–123).

Debunking explanations, and their general plausibility – at least in structure, if not in empirical detail – threatens to make the postulation of a separate realm of moral truths redundant. This redundancy would

---

93 It has recently been claimed that the error theorist cannot use such debunking accounts against our moral beliefs without also employing them against our mathematical beliefs as well (Clark-Doane 2012, cf. 2014). If it turned out that the moral error theorist could not run her argument against moral features without expanding her target to also include mathematical features and knowledge, she would of course have to choose between either so expanding or retracting the argument against moral features. I will not pursue this issue here, other than to note that although being forced to also embrace an error theory about numbers would make moral error theory more costly, it would not necessarily make the theory unacceptably implausible. In what follows, I will operate on the assumption that moral error theory is not necessarily committed to an error theory about numbers; but if it turns out that it is, we might have to deduct some plausibility points from the theory, though not, I think, a critical amount.
obtain in virtue of the fact that such causal accounts give us an answer to the question of how we could come to have the practices and beliefs we have, entirely independently of their truth. Furthermore, such accounts bring out the reason why we feel so strongly about questions of morality, and even why one might feel a very strong aversion to the view that there are no moral truths. If moral features in addition are metaphysically queer, in that they fulfil (most of) (i) through (v), it starts to look like the error theorist has a very strong case against the non-naturalist.

3.1.1.3 The plausibility claim
We have now seen both how to charge moral features, such as irreducible normativity, with being metaphysically queer, and how to give an account of our moral judgements and beliefs that does not assume the existence of moral truths or features. What remains is to put these two – the queerness argument and the debunking explanation – together. What we get is that the moral features in question are metaphysically queer to the point of being ontologically suspicious and that we would lose nothing of theoretical value by rejecting the existence of moral truths as the non-naturalist understands them. Taken together, it certainly seems more plausible to reject the existence of moral truths and features than to accept them.

However, maybe it is possible for the non-naturalist to try another strategy to justify the postulation of moral truths. This alternative strategy relies on a point I mentioned earlier, namely that theory choice relies on a holistic evaluation of all relevant evidence. The big question is therefore, which theory is the more plausible overall – moral error theory or non-naturalist realism? One way to try to tip the scales in favour of non-naturalism is to take our ordinary convictions into account. If the abandonment of (non-naturalist) moral realism would be sufficiently implausible, this could conceivably be thought to outweigh any of the theoretical benefits that moral error theory manages to score in the ways outlined above.

As Olson (2014) helpfully notes, A.C. Ewing, for instance, signals that our confidence in basic moral beliefs appear to be epistemological bedrock, when he rhetorically asks in response to the possibility of there being no moral truths: “Do I not know that it would be wrong of me to go into the street and torture the first person I meet even if I happen to be so constituted that I should enjoy watching his suffering?” (Ewing 1947: 30, cited in Olson 2014: 139). Ewing further held that it is not possible to “seriously believe for an hour in the emergencies of daily life” in a theory that makes such a claim (1947: 32, cited in Olson 2014: 139). He took this as a severe theoretical flaw, and thinks that, as a methodological restriction on philosophical theories, we should “demand of a philosophy that it should be in accord with what we cannot possibly help believing in ordinary life” (Ewing 1947: 32, cited in Olson 2014: 140).

This methodological demand is reminiscent of one David Lewis has made in more recent times, which he terms “a simple maxim of honesty”: “[N]ever put forward a philosophical theory that you yourself cannot believe in your least philosophical and most commonsensical moments” (1986: 135). The question of whether any given proponent of a given theory could believe it even in her least philosophical moments
seems difficult to answer with certainty (at least without checking). In any case, I do not always think it makes sense to hold this maxim to be applicable to a philosophical view, or, for that matter, a scientific theory. For instance, when it comes to moral error theory, it seems odd to subject it to this demand, since the theory itself predicts there to be deep, psychological inclinations and mechanisms in us that incline us towards making and subscribing to moral judgements. As such, I don’t think it can be taken to be a knockdown argument against a view that otherwise has sufficient evidence backing it up, that it is difficult to believe it in one’s least philosophical moments.

For instance, I do not think it should be held against a theory that holds the universe to be a simulation, or which claims that we are brains in a vat, that we have difficulty believing them in our least philosophical moments. More convincingly, even if a visual scientist was caught off guard and thought that the two lines in a Müller-Lyer illusion was different lengths, that would not disprove the underlying theory claiming them to be the same length. In other words, in cases where there is a plausible, even empirical, account of why the proponent of the view could not believe the theory in question (in his or her least philosophical moments), I think we can safely disregard this maxim.

Even if the above reasoning is correct, there is another way to argue against moral error theory on the basis of our commonsensical belief in moral truths. Enoch (2011: 117–121) pursues such an argumentative strategy against moral error theory, and argues against it by employing moral claims as premises. At first, this might be thought to beg the question against the proponent of error theory, as the non-naturalist would then employ a moral premise in an argument against a view that denies the existence of true moral claims. However, in order to see how such an argumentative strategy can still be legitimate, let us take a closer look at one of Enoch’s arguments, which he calls the Simple Moorean Argument:

The infliction of horrible pain on random victims is morally wrong; therefore, it’s true that the infliction of horrible pain on random victims is morally wrong; therefore, some simple moral judgements are true; therefore, a metaethical error theory is false. (Enoch 2011: 117)

This can helpfully be construed as follows:

**THE SIMPLE MOOREAN ARGUMENT**

**Moorean premise**

It is a fact that inflicting horrible pain on random victims is morally wrong.

**Moorean conclusion**

Thus, there is at least one moral fact.

**Conclusion**

Thus, moral error theory is false.
Such Moorean arguments are, of course, not knockdown arguments. However, I do think they provide a real challenge to the error theorist. What such arguments help bring out is that when considering the plausibility of moral non-naturalist realism vis-à-vis error theory we must consider which view we have the most confidence in, as expressed by the parts of the simple Moorean argument:

\[\text{Are you more confident in its premise (The infliction of horrible pain on random victims is morally wrong) or in the denial of the conclusion (error theory)? If, like me, you are more confident in the former, it seems like you are entitled to conclude to the denial of error theory. (Enoch 2011: 119)}\]

It seems fair enough that, in general, if we are to give up the belief that anything is morally wrong, we should be required to be more confident in the reason for abandoning this belief, than in the belief that is being abandoned. This is why it is a challenge to moral error theory. However, it should be obvious how this challenge can be answered. After all, we have just looked at the general structure of accounts that explain how the belief that makes up the Moorean premise has come about, and why we believe it so strongly and, consequently, are rather reluctant to give it up. When discussing how we should evaluate the simple Moorean argument, Enoch says that

\[\text{a presumption in favor of common sense … is all the Moorean methodology comes to, at least as I practice it. Questions may then arise, of course, about the strength of the presumption. I will want to insist that the strength may vary from context to context, and that at times it may be quite close to maximal. (Enoch 2014: 863, my emphasis)}\]

If Enoch means that the strength (i.e. the presumption in favour of) the Moorean premise in the moral case should be “quite close to maximal”, it would mean that he thinks that we should take our belief in the Moorean premise as almost certain simply because it is commonsensical, and despite very powerful considerations to the contrary. This, then, is another instance of the proponent of non-naturalism simply digging in his heels. While that might sometimes be legitimate, a claim to the effect that we should accept our pre-theoretical and commonsensical opinions at more or less any cost seems less than convincing, to put it mildly. I am not sure what more one can say to the non-naturalist who claims that almost nothing – not even the presence of a queerness argument and a debunking explanation – would change her faith in commonsensical opinion. As Joyce comments in a slightly different context, such non-naturalists might be thought to

\[\text{have about them an air of slightly desperate conservatism: an anxious determination to ensure that popular belief systems turn out as true. (2008b: 265)}\]

While one can very well admit to a presumption in favour of commonsensical opinion, it seems difficult to defend anything more than a defeasible presumption. The details of such a presumption is the centre of much

---

94 See Miyata (2015) for a detailed look at Moorean arguments in metaethics, although primarily as directed against subjectivist theories.

95 Joyce (2014) seems, for various reasons, to think that there is something fishy about such arguments. I do not share his scepticism on this score.
debate in many subdisciplines of philosophy, and is not something I will venture into here, other than assent to Lewis’s position on the matter:

Common sense has no absolute authority in philosophy. It’s not that the folk know in their blood what the highfalutin’ philosophers may forget. And it’s not that common sense speaks with the voice of some infallible faculty of ‘intuition’. It’s just that theoretical conservatism is the only sensible policy for theorists of limited powers, who are duly modest about what they could accomplish after a fresh start. Part of this conservatism is reluctance to accept theories that fly in the face of common sense. But it’s a matter of balance and judgement. Some common sense opinions are firmer than others, so the cost of denying common sense opinion differs from one case to the next. And the costs must be set against the gains. (Lewis 1986: 134)

I take it that whatever one makes of the Simple Moorean Argument, the availability of an argument from queerness that I have presented here, which includes a debunking explanation, shows that we have a highly plausible alternative account of why we believe the Moorean premise so strongly. The availability of such a plausible alternative should itself lead to our confidence in the Moorean premise being severely weakened. Weakened enough, in any case, for the Simple Moorean Argument to not have much force at all against the error theorist. As Enoch himself points out:

At the end of the day, you have to ask yourself what it is that seems most plausible to you, given the entirety of evidence and arguments available to you. At the end of the day, then, this will be the test for the error theory as well. (Enoch 2011: 121, my emphasis)

To me, at least, and given the previous considerations, it seems like moral error theory comes out on top in terms of plausibility points. However, my goal in this thesis is not to show that moral error theory is obviously true, but only that it seems to score, on the whole, slightly more plausibility points than its main competitor. I will therefore leave these matters here. I hope to have shown that there is a strong and quite plausible case both for claiming that moral discourse has a core commitment to irreducibly normative favouring relations and that such relations seem to be objectionably queer, in the sense that they are both metaphysically queer and theoretically redundant. We might therefore be tempted to conclude with Mackie:

Considerations of [queerness] suggest that it is in the end less paradoxical to reject than to retain the common-sense belief in the objectivity of values, provided that we can explain how this belief, if it is false, has become established and is so resistant to criticism. (Mackie 1977: 42)

In any case, it should be admitted that it is not obvious that the non-naturalist scores the most plausibility points, and that even if she does, she is not likely to win by much. Enoch makes this clear repeatedly by claiming that error theory is the strongest competitor to his own preferred non-naturalism (e.g. 2011: 121), and that the race is pretty close. Given the plausibility of the view, it seems very much worth looking into where moral error theory, if true, would take us. This exploration will be the task for the remaining chapters.
4 Generalization I: Practical reasons

[The arguments for a moral error theory have not threatened hypothetical imperatives.]

Joyce (2001: 177)

4.1 Do the arguments for moral error theory generalize?
In the two preceding chapters, I have presented what I take to be the strongest and most viable version of moral error theory and the argument from queerness that supports it. I have also tentatively argued that such an error theory seems plausible on its own, and that it in addition succeeds in scoring more plausibility points than what I have argued is its main competitor non-naturalism. This concludes my discussion of moral error theory as such.

The time has now come to pick up on Mackie’s exploratory comments about the potential reach of the arguments employed on behalf of moral error theory. He (1977: 15) pointed out that there seems to be strong similarities between moral value and value of other kinds. Mackie’s own contention was that his arguments against moral value would unproblematically generalize to target other domains, and successfully so. The question we now face is therefore: how far-reaching are the theoretical consequences of the arguments for moral error theory?

In exploring this question, it is worth having in mind that the answer will depend, at least to some extent, on the specific version of moral error theory under consideration. As we saw in Chapter 2, many error theorists have claimed to have categorical reasons as their target, and that such reasons are (mostly) confined to the moral domain (cf. Joyce 2015: §4). Despite such oft-repeated claims, I have argued, and will further support in this and the next chapters, that their target is better understood as irreducible normativity. My contention is therefore that error theorists who have focused on categorical reasons or some similar notion should instead understand their target to be irreducible normativity, since this notion better captures the thing that such theorists take themselves to be attempting to capture – the inescapability and normative force of moral reasons. For this reason, it is not only the specific argument on behalf of moral error theory that I have presented here that we should see this investigation as concerning, but rather all error theoretic arguments that target similar notions, such as categorical moral reasons or objective prescriptivity.

---

96 See Olson (2014: Chapter 5) for some arguments for moral error theory that generalize differently than the one I consider here.
97 Joyce (2001) argues that there are no categorical practical reasons either.
98 As I mentioned in note 45 in Chapter 1, some theorists claim that there are categorical moral reasons in the sense that all agents, in virtue of being rational agent, share a set of ends which give rise to moral reasons. While I will not argue for it here, I do believe the arguments of this chapter show that such reasons too would be committed to irreducible normativity. In this way, error theorists are safe to opt for irreducibly normativity instead of categorical reasons.
When it comes to investigating to what extent the arguments for moral error theory generalizes, it helps to have the spelled-out version of the argument from queerness, and the resulting substantial claim, as it was argued for in Chapter 3.

AN ARGUMENT FROM QUEERNESS

Queerness claim
Irreducibly normative favouring relations are metaphysically queer.

Debunking explanation
There are highly plausible accounts of why we are inclined to believe that there are irreducibly normative favouring relations, accounts that in no way depend on such relations actually obtaining.

Plausibility
Given all available evidence, it seems more plausible to reject the existence of irreducibly normative favouring relations than to accept them in our ontology.

Conclusion (substantial claim)
There are no irreducibly normative favouring relations.

This argument, it should be clear, does not limit itself to the moral domain. So far, we have only looked at irreducibly normative favouring relations as a feature of moral reasons, but there does not seem to be anything specifically moral about this feature. The trouble with such favouring relations is rather their irreducible normativity. Because of this, it seems straightforward that this argument for moral error theory generalizes beyond the moral domain and to all irreducibly normative entities; irreducible normativity will be equally objectionable independently of which domain it appears in. One might, however, wonder whether the debunking explanation will successfully generalize to other normative domains than the moral. I will return to this question in Chapter 6.

That moral error theory generalizes this broadly should be something of a surprise. For instance, one of the foremost defenders of the view in the past 15 years seems somewhat puzzled about why an error theorist would move from moral error theory to a metanormative\textsuperscript{99} error theory:

I don’t think that many metaethical theories deny the existence of normative truths. Not even the typical error theorist denies them. Certainly it is possible to be an error theorist about normativity across the board, but I don’t know of anyone who endorses the view. … [P]eople like Mackie (1977) and myself (2001) … think that moral normativity is a defective concept, but happily allow that other normative frameworks are trouble-free. (Joyce 2014: 844)\textsuperscript{100}

\textsuperscript{99} This view also goes under the name normative error theory.

\textsuperscript{100} From what I can gather from the context, Joyce has irreducible normativity in mind in this quote. That no one is a metanormative error theorist about all of normativity, where “normativity” is taken to include also reducible forms (e.g. something being normative in a rule-implying or norm-relative sense) is not very surprising – as we have seen,
As far as I know, the list of error theorists about irreducible normativity is indeed rather short: Olson (2014) is explicitly committed to it and Streumer (forthcoming) also defends the view.

However, if what I have argued in the previous chapters is correct, Joyce should himself become a metanormative error theorist, since the favouring relations framework we have looked at seems to be a substantial improvement upon his framework of strong and weak categorical reasons. Furthermore, there does not seem to be any good reason to block the generalization from the moral to the irreducibly normative. The only reason I can think of is in order to block the view from generalizing. That might admittedly seem tempting, since allowing error theory to generalize to other domains is likely to multiply the possible objections to it tenfold. For instance, many of the challenges we have looked at in the previous chapter where the moral error theorist arguably comes out ahead of the non-naturalist, might play out very differently if the non-naturalist’s competitor instead takes the form of a metanormative error theorist. I am therefore not at all sure whether it is a good thing for the error theorist that the view so generalizes. 101 Nonetheless, I can see no way of blocking the spread of error theory from the moral domain to other normative domains that is not simply ad hoc.

Besides, even if blocking the spread of error theory was found to be acceptable, there is still the fact, as I will show in this and the next chapter, that non-moral normative domains have peculiarities similar to those within the moral domain. That the framework of irreducibly normative favouring relations can account for this in a unified way seems like another, independent reason to prefer the favouring relations framework. As Mackie himself says when discussing different problematic normative domains, “there would be at least some initial implausibility in a view that gave the one a different status from the other” (1977: 15). 102

What seems to follow from a sufficiently strong and well-motivated argument for moral error theory is, therefore, a global, metanormative error theory. (At least when the “normative” part of “metanormative” is understood as “irreducibly normative”, which is how I will understand it in what follows. 103 ) In short, moral error theory takes us directly to an error theory about irreducible normativity.

The literature which discusses a form of error theory that generalizes to target the irreducibly normative more generally is largely unified in being deeply sceptical of its success, if not its internal coherence (see e.g. Cuneo 2007: Chapter 4; Stratton-Lake 2002: xxv–xxvi; Parfit 2011b: 293, 522, 619; Enoch 2011: Chapter 3). The few who do not see the prospects for a metanormative error theory to be as dim are often unsatisfyingly brief and circumscribed in their treatment of the view (e.g. Bedke 2010; Olson 2014; such notions do not seem problematic. Note also that Mackie, as we have seen, does in fact have a bone to pick with at least some other normative domains than the moral (e.g. aesthetic value; Mackie 1977: 15). 102 101 We will look at the plausibility of a metanormative error theory in Chapter 6. 103 Mackie is here talking about moral value and aesthetic value, but the point holds for other similar normative domains as well. 101 I will, however, continue to distinguish between reducible and irreducible normativity when “normative” is not used in this construction.)
If you are going to champion a theory as radical as metanormative error theory, you had better know perfectly well what you are giving up, and, perhaps equally importantly, what will remain. Let us therefore begin to investigate what specific entities such a generalized error theory would target besides moral reasons.

We have seen that the favouring relations framework seems to capture the conceptual structure and semantic commitments of not just moral reasons, but also of other kinds, such as institutional reasons. We also saw that not all forms of reason talk are committed to irreducibly normative favouring relations. In other words, what metaphysical commitments any particular type of reasons gives rise to must be investigated on a case-by-case basis. One way of investigating the scope of a metanormative error theory is therefore to look closer at the conceptual structure and semantic commitments of other forms of reasons in order to see if they entail irreducibly normative favouring relations.

We have already looked at the distinction between categorical and hypothetical reasons, and we have found that some categorical reasons entail irreducibly normative favouring relations (e.g. moral reasons), while some do not (e.g. reasons stemming from the rules of etiquette). What about hypothetical reasons? These reasons have often been held to be metaphysically unproblematic by error theorists, and most other philosophers. One would therefore not expect them to be a potential target for even a metanormative error theory. However, as we glimpsed in the previous chapter, they do seem to possess strong normative force. Furthermore, there have been brief treatments indicating that at least some hypothetical reasons might involve irreducible favouring relations (e.g. Olson 2014: 152–155; Bedke 2010: 48–51). It seems worthwhile, therefore, to expand on these treatments and provide an in-depth look at the commitments of the reasons of practical rationality, with a particular focus on the type of normativity they are committed to. While my main focus will be the nature of hypothetical reasons, I think it will be helpful to situate such reasons within the landscape of practical reason more generally.

### 4.2 The reasons of practical rationality

In Chapter 2, we applied the framework of favouring relations to both moral and institutional reasons. These, however, are not the only types of reasons that this framework is used to account for. Such relations are taken by many to be a feature of normative reasons generally:

> It is widely agreed that when someone has a reason to perform an action there is something favouring that action. It also widely agreed that whatever so favours the action is thereby, in some sense, 'reason giving'. Beyond that, agreement stops. (Lillehammer 2010: 17; cf. Scanlon 1998: 19)

---

104 Streumer has a forthcoming book, *Unbelievable Errors: An Error Theory about All Normative Judgements*, which might somewhat remedy the situation.
In investigating the broader class of reasons, a natural place to start is to take a closer look at the reasons that are taken to spring from practical rationality, which may be loosely defined as “the general human capacity for resolving, through reflection, the question of what one is to do” (Wallace 2014).

One simple answer to the question “What am I to do?” is “What you should do!” Of course, the question then morphs into “What should I do?” A plausible answer here is “You should do that which you have (good or most) reason to do”. In short, then, the question of what we are to do can naturally be seen as a question of what we have reason to do. So, what do we in fact have reason to do? As we have already seen, in many situations we usually take moral reasons to have a claim on what we ought to do, but having already looked at such moral reasons in detail, I will in this chapter confine myself to an investigation of non-moral practical reasons. What other practical reasons are there? Perhaps unsurprisingly, there are countless accounts that attempt to answer this, and as with moral reasons there is much disagreement over how to understand the nature of practical reasons more generally.105

One central distinction when it comes to accounts of practical rationality and reasons is that between instrumental and non-instrumental accounts. This distinction separates those accounts which hold our practical reasons to stem simply from our ends together with means–ends relations, and those that hold that not all (or no) practical reasons are of this form.106 In this chapter, I will focus on instrumentalism about practical rationality and reasons, which is the “default view in the field, and probably among philosophers in general” (Millgram 2001: 4). Instrumentalism is an umbrella account that has many sub-varieties, some of which we will encounter below.

Since the issues with which we will be concerned are about the fundamental nature of reasons, it does not matter much which form of practical instrumentalism we look at. I will therefore illustrate my points in this chapter with rather crude versions of practical instrumentalism. That these are not very plausible versions need not detain us, as I will argue in section 4.5 that all the points I make with regards to such simple forms of instrumentalism also holds for more advanced versions of the view, and even Williamsian varieties that connect hypothetical reasons closely with optimally rational deliberation.

According to a simple form of practical instrumentalism, the only reasons for action that practical rationality gives rise to are hypothetical, in that they depend on the ends of the agent together with means-end reasoning.108 The only source of practical reasons, in short, is the following principle.

---

105 For an overview of current views of practical reasoning, see Millgram (2001).
106 This distinction closely parallels the debate between internalist and externalist views about reasons, a dispute that is not necessary to get into for our purposes here. For an overview, see Finlay and Schroeder (2012). The internalist will be a practical instrumentalist, while the externalist will not.
107 The names Humeanism or Humean instrumentalism are often used as equivalent with instrumentalism (as in some of the quotes in this chapter). However, as the question of whether Hume was Humean in this sense is contested, I will avoid this use (cf. Millgram 1995).
108 In fact, some theories of morality, such as contractarianism, holds moral reasons to be just such hypothetical reasons, in that, despite appearances to the contrary, there is never a conflict between what is morally required of us
An agent, $A$, has a reason to $\Phi$ iff $A$ has an end, $E$, which $\Phi$ing will satisfy.

Let me note two complications with this principle. Note first that instrumentalist accounts do not really say whether you should $\Phi$ or abandon $E$ (cf. Wallace 2014: §4). In the following, I will omit the qualification that one can always rationally choose to eliminate the desire, instead of pursuing the end. Second, note that there is a difference between saying that an agent has a reason to do that which the agent believes to be a way of satisfying $E$ (which might frustrate it instead), and that the agent has a reason to do that which in fact would satisfy $E$. I will not go into how one can connect this objective/subjective distinction, but simply deal with the principle as stated, which is in terms of what might be called objective reasons (cf. Joyce 2001: 53).

Instrumentalism, at least in its basic forms, is furthermore committed to the view that “the choice of means to our ends can be more or less rational, but our choice of ends themselves can’t” (Fehige 2001: 49). On such an instrumentalist view, an agent is rational to the extent that she is guided by the practical reasons the instrumental principle begets (cf. Joyce 2001: 54). Practical instrumentalism has long been popular, though it has certainly also received sustained criticism. However, what is most interesting for our purpose is that it, and the practical reasons it gives rise to, seem minimal as far as metaphysical commitments go. In fact, it has seemed to many, including Foot (1972) and Mackie (1977), to be the only way of capturing “real reasons” without subscribing to metaphysically costly notions such as categorical reasons. That this notion of rationality and reasons seems like a minimal commitment is also endorsed by R. Jay Wallace:

Instrumental rationality, in its most basic form, instructs agents to take those means that are necessary in relation to their given ends. In the modern era, this form of rationality has widely been viewed as the single unproblematic requirement of practical reason. (2014: §4)

On the other hand, as we saw in the Chapter 2, hypothetical reasons nonetheless seem to possess strong normative force – they have substantial deliberative weight. In this sense, they are like moral reasons and unlike, say, institutional reasons. Furthermore, non-instrumentalist theories of practical rationality usually

---

and what this form of instrumental practical reasons provides us with (e.g. Gauthier 1986). I will not spend time discussing such views, as I argued in Chapter 2 that moral reasons do not seem to fit the profile of such hypothetical reasons (see also the Appendix).

109 See Joyce (2001: 53–) for one way to connect these two types of reasons. Again, my arguments in this chapter will apply indecently of how, and whether, one chooses to do this.

110 For some discussion of practical instrumentalism and criticisms of it, see the essays in Millgram (2001). For a selected list of proponents of the view, see Fehige (2001: 70 n.1).

111 Although I have previously argued and will continue to argue that the categorical-hypothetical distinction is not very helpful in separating metaphysically troublesome from trouble-free reasons, I will nonetheless employ it at certain points in this chapter and the next, in order to preserve terminological continuity with the literature on the topic.
add other elements to it, and are often themselves committed to the instrumentalist principle in some qualified form or other.

The instrumentalist picture, then, is one way to think about practical rationality and reasons, and one which seems to make minimal metaphysical commitments. Another is championed by the non-instrumentalist, who holds that other things than an agent’s ends can determine what we have reason to do. Lillehammer (2003) names one such type of view substantivism, which holds that what grounds our practical reasons are often simply the given action or end itself, and that individual ends, therefore, can be subject to rational scrutiny. Such a view of ends is familiar from the moral case, where the claim that we have a reason not to torture animals for fun is true in virtue of the type of action it is. Similarly, it could be argued that we have a reason to avoid self-harm or overeating simply because these things are, in some sense, bad for us. This badness does not need to stem from what one wants or desires, and even agents who take up these things as their ends would not, on this view, have a practical reason to carry them out. These agents would have such a reason, however, according to the proponent of practical instrumentalism.

I will not discuss such substantivist and similar views, but I think it should be clear from the arguments of this chapter that it applies with equal, if not stronger force to such views. Our focus, therefore, will be on the nature of hypothetical reasons, with a special focus on what kind of favouring relations they involve, if any.112

It has been customary both in metaethics and, more generally, in the theory of action, to hold that while categorical reasons are metaphysically suspicious, hypothetical reasons are ontologically respectable through and through. We have already looked at Foot (1972), who expressed just this view. As we saw, she held that in order to save morality from being committed to what she thought of as a philosopher’s fantasy – categorical reasons – we should discard the notion of categorically reason-giving requirements and imperatives altogether. Instead, we should accustom ourselves to thinking of reasons for action, including moral reasons, as hypothetical reasons.

Among the many philosophers who have subscribed to this train of thought, we find most error theorists (e.g. Mackie 1977: 75; Joyce 2001).113 Let us therefore take a closer look at the structure of this type of reasons to see whether the confidence in their metaphysical acceptability is borne out.

In order to get a grip on the structure of the hypothetical reasons involved in practical rationality, it is instructive to take a look at how they are talked about by those who hold them to be less metaphysically mysterious than their categorical counterparts. Mackie, for instance, uses a number of phrases when

---

112 I think the favouring relations framework, and my discussion of hypothetical reasons in this chapter, can be quite easily adapted to also target such views.
113 See Olson (2014: 152–155) for an exception.
describing hypothetical reasons, and these are typical of the discussion around hypothetical reasons in general.\textsuperscript{114}

‘If you want $X$, do $Y$’ (or ‘You ought to do $Y$’) will be a hypothetical imperative if it is based on the supposed fact that $Y$ is, in the circumstances, the only (or the best) available means to $X$, that is, on a causal relation between $Y$ and $X$. (Mackie 1977: 27–8)

The reason for doing $Y$ lies in its causal connection with the desired end, $X$; the oughtness is contingent upon the desire. (1977: 27)

The reason for doing $Y$ is contingent upon the desire for $X$ by way of $Y$’s being a means to $X$. (1977: 29)

The general idea behind hypothetical reasons seems clear enough. By desiring something, and recognizing that some behaviour is a way of efficiently attaining that something, one would seem to thereby have a (normative) reason to act accordingly. This, after all, is what being practically rational is all about according to the practical instrumentalist. Nonetheless, one question that quickly arises is how such talk of reasons being “created”, “lying in” and being “contingent upon” is supposed to be cashed out. One way it cannot be cashed out, if it is to remain a metaphysically unproblematic, is in terms of irreducible normativity. So, there must be some other way of accounting for the structure of hypothetical reasons than an appeal to irreducible normativity.

In order to look at such reasons in more detail, let us focus on one of the simplest varieties of practical instrumentalism. Such a view holds that an agent has a reason to do something if and only if doing so would satisfy some present desire that agent has. On this account, desiring to drink and seeing a glass of clean water on the table in front of you gives you a reason to take a sip. Such hypothetical reasons are therefore prescriptive, in that they tell us what we should do. And since one can always, at least in principle, abandon the desire, it is clear that they are not inescapable – such reasons are evaporating.

As we have seen earlier, prescriptivity and evaporability are not all that interesting in themselves, and the most important goal of our investigation of hypothetical reasons is to see what provides hypothetical reasons with their strong normative force. When we think of this type of reasons – the ones that stem from practical rationality – they seem to be paradigmatic of reasons having strong normative force (Joyce 2001). What could provide more deliberative weight than reasons stemming from what you want to do?

In order to see where such reasons get their normative force, consider the question of what an instrumentalist account of normative reasons is committed to. First of all, it seems undeniably committed to a large number of particular hypothetical imperatives and reasons, as in all the particular (i.e. filled-in) instances of “If you desire $D$, and $\Phi$ing will satisfy $D$, you have a reason to $\Phi$”. This observation, however, can quickly lead to the question of whether an instrumentalist account of practical rationality is not also committed to a general and even stronger imperative.

\textsuperscript{114} The quotes are helpfully collected in Olson (2014: 152).
4.2.1 Do hypothetical reasons imply categorical reasons?

Instrumentalists seek to populate practical rationality with only hypothetical reasons, and to do so without in any way relying on categorical reasons. However, some philosophers have accused instrumentalist accounts of simply moving the categorical reasons out of sight, rather than managing without them (Korsgaard 1996; Hampton 1998: 125–206; cf. Lillehammer 2003, 2007: Chapter 3).

The reasoning behind this is as follows. If one accepts all the particular hypothetical imperatives that a practical instrumentalist is committed to as valid, one would seemingly in addition have to accept at least one categorical norm of practical reason, namely that an agent ought to promote the means to whatever ends that agent presently has. In short, one would seem to be committed to a categorical imperative of the following form: “Take the means to satisfying your present desires!” As we have seen, (strong) categorical imperatives give rise to categorical reasons, and the instrumentalist would thereby be committed to categorical reasons. To see this, note that the reason that an agent has to take the means to his present ends obtains independently of the particular ends of any particular agent.

This is certainly one way to answer the question of how hypothetical reasons get their strong normative force – by springing from a categorical requirement. If the argument is successful, the consequence is that since we had better accept hypothetical reasons, we should therefore also allow for the existence of categorical reasons. Thus, instrumentalism about practical reason would be false. Opting for the second horn of this dilemma would mean to give up on not only categorical reasons, but hypothetical ones too, and it will thus equally lead to the falsity of practical instrumentalism.

4.2.1.1 Joyce’s answer

Joyce (2001: 115–123) has responded to this challenge by attempting to clarify the structure of imperatives in a way that allows the instrumentalist to avoid commitment to categorical reasons. He claims that there are two ways to interpret reasons claims, and uses the following example to make his point (Joyce 2001: 121):

Take what seems to be a straightforward [example of a conditional reason]: one’s reason to save a drowning child if one exists. There are two readings:

(i) If there exists a drowning child, then S has a reason to save him/her.
(ii) S has a reason to save a drowning child if one exists.

In (i), the claim made is that S has a reason to save the drowning child only when such a situation as specified in the antecedent in fact obtains. In (ii), on the other hand, it is being claimed that S always has a certain conditional reason, namely a reason to save a drowning child if one exists. Joyce finds (i) the more intuitive reading by far:

115 Joyce frames this in terms of a discussion of internal vs. external reasons, but, again, it is not necessary to introduce that dimension here.
It is far more natural to say in general S has no reason to save any drowning children, but when there exists a child drowning, then (and only then) does S get a reason to save him/her. The consequence of disagreeing with this is to lumber all of us with all sorts of bizarre reasons: you, right now as you read this, would have a reason of the following sort: to push the purple eject button on the panel in front of you if there is an aggressive alien in your cockpit, if you are a thirty-first-century pilot defending Earth from Plutonian invasion. (Joyce 2001: 121)

The alternative to accepting the first reading, then, is to embrace an explosion of reasons, where we presently have conditional reasons to do all sorts of things. It seems easy to agree with Joyce that such conditional reasons “are very shady customers” (Joyce 2001: 121). Importantly, the distinction drawn in the example above generalizes, and allows for the conclusion that there are no such conditional reasons. All reason claims of the form “S has a reason to Φ if C obtains” should therefore, according to the instrumentalist, be read as “If C obtains, then S has a reason to Φ,” and not as “S has a reason to Φ-if-C-obtains” (2001: 121).

Joyce thinks this reading of reason claims is enough to save hypothetical reasons from somehow entailing categorical reasons, since the particular hypothetical reasons that the instrumentalist are committed to only arise when the antecedent condition is satisfied. On his understanding, then, a hypothetical reason is simply the claim that if a certain situation obtains, and if the agent stands in a particular relation to an action, then the agent has a reason to perform that action. There is no more general imperative or reason, which always holds.

The Humean need not, and should not accept the following imperative: “You ought to (Φ if you want X and Φing is the best means of achieving X).” Rather the Humean endorses this “If (you want X and Φing is the best means of achieving X), then you ought to Φ.” (Joyce 2001: 122)

The difference Joyce introduces, then, is between what has been termed wide and narrow scope imperatives. Wide scope imperatives tell us, categorically, that you ought to Φ-if-C-obtains. A narrow scope imperative, on the other hand, only tells you that if certain circumstances obtain, then you ought to do something: If C obtains, then you ought to Φ. The difference, as we saw above, is that this latter view of imperatives does not entail that one always has a reason to do something, only that such reasons obtain when certain circumstances are in place. Furthermore, it is not the case that reasons obtain independently of the ends of the agent, as in the wide scope reading. Rather, in the narrow scope reading of hypothetical imperatives, it can truly be said that the reason depends for its existence on the ends of the agent.

In what follows, I will grant Joyce that his proposed reading of hypothetical imperatives saves his account from the challenge. There are two reasons for this. First, if it does not manage to save the instrumentalist, then hypothetical reasons would entail strongly reason-giving categorical imperatives, and therefore categorical reasons. And as I have claimed, but not argued, such categorical reasons would be
targeted by arguments for moral error theory. \footnote{At least on the assumption that strongly reason-giving categorical practical reasons are committed to irreducible normativity, a claim I will not argue for at length here, but which I do think can be quite easily secured. Some forms of strong categorical reasons can also be argued against in other ways than by taking them to be committed to irreducible normativity, for instance if they assume that all (rational) agents must share certain ends, where this is not, in fact, the case (cf. the Appendix).} Second, if it \textit{does} manage to save hypothetical reasons, we can proceed, as we shall, with looking at their structure in light of the favouring relations framework.

4.3 Applying the framework of favouring relations to hypothetical reasons

Having presented a way to conceptualize normative reasons in Chapter 2, let us now attempt to apply it to hypothetical reasons and see if we cannot say a bit more about their structure. Let us take a paradigm case of a hypothetical reason, and the scenario surrounding it.

\textit{THIRSTY} You desire water. You see that right across from you is a water cooler with clear, clean drinkable water, which is available to you at no cost.

If we try to break 	extit{Thirsty} up into its elements in accordance with how we have done previously, we might get the following partitioning: There is a fact, namely the psychological fact that you desire water, and there is an agent, you, and a course of behaviour, getting a glass of water from the water cooler. \footnote{One might not want to use phrases such as “the fact that you desire water”, but rather “the desire for water”. As I said when I introduced the framework, the thing which does the favouring – the consideration, or the ground – can be different things, though I usually employ “fact” to capture this role. See Lillehammer (2010) for an argument that “the fact that you desire” and “you desire” are interchangeable in terms of playing the role of a consideration in a reason relation.} Let us again operate with the simple form of practical instrumentalism introduced earlier, which says that an agent has a reason to do something, insofar as it would satisfy some present desire of the agent.

One question that such hypothetical reasons seem to invite is how a normative reason can arise from the existence of a psychological fact such as a desire (or some other end of an agent). One natural way to explain this is, unsurprisingly, is to appeal to favouring relations. Let us therefore apply the favouring relations framework. We can then construe hypothetical reasons as follows.

\begin{center}
\textbf{The conceptual structure of a hypothetical practical reason}
\end{center}

\begin{center}
\begin{tabular}{ccc}
\textit{Fact} & \textit{Circumstances} & \textit{Agent} & \textit{Behaviour} \\
You desire water & Available for free & You & Getting water \\
\end{tabular}
\end{center}

\begin{center}
\textit{A hypothetical reason}
\end{center}

There is a hypothetical reason for you to get water.
The suggestion here, then, is that the psychological fact that an agent desires water to favour the agent performing the action of obtaining and drinking a glass of water. This seems to be natural interpretation of a hypothetical imperative, and of the ways in which we talk about them.

The question that now arises is what kind of favouring relation is involved. In order to settle this, we should look at how such reasons are understood in the discourses involving them. With respect to institutional facts such as the rules of games, we saw that they involve favouring relations that are reducible to a correct-according-to relation. This, however, does not seem like a plausible candidate for the favouring relation involved in hypothetical reasons. It seems that, unlike institutional reasons, you really do have a substantive reason to comply with hypothetical reasons, especially since they are usually taken to spring from the dictates of practical rationality (cf. Joyce 2001: Chapter 3).

In *Thirsty*, it is not the case that the only thing that gives you a reason to drink is that it is correct according to practical rationality, in the sense that there obtains a correct-according-to relation. Of course, it is correct according to practical rationality, but in addition, there seems to be something beyond that, something stronger that favours you acting in accordance with practical rationality. However, if the favouring relation involved in hypothetical reasons is not reducible to a correct-according-to relation, what does it then consist of? The relation seems surprisingly similar to the moral favouring relation, and most of the things said about it in Chapter 2 seem to hold here as well. If the favouring relation involved in hypothetical reasons is not irreducibly normative, there must therefore be some other way for desires (or other ends) to favour certain courses of behaviour. But I have no suggestions for what such a reducible favouring relation could be.

As far as I can tell, there is no other way to explain how psychological states like desires give rise to reasons with strong normative force, other than by explaining it in terms of irreducible favouring. With respect to hypothetical reasons, then, it looks like the favouring relation involved is irreducibly normative – there is nothing further we can reduce it to, as we could with institutional reasons. As Bedke puts it:

If the metaphysical objection against ‘objectively binding’ reasons holds, it applies with equal force to end-given reasons [i.e. hypothetical reasons], for end-given reasons claim that there are favourers and disfavours, too; they just claim, to take a representative Humean view, that the only favouring consideration, F, is that an action would advance one’s ends. (2010: 48)

Therefore, if we want to understand hypothetical reasons in terms of favouring relations, as it seems we should, then we must understand them in terms of *irreducibly normative* favouring relations. If this is the case, error-theoretic arguments against the irreducible normativity of moral reasons generalizes to hypothetical reasons as well.
At this point, one might begin to wonder if there has not been made a mistake. We might, for instance, think with Joyce that “the fact that one might thus squeeze a [metaphysically objectionable] reason from radical Humean instrumentalism should arouse suspicion” (2001: 120). Indeed, this seems like an incredibly surprising turn of events – that practical instrumentalism, which is usually taken to be one of the few, if not the only, metaphysically low-cost options for “real reasons” should turn out to be conceptually committed to such a metaphysically loaded and objectionable notion. That it has taken so long to recognize this should border on the unbelievable.

I am sensitive to this objection, though, of course, it does not carry much weight in itself. In what follows, I will therefore look at ways of trying to account for hypothetical reasons in a way that does not commit them to irreducibly normative favouring relations.

4.4 The prospects for hypothetical reasons without irreducible normativity

In trying to see what could have gone wrong, we might start by thinking again about how hypothetical reason talk functions in ordinary discourse. On this note, Joyce sketches the following scenario:

Consider, for example, someone who asks, “Why ought I get on the train?” The Humean thinks that an adequate answer may be “Because you desire to go to London, and getting on the train is the best means of achieving it.” But the person continues with a second question: “Why does the mere fact that I desire something give me a reason for performing an action that will lead to its satisfaction?” Now at this point what the Humean should not do is trot out a more general imperative: “Because you ought to follow this principle: if you desire X, and Φing is the best means of achieving it, then you ought to Φ.” This would be futile, for the mere fact that the person needed to ask the second question shows that she is not going to accept this as an answer. Rather the Humean thinks that the questioner does not properly understand what the word “reasons” means. (Joyce 2001: 122–123)

This line of thought might explain what has gone wrong in the above analysis. When it comes to hypothetical reasons, in virtue of the meaning of the word “reason”, it is simply not legitimate to push the question of why you should do something beyond the information that you have an end (desire or otherwise) which would be fulfilled by Φing.119 Seemingly one could always perform the above line of reasoning against an instrumentalist account “But why does my ends in some way tell me what to do?”

One way to escape this incessant questioning, which seems to bottom out in an appeal to irreducible normativity, is the following: The word “reason”, at least in one legitimate sense, simply means something like taking effective means towards one’s ends. If this were the case, one might be inclined to think that there is no mystery involved, and no necessity in an appealing to irreducibly normative favouring relations. I will try to spell out an instrumentalist reply of this sort as clearly as I can, in order to see whether we might not capture the notion of a practical reason in this metaphysically lightweight way.

118 Joyce is here concerned with external reasons, but again, the point remains the same.
119 While this might seem implausible on this crude account, I push the same point against a more advanced form of instrumentalism in section 4.5.
I have, in Chapter 2 and the Appendix, said something about why it does not seem possible for moral reasons to be reduced to empirical and natural phenomena. However, the same might not be true of hypothetical reasons. It might even seem like they should in principle be easier to reduce, since our concept of a hypothetical reason, at least at first glance, seems much less metaphysically extravagant than, for instance, that of a moral reason. Let us therefore investigate the possibility of reducing hypothetical reasons in a way which does away with the irreducible normativity and which only appeals to metaphysically unobjectionable elements.

As a first approximation, we could attempt to claim that the structure of hypothetical reasons is something like this:

What it is for \( F \) to be a [hypothetical] reason for \( A \) to \( \Phi \) is for \( A \) to have a goal with content \( F \), and for \( \Phi \)ing to further that goal. (Bedke 2010: 50)

This mirrors the claim we looked at above, that all there is to being a (hypothetical) reason, is to say something about what would be a way of satisfying an agent’s ends. Let us see if we can give an account of the reason involved in \textit{Thirsty} discussed earlier based on this understanding of hypothetical reasons, which I will call the empirical conception of hypothetical reasons (cf. Olson 153–155).

\begin{center}
\textbf{The conceptual structure of an empirical hypothetical reason}
\end{center}

\begin{center}
\begin{tabular}{ccc}
\textit{End} & \textit{Agent} & \textit{Behaviour} \\
Desire for water & You & Getting water \\
\end{tabular}
\end{center}

\begin{center}
\begin{tikzpicture}
\node (A) at (0,0) {Desire for water};
\node (B) at (1,0) {You};
\node (C) at (2,0) {Getting water};
\draw[->] (A) -- (B);
\draw[->] (B) -- (C);
\end{tikzpicture}
\end{center}

\begin{center}
\begin{align*}
The \textit{action} \text{ will (causally) satisfy or further the end}. \\
\begin{minipage}{0.5\textwidth}
\textit{An empirical hypothetical reason} \\
A hypothetical reason to get a glass of water.
\end{minipage}
\end{align*}
\end{center}

On this account, then, all that is involved in being a hypothetical reason is for some action to stand in a certain (causal) relation to an agent’s ends. What is required is that the action is such that it would (help) satisfy the end of the agent. This account of hypothetical reasons seems to manage without any form of favouring relation, and it does not seem to involve any metaphysically objectionable content. As such, empirical hypothetical reasons do not immediately seem to fit into the favouring relation framework. This would seem to be an example of an account of a normative reason that manages to reduce what it is to be a reason with strong normative force to something other than irreducible normativity. In fact, it would seem to reduce the normativity involved to a causal relation. The question now becomes whether this is a plausible account of the way hypothetical reasons are used in ordinary discourse.
It does indeed seem to capture at least one way that we ordinarily use the term “reason” (Olson 2014: 153–155). Let us look at some examples:

John has a reason not to confess to the police.
Thieves have a reason for wearing gloves.
Putin had a reason for invading Ukraine.

These are cases where the word “reason” can be interpreted simply to mean that the actions are efficient means to satisfy some end the agent has. All that is being said, arguably, is that performing the action *would be an efficient means* to satisfying one or more of their ends. In these cases, this should not be taken to imply that there is some irreducibly normative favouring relation involved – one is not necessarily saying that their motivational states in some sense favours the associated action, or that there are moral reasons in play (there may very well be, but that is orthogonal to *this* use of “reasons”). Rather, all that is being said on this use, is the empirical claim that the actions are, in fact, a means to satisfy one or more of the agent’s ends. It therefore seems that we have found a linguistically legitimate use of “reason”, where this is all that is being said. On this empirical account of hypothetical reasons, then, there is no irreducible normativity, and it will therefore be no problem for the error theorist, or anyone else, to accept the existence of *this* kind of hypothetical reasons.

If this covers *all* uses of hypothetical imperatives, then it indeed seems like we have managed to give a naturalistically acceptable and metaphysically unproblematic account of such reasons. Unfortunately, there seem to be considerations that pull in the opposite direction.

### 4.4.1 The objection from normativity

I believe there is a certain intuition underlying instrumentalism to the effect that if anything gives us a reason to act, it’s our ends. *If* you are thirsty, *then* you have a reason to get something to drink. If someone asks further questions about why this constitutes a reason to do it the instrumentalist could point to the empirical account outlined above, and simply claim that *that* is how we actually do use the word reason. Or less, presumptuously, that *if* an empirical account is the only way for hypothetical reasons to avoid irreducible normativity, then it is the only *defensible* form of reason talk.

It is therefore worth asking whether such an account manages to capture everything we associate with practical reasons. To see why this is a pressing question, consider the elements that the empirical account takes to be sufficient for there to obtain a hypothetical reason. As we saw, what is needed is an agent with some kind of end, where this might, for instance, be a psychological state, such as a desire and a course of action which would satisfy the desire. In cases where these things are present, it is linguistically legitimate
to speak of the agent having a reason for pursuing the action, in the sense that the action would (causally) help fulfil an end of the agent.

As I said in Chapter 2 with respect to moral reasons, these ingredients taken together do not seem to be sufficient, conceptually speaking, for there to be a normative reason in play. As I have defined them, a normative reason must, at the very least, tell us what to do. However, it is difficult to see in what sense empirical hypothetical reasons tell us to do anything at all. Say that you have a desire, and that there is a course of action that would help satisfy your desire. Does this actually tell you to do something? Not as far as I can tell. What is missing, of course, is the sense in which this fact – that some course of action would help fulfil an end of yours – favours you acting in that way. Without this element, it is difficult to see how one can cook up a normative reason from the ingredients of an empirical hypothetical reason. And remember, the bar for being a normative reason has been set stupendously low; all that is needed to be a normative reason is to be prescriptive – to tell us what we should do, in some sense or other. Even a rulebook for a game you just made up is capable of fulfilling this, in that the rules in it would give rise to institutional reasons.

The instrumentalist, I suspect, would not be impressed by this line of reasoning. They would claim that we are still missing the point. To be a normative reason, the instrumentalist could claim, simply is to be an empirical hypothetical reason, and therefore, to be a normative reason – a real reason to do something – just is for something to be an efficient way of fulfilling one of your ends. The instrumentalist about practical reason could further claim that the empirical understanding of hypothetical reasons even has the virtue of explaining what it is for something to have strong normative force. In response to this, we might have the lingering feeling that something has been explained away, as expressed by Bedke:

> It seems to me that we would then ignore a concept we are all stuck with and that we all want to retain: that of a normative reason. And to talk about normative reasons one must claim, at the very last, that certain facts about available courses of action normatively relate to actions and attitudes. (2010: 50)

But again, the proponent of the naturalistic reduction might throw up their hands and claim that “this is all the normativity there is!” As is often the case when considering whether naturalistic reductions are sufficiently satisfying, it can come to an impasse, with proponents simply saying, “Yes, we’re satisfied” and opponent’s saying “No, there’s something important missing!” In the following, I will try to present a way in which we might try to break up this stalemate, and one that proceeds on terms the proponent of naturalist reductions can accept.

4.4.2 Competing reductive accounts and normative disagreement

In order to flesh out the notion of a empirical hypothetical reason we must get more specific. So far, we have only looked at one way of cashing out such a reduction of hypothetical reason talk, namely that the only reasons there are for an agent to do something is to do that which would satisfy an end of an agent. This, however, is too vague. In order to get a clearer picture of what reasons we actually have, we must get clear on which of our ends it is that gives rise to reasons. Fortunately, there is no lack of candidates. The crude
view we have already looked at, which says that one has a reason to do what would satisfy one’s actual, present desires, we can call present aim instrumentalism.120

**PRESENT AIM INSTRUMENTALISM**

$A$ has a reason to $\Phi$ iff $A$ has a present desire, $E$, that $\Phi$ing will satisfy.

This theory of normative reasons tells us that what we have reasons to do, are those actions that would satisfy our occurrent desires. As we have seen, an unqualified “ought” or “should” can be understood to simply be equivalent to what we have most reason to do. If we only consider practical reasons, present aim instrumentalism (PAI) therefore tells agents that they should maximize the satisfactions of their present desires.

A rival view to present aim instrumentalism is complete life instrumentalism.

**COMPLETE LIFE INSTRUMENTALISM**

$A$ has a reason to $\Phi$ iff $\Phi$ing will best satisfy $A$’s desires during the course of $A$’s complete life.

Complete life instrumentalism (CLI) tells us that what we have reason to do is that which would maximise the satisfaction of your desires over your complete life. Therefore, if you have a very strong desire for something today, but the present satisfaction of which will frustrate some even greater desires that will come along later, then you should abstain from acting on your present desire, despite it being the strongest desire you possess right now.

These, then, are two competing accounts of what we have reasons to do. While they may sometimes give the same answer, they are likely to often disagree about what a given agent should do – there will be normative disagreement between them. It will be instructive, I believe, to study such normative disagreement between competing instrumentalist accounts in order to get a clearer view of what they are committed to. Let us therefore look at a particular case of such disagreement.

**CAKE**

You see that there is a delicious cake on the table in front of you, with a note that says, “Help yourself!” You have a very strong present desire for cake. Choosing to act on the desire to eat the cake will frustrate (or rather, contribute to frustrating) an even stronger future desire, the desire not to become overweight.

120 As we are not going to evaluate how good these theories are in themselves, but only look at some interesting features of setting them up side by side, it does not matter that they are not very plausible. I adopt the general argumentative strategy employed in this section from Bedke (2010).
In this scenario, the present aim instrumentalist would claim that what would maximize the satisfaction of your actual, current desires is to eat the cake, and that, consequently, what you have most reason to do – what you should do – is to eat the cake. The complete life instrumentalist, on the other hand, will insist that what you really have most reason to do is to not eat the cake, as this will lead to a greater overall degree of desire-satisfaction during the course of the agent’s lifetime.

There is, then, real disagreement here. Remember, these two accounts are meant to be ways to specify how we should understand empirical hypothetical reason. In essence, we’re asking which ends and which causal relations gives rise to reasons. With this in mind, it is instructive to notice that the disagreement between PAI and CLI is not about what you are motivated to do, what you think you should do or any other such empirical claim (Bedke 2010: 49). Furthermore, it does not seem like the discussion is about what the word “reason” means; it would seem that the proponents of the competing accounts are perfectly agreed on a general sense of the word, namely that we, in some sense to specified, have reason to do that which satisfies our ends. What they are disagreeing over is, quite plausibly, which things have reason-giving powers and, consequently, what reasons there actually are. As they are instrumentalists, they agree that it is only ends that have such reason-giving powers, but they disagree over which ends should be taken into account. How can we decide this?

Well, it would seem that in order to decide on whether we should be “present aimers” or “complete lifers” we would need to discover what we really have reasons to do. Now, there does not seem to be any empirical way of settling this debate, as what they are disagreeing about is not empirical details – they are, after all, in complete agreement concerning the relevant details. Furthermore, they agree completely about what would be correct according to CLI or PAI. They are not disagreeing about that. Again, it seems that the disagreement boils down to the question of what we really have reasons to do.

If the instrumentalist wants to answer this question, it appears necessary, at some point, to argue that certain facts, such as a certain restricted set of ends, in some way counts in favour of fulfilling them. However, as soon as the instrumentalist does this, we are no longer only entertaining an understanding of hypothetical reasons where the features involved are an end, an agent, a behaviour and a causal relation. What we then get, of course, is the picture of hypothetical reasons I laid out in 4.3, where some set of ends favours an agent acting in a certain way in a given context. The instrumentalist, therefore, does not seem capable of escaping the favouring relations framework by reducing all hypothetical reasons to some empirical configuration of elements.

In short, hypothetical reasons seem to possess strong normative force, but this, it has now become clear, is because the favouring relation involved in such reasons are in fact irreducibly normative.

---

121 As always, there is the possibility of eliminating the desire. However, as we all know, that is no simple task when confronted with a delicious piece of cake.
Furthermore, those hypothetical reasons which do not involve irreducibly normative favouring relations are, in fact, not normative reasons at all.

4.5 What about other accounts of practical reasons?

In the above, I have only looked at a few accounts of practical reasons, and what kind of structure they impart to such reasons. It might therefore be objected that there is some other account of practical reasons which escapes my charge that it is best understood in terms of irreducibly normative favouring relations. I will deal with two variants of this argumentative strategy. The first of these takes instrumental rationality to not give rise to practical reasons at all, and the second defends a theory of practical reasons where such reasons are taken to have a close connection to rational deliberation.

As we have seen, there does seem to be one interpretation of hypothetical reasons where they are metaphysically unproblematic, namely when they are taken to be empirical claims about when a certain course of action would, causally, lead to the satisfaction of an end. However, on such an interpretation, hypothetical reasons are not normative reasons, since they do not tell us what we should do. As we saw in the case of normative disagreement between competing accounts of empirical reductions, something more must be added, namely favouring. One strategy then, is to accept this, and also to accept the conclusion that the instrumental principle that underlies practical instrumentalism does not give rise to practical reasons.

Intention reasoning takes you from your intention of buying a boat and your belief that borrowing money is the only means of doing so, to an intention to borrow money. But it does not determine that you ought to borrow money, nor even that you have a reason to borrow money. (Broome 2002, cf. 1999, 2004)

One way of cashing out such a claim, is by holding the instrumental principle “If you have an end, E, and believe that by Φing you will satisfy E, you ought to Φ” to only provide a normative requirement, without necessarily being reason-giving. That is, it could be claimed that since, when one has an end, E, and knows that Φing would effectively bring about E, that does not give you a reason to Φ. Instead there is a general normative requirement that you cannot rationally both desire E and not Φ – you either have to give up the desire, or perform the action in order to satisfy the requirements of rationality (Broome 2002; cf. Wallace 2014: §4). I will discuss two ways in which a theorist can build on this claim, either by denying that there are practical reasons at all, or taking practical reason to fall out of a more complicated set of rational requirements than those we have looked at so far.

First, in order to escape the irreducible normativity of hypothetical reasons, the theorist might stop after this claim, and embrace the conclusion that there are no hypothetical reasons. If we continue to focus on an instrumentalist account of practical rationality, we can further take it for granted that such a theorist does
not believe there to be any non-hypothetical reasons either. As a result, this first option leads a theorist to hold that there are no practical reason at all.

Instead of taking rationality to engender practical reasons, such a theorist could instead opt to only take there to be constitutive and regulative requirements of rationality. Constitutive requirements of rationality would specify what it would take to be counted as a rational agent. These are requirements you must fulfil in order to count as a rational being at all. Such requirements could for instance have to do with, say, certain core elements of practical reasoning as outlined above. Secondly, such a theorist could also accept that there are regulative rules of rationality, rules that determine whether an agent is more or less rational.

Such an account, then, takes there to be no practical reasons, but only structural or constitutive and regulative requirements for being (more or less) rational. On a particular understanding of what it is to be a requirement of rationality, this would amount to a non-normative, empirical theory of rationality and rational requirements. Being rational would therefore be similar to being blue: some things are, some things aren’t (and some things are it some of the time, but not always). This, in short, would be a form of nihilism about practical reasons – there are no practical reasons, and so, all practical reasons claims would be false (cf. Millgram 2001: 3–4). We will look closer at this view in Chapter 6.

Instead of opting for the view that the instrumental principle does not give rise to practical reasons at all, it would be possible to hold that practical reasons will follow from rational requirements:

> [W]hat one has reason to do, on this view, is what one would desire or intend to do if one was fully rational (i.e. fully in compliance with the wide-scope structural requirements that govern one’s attitudes in combination). (Wallace 2014: §4)

A theorist might therefore claim that practical reasons are not some mysterious force, but rather something that simply springs from rational requirements, which are a feature of human rationality. The reason we have missed this fact so far, this theorist might claim, is that we have focused on too simple models of practical reasons.

As far as I can see, however, the more one builds into the models of practical rationality, the clearer it becomes that they are committed to a form of reasons that involve irreducibly normative favouring relations. Let us therefore look at one such comparatively more advanced view of practical reasons which connect them more closely with rationality. Bernard Williams (1979) introduced the idea that what reasons we have, are determined by the desires we would have after a procedure which takes our current, actual motivational set as its input, and runs them through a sound deliberative route which involves the agent having no false beliefs, all relevant true beliefs, and optimal instrumentally rational deliberation.

On this view, no matter how the details are spelled out, we should quickly be able to glean the most natural interpretation of it. After all, the desires in question are not the ones you have now, but they are, rather, the desires you would have after running them through some deliberative procedure. Facts about this
counterfactual situation therefore seem to be what gives rise to practical reasons on this account. I therefore suggest that we can construe this account as follows:

**The conceptual structure of an advanced hypothetical reason**

<table>
<thead>
<tr>
<th>Fact</th>
<th>Circumstances</th>
<th>Favours</th>
<th>Agent</th>
<th>Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>A sound deliberative route would lead you to desire x, and doing y will bring you x.</td>
<td>-</td>
<td>You</td>
<td>Doing y</td>
<td></td>
</tr>
</tbody>
</table>

A hypothetical reason

You have a hypothetical reason for doing y.

This quick sketch of a more advanced instrumentalist view should suffice to show that the claim that hypothetical reasons involve irreducibly normative favouring relations cannot be avoided by adding to the basic instrumentalist theories we have looked at.

### 4.6 Summary

In this chapter, we have seen that moral error theory, suitably specified, generalises beyond the moral domain and to all irreducible normativity. We have seen that such a metanormative error theory would target the domain of normative, practical reasons.

We have found that hypothetical reasons can be separated into two classes. First, there is an interpretation of them where they are purely empirical claims. This sense is indeed unproblematic, although, as we have seen, such reasons are in fact not normative reasons, in that they don’t seem to tell us what we should do. Second, there is a form of hypothetical reasons that do tell us what we should do, but these do entail irreducible normativity.

We can therefore conclude that some hypothetical reasons do entail irreducibly normative favouring relations, while some do not. This further underlines my claim that the hypothetical-categorical distinction does not carve up the space of reasons into those that are metaphysically problematic and those that are not, as many metaethicalists, including both naturalists and error theoreticians, have previously thought.

Instrumentalist theories that subscribe only to the first, empirical, type of hypothetical reasons can be accused of not being able to account for our concept of a normative practical reasons at all. Instrumentalist theories that do embrace normative hypothetical reasons, on the other hand, seem to be committed to irreducible normativity.
In short, then, all, or at least most, normative practical reasons seem to entail irreducible normativity. Such reasons would therefore be targeted by a metanormative error theory. Whether this should lead us to accept the existence of irreducible normativity, or deny that there are many, if any, true practical reason claims will be picked up in Chapter 6.

5 Generalization II: Epistemic reasons

The epistemic nihilist is a disenchanted realist.

Cuneo (2007: 116)

5.1 The reasons of epistemic rationality

In the last chapter, I argued that hypothetical reasons that issue from practical rationality fall victim to a generalized error theoretic argument. More specifically, all hypothetical reasons that deserve to be called normative fell pray it.\(^{122}\) While such practical reasons, reasons for action, are perhaps one of the first types of reasons that comes to mind, there are also other types of reasons that are worth taking a closer look at. One such alternative type of reasons does not deal with what we have reason to do in the practical sense of performing bodily actions and the like, but rather with what we have reason to believe. Such reasons do not provide us with reasons for performing actions, but rather with reasons for epistemic or doxastic behaviour.\(^{123}\)

In this chapter, I will therefore leave issues about practical rationality and reasons for actions behind, and look at this other form of reasons concerning what we have reason to believe. Such reasons are commonly taken to issue from epistemic rationality, a rough characterization of which is given by Thomas Kelly as “the kind of rationality which one displays when one believes propositions that are strongly supported by one's evidence and refrains from believing propositions that are improbable given one's evidence” (Kelly 2003: 612).

5.1.1 Epistemic reasons and ordinary discourse

The kinds of reasons that issue from epistemic rationality therefore have to do with what we have reasons to believe, and as such, they are usually called epistemic reasons or reasons for belief.\(^{124}\) While the term

\(^{122}\) As we saw an empirical, non-normative interpretation of such reason claims that avoids irreducibly normative favouring relations. However, not all, or even most, uses of hypothetical reasons can be interpreted in this manner.

\(^{123}\) In talking of doxastic behaviours, I adopt the use of the word “behaviour” in “its widest sense, to stand for either the intentional or non-intentional activity of agents” (Cuneo 2007: 59). Furthermore, I make no assumptions about debate concerning the truth of doxastic voluntarism, which is, roughly, the view that “people have voluntary control over their beliefs” (Vitz 2015). The central arguments of this chapter should hold regardless of ones position on that issue.

\(^{124}\) While it is common to belief as the paradigmatic epistemic state, it is possible to argue that the scope of epistemic rationality is broader than this (Cuneo 2007: 71–73). In any case, I will focus on beliefs.
“epistemic reason” is unlikely to be used all that much, if at all, in ordinary discourse, the concept certainly seems to be found in a surprisingly large number of ordinary contexts. In order to demonstrate this, I will first provide some examples of paradigmatic classes of epistemic reasons, and I will also say something about what features ordinary discourse ascribes to such reasons.

The first class of cases where we, arguably, think there are reasons for adopting a certain doxastic behavior involves perception. When asked why you believe there to be a water leakage, it would appear that pointing out that you see the water leaking out of the pipes provides you with a reason for this belief; the observation seems to give you a reason to believe that there is a water leakage. In short, beliefs formed on the basis of how something appears to us under normal circumstances seems to be supported by a (defeasible) reason.  

Another class of everyday examples of epistemic reasons spring out of the nature of logical reasoning and deductive inferences. It would seem that believing that if \( p \) and also believing that if \( p, \; \text{then} \; q \) would provide you with a reason to believe \( q \). If you believe that the light is on, and you also believe that if the light is on, then John is home, then you have a reason to believe that John is home.  

One interesting fact about such epistemic reasons and the discourse involving them is the surface similarities they share with moral and practical reasons and discourse. In fact, our discourse about epistemic reasons is in many ways surprisingly similar to moral and practical discourse (cf. Alston 1988; Feldman 2004a). While we might say that someone “ought to \( \Phi \)” in the ethical or prudential case, we might also say that given the information available to a person, they “ought to believe that \( p \)”, or that given certain scientific evidence, a scientist “should not believe that \( p \)”. There is also corresponding reason talk, as when we say that an agent has a “reason to believe that \( p \)”, or that “the evidence indicates that we should not believe that \( p \)”.

Richard Feldman brings out these analogies between moral and epistemic discourse clearly:

In the case of ethics, we say that a person ought to perform a certain action, that someone should not do a certain thing, that people have obligations to act in some ways, that they have rights and duties, and that they deserve praise or blame for what they have done. We make seemingly analogous epistemic judgments about beliefs. We say that a typical well-informed contemporary American ought to believe that the Earth revolves around the Sun and should not believe that the Earth is flat. We say that a person has a right to believe one thing and perhaps a duty to refrain from believing something else. We sometimes praise those who believe the things they should and we criticize those who fail in their believings. (Feldman 2004a: 166)

125 The exact relation between appearance, or more generally, perception and what we have reason to believe is a source of much debate. Nothing of importance for this chapter hangs on the details of resolving this issue. For an overview of the debate about perceptual justification of beliefs, see Silins (2015). Even if perception did not give rise to reasons for belief, we could simply substitute appearance examples with examples such as those in the next paragraph. Alternatively, we could take our ordinary conception to hold that perception provides reasons for belief, but that this conception is erroneous.

126 Of course, if the conclusion is sufficiently implausible, it may instead make you want to reconsider the premises (Harman 1973). I will not include this qualification from here on, but take it as read.

127 It also seems natural to take the relation between what we ought to believe and what we have (a) reason to believe to be more or less the same as what I said about the relation between “ought” and moral reasons in Chapter 2.
To further illustrate these similarities, remember that moral and practical considerations can favour or disfavour courses of action. From a cursory glance at epistemic reasons, they seem to share such structural features. It would seem that, say, believing that \( p \) and if \( p \), then \( q \) favours believing that \( q \). Similarly, if the agent believed both \( p \) and if \( p \), then \( q \), but rather believed that not-\( q \), it would be natural to say that having the former beliefs disfavours believing that not-\( q \). Why do we think there are such epistemic reasons? One plausible answer is that we take such epistemic reasons to be what we have reason to believe “from the perspective of representing reality aright, gaining true belief, knowledge, understanding, or the like” (Cuneo 2007: 117). In other words, facts about truth, justification, knowledge and evidence appear to favour (or disfavour) certain doxastic behaviours, that is, behaviours such as forming, abandoning or suspending beliefs.\(^{128}\) These apparent similarities should be enough to motivate us to investigate whether epistemic reasons are best understood in terms of such favouring, and if they are, what kind of favouring relation they involve.

Just as with practical rationality and the reasons that spring from it, there are competing accounts of what exactly epistemic rationality consists in, and about the nature of the reasons that issue from it. For instance, some philosophers have argued that epistemic rationality is, in fact, just a form of instrumental rationality, and that it therefore only gives rise to instrumental (i.e. hypothetical) reasons.

Epistemic rationality ... is simply a species of the genus instrumental rationality ... Epistemic rationality, no less than any other sort of rationality, is a matter of integrating ends and means ... Good reasons are instrumental reasons; there is no other sort. (Laudan 1990: 318, cited in Kelly 2003: 614)

If all epistemic reasons were in fact hypothetical reasons it would not be necessary to devote more space to the discussion of them, as the arguments of the last chapter would be directly applicable. However, as I will show below, there is evidence that epistemic reasons cannot be subsumed under the heading of hypothetical reasons so easily.\(^{129}\)

Furthermore, as it is reasons themselves that are my main concern, I will not dwell on the nature of epistemic rationality over and above carrying out an exploration of what are relatively uncontroversially taken to be epistemic reasons, in the sense that they are reasons to form, abandon or suspend belief. My main concern in this chapter is to explore the nature of the kind of epistemic reasons that ordinary discourse is committed to.

\(^{128}\) For our purposes, it is the interplay of evidence, reasons and agents that is most interesting. In contemporary epistemology it is just as usual to simply focus on the relations that hold between epistemic states. This is highlighted in the difference between personal justification, “\( A \) is justified in believing that \( p \)”, and doxastic justification, “\( A \)’s belief that \( p \) is justified”. These levels can sometimes come apart. For a discussions of such different levels of focus, see Littlejohn (2012:5–8).

\(^{129}\) In section 5.3.3 I will return to discuss the possibility of reducing epistemic reasons to hypothetical reasons. See also Kelly (2003) for a critique of understanding epistemic rationality as a form of instrumental rationality.
We have already looked at some examples of how epistemic reasons figure in everyday contexts, but we have not yet looked at the features of these reasons. As I suggested in Chapter 2, normative reasons can helpfully be understood in terms of whether or not they possess the following three features: prescriptivity, inescapability and strong normative force. In The Normative Web (2007), Terence Cuneo conducts an investigation of epistemic facts and reasons and in particular of their similarities to moral facts and reasons. Cuneo has therefore attempted to capture what he calls the “commonsensical conception of epistemic facts” (2007: 56–59). As facts for Cuneo are reasons (in the property sense), and reasons entail reason relations, we can safely translate his talk of epistemic facts to talk about epistemic reasons.

According to Cuneo, our ordinary conception of epistemic reasons contains, among other things, certain platitudes to the effect that they are authoritative. Cuneo claims that such reasons are authoritative in at least two ways. First, our ordinary conception of epistemic facts is committed to them being prescriptive “in the sense that they are, imply, or indicate reasons for agents” (2007: 58). This aligns fairly well with our previously introduced notion of prescriptivity, where it was taken to apply to anything that tell us what we should do. And, by Mackie’s platitude, things that tell us what we should do “are, imply, or indicate reasons for agents”. It is therefore easy to agree that since epistemic reasons tells us what we should do, they are prescriptive in this sense. In order to illustrate the prescriptivity of epistemic reasons, Cuneo provides the following example:

It appears, for example, to be a conceptual truth that, if a certain plan of inquiry is manifestly irrational, then its being irrational provides appropriately situated agents with an epistemic reason not to engage in it. (2007: 58)

In other words, the fact that some intended plan is epistemically irrational seems to disfavour going ahead with it. This much should be relatively uncontroversial. It seems like most parties to the debate can agree that we ought to avoid epistemic irrationality from the epistemic point of view. This does not entail that epistemic reasons are any stranger than, say, institutional reasons, where the institution in question is, say, the norms or rules of our epistemic practice.

The second sense in which Cuneo claims our ordinary conception of epistemic reasons is committed to their being authoritative has to do with them inescapably (i.e. categorically) governing our conduct (2007: 59). As he puts it:

The fundamental idea in this case is that some epistemic facts are, imply, or indicate categorical reasons for agents to behave in certain ways regardless of whether these agents care about conducting their behaviour in a rational way. (2007: 59)

---

130 Cuneo often talks about facts since he uses “reason” in the property sense discussed in Chapter 2 (Cuneo 2007: 65), where a reason is a consideration (e.g. a fact) which favours something. A fact, on this conception, can therefore be a reason (in the sense that the fact favours something). As all reasons in this sense entail that there is a reason relation (i.e. the reason is someone’s reason etc.), epistemic facts therefore entail epistemic reasons.
Cuneo therefore thinks that, for instance, an epistemic reason to believe a proposition when having compelling evidence in favour of doing so obtains independently of the agent’s ends. In this sense, then, epistemic reasons are therefore like moral reasons and unlike hypothetical reasons in being inescapable. However, as we have seen, not all inescapable reasons are metaphysically troublesome. The profile we have sketched of epistemic reasons so far in fact seems to match the profile given for institutional reasons in Chapter 2.

The important question about epistemic reasons is therefore whether they are in possession of strong normative force (e.g. by being irreducibly normative). Of all the kinds of reasons we have looked at so far, all reasons that have had strong normative force have been committed to irreducible normativity. Unless we find a counterexample, this supports my claim that we should take the possession of such force to simply consist in their being irreducibly normative. In order to get at the question of whether epistemic reasons possess strong normative force, and whether this is in virtue of being committed to irreducible normativity, we must take a short detour and look at a concept that is central to epistemology and closely related to reasons for action.

5.1.2 Epistemic justification and ordinary discourse

We have now seen how we might anchor epistemic reasons in ordinary discourse and looked at some of their features. However, when explicitly discussed under the heading of epistemic reasons, such discussions take place in academic literature, within epistemology. In order to link up our description of our ordinary conception of epistemic reasons with the literature on epistemic reasons and related notions, I will now introduce and discuss an important and relevant epistemological concept, namely epistemic justification.

In ordinary discourse, and in the epistemological literature, when it comes to the question of what we have reason to believe, this is often connected to of what we are justified in believing. Because of this, I now wish to briefly sketch an answer to what we mean by “justified” when it comes to belief in ordinary settings. In order to do this, it will be helpful to think of a case where we would speak of someone as being justified or unjustified in believing that $p$.

Failing grade

A student has handed in a draft of a term paper to the course teacher very late. The teacher managed to look it over only in time to hand it back to the student the day before it’s due. In her comments, the teacher writes that the paper, as it is, will certainly not get a passing grade. The student does not have time to revise it, but nonetheless believes that the paper will get a passing grade. There is no other evidence available to the student that supports this belief.

131 There is, of course, a close parallel with ethical and practical forms of justification also here, such as when we talk of whether someone acted in a justified manner or not.
Is the student epistemically justified in believing that the paper will get a passing grade? Clearly not. The student is not justified, because there is no evidence available to the student that supports the belief, and in fact, the student possesses strong evidence to the contrary. The student should therefore not have believed that the paper would get a passing grade, and can be held to be epistemically blameworthy for doing so. With this brief sketch in place, let us now try to relate this ordinary conception of justification to that which we find in the philosophical literature.

In epistemology, discussion about what makes a belief epistemically justified has raged for a long time. Such a question is also relevant to our enterprise here, because discussions about what we *have reason to believe* are tightly connected to questions about what we are *epistemically justified* in believing. In short, a part of the enterprise of epistemology has consisted in explaining what we *should* believe, often indirectly through deciding what we are epistemically justified (from here on, simply “justified”) in believing. This is often made explicit, as Pollock does in *Contemporary Theories of Knowledge*:

> I have taken the fundamental problem of epistemology to be that of deciding what to believe. Epistemic justification, as I use the term, is concerned with this problem. Considerations of epistemic justification guide us in determining what to believe. We might call this the “belief-guiding” or “reason-guiding” sense of “justification”. (Pollock 1986: 10)

A conception of justification that takes such a notion to have a belief-guiding normative force, and which also gives rise to obligations and duties of various sorts, is commonly called the *deontological conception of justification*.

> The terms, ‘justified’, ‘justification’, and their cognates are most naturally understood in what we may term a “deontological” way, as having to do with obligation, permission, requirement, blame, and the like. (Alston 1988: 257)

This conception of justification and belief-guidance is not new. The historical lineage arguably goes back arguably to Plato, and continues through Descartes, who after all named his epistemological treatise *Rules for the Direction of the Mind* (1985), and to Locke who explicitly endorses this conception:

> [H]e that makes use of the light and faculties God has given him, and seeks sincerely to discover truth by those helps and abilities he has, may have this satisfaction in doing his duty as a rational creature, that, though he should miss truth, he will not miss the reward of it. For he governs his assent right, and places it as he should, who, in any case or matter whatsoever, believes or disbelieves according as reason directs him. (Locke 1959: 413, my emphasis)\(^{132}\)

This should not be surprising. After all, it arguably springs out of the roots of epistemology as a discipline to be concerned with just such questions as what one has reason to believe, and, in the end, what one should believe. Neither is this just an historical phenomenon, as such goals still shape much of contemporary

---

\(^{132}\) Note that as Locke can be seen as holding a divine command theory of morality, it might be that not sincerely seeking truth is not only epistemically blameworthy, but morally blameworthy as well (Chignell 2010: §1.2).
epistemology. For instance, a deontological conception underlies many explicit definitions of what justification is:

One is justified in being confident that \( p \) if and only if it is not the case that one ought not to be confident that \( p \); one could not be justly reproached for being confident that \( p \). (Ginet 1975: 2)

This then, has been a traditional approach to both epistemology and justification, and it is closely aligned with the notion of justification that we find in ordinary epistemic discourse (Alston 1988). A general definition of such deontological accounts can be given as follows:

\[
\text{NON-DEONTOLOGICAL JUSTIFICATION} \quad S \text{ is justified in believing that } p \text{ if and only if } S \text{ believes that } p \text{ while it is not the case that } S \text{ is obliged to refrain from believing that } p. \quad \text{(Steup 2005: §2.1)}
\]

An agent can therefore only be justified in believing something if the agent is not, for some reason, epistemically obligated \textit{not} to believe it. To see how such a conception can relate to epistemic reasons, recall the \textit{Failing Grade} scenario from earlier. The student seems epistemically obligated to \textit{not} believe that he will get a passing grade. This, in turn, explains why he is not justified – because he was obligated not to, epistemically speaking.

There are different ways in which beliefs may be justified, for instance by perception or by the logical inference, and this in turn may be seen as giving rise to reasons for belief. Hartry Field paints a vivid, if tongue in cheek picture of how this is supposed to work with respect to perception:

A common picture of justification among epistemologists is that typically when a person is looking at something red, her sense impressions pump in a certain amount of justification for the belief that there is something red in front of her; but that there can be contrary considerations … that may pump some of this justification out. In addition, the justification provided by the senses can be fully or partially undercut … so that not all of the justification gets through. On this picture, the job of the epistemologist is to come up with an epistemological dipstick that will measure what overall level of justification we end up with in any given situation. (Field 2009: 250)

However, not all contemporary accounts of justification adopt the deontological understanding of epistemic justification. Since the deontological conception of justification seems to match our ordinary concept of epistemic justification, I will investigate the relations between agent, justification and epistemic reasons first. I will return to non-deontological conceptions of justification in section 5.3.2.

5.2 The structure of epistemic reasons

Having now looked both at some of the features of our ordinary conception of epistemic reasons, and at the way in which deontological accounts of justification are related to them, let us move on to try do dissect the structure of such epistemic reasons, and see if we can relate them to the favouring relations framework. This
will also allow us to consider what degree of normative force such reasons possess, and whether they are committed to irreducible normativity.

As we have seen, a helpful way to begin dissecting the structure of a type of reasons is to look closer at a paradigmatic example of such a reason and the situation involved. I will use one involving appearance, in line with the type of examples mentioned earlier.

**COFFEE CUP**

You sit at your desk in a brightly lit room, you have slept well and are otherwise in a normal, attentive state of mind. Your eyes focus on a cup of coffee that stands on the side of the desk in front of you. You form the belief that there is a cup on the desk in front of you.

If we grant that the situation outlined above is paradigmatic for situations involving epistemic reasons, let us look at how we should construe the nature of the reason involved. First, there is the fact that I perceive a cup of coffee, and that this happens under normal circumstances where I am awake, attentive and not under the influence of, say, hallucinogenic drugs. In response to such a perception it is likely, as a matter of psychological fact, that I form the belief that there is a coffee cup on the side of the desk in front of me. However, this causal fact does not at first glance seem to capture what we mean when we say that seeing the coffee cup gives me a reason to form a belief, or, if I have previously formed the belief, maintain it. Rather, we would be inclined to say that forming or maintaining the belief is *justified* on the basis of the observation, and that believing it is not only something we, as an empirical fact, is statistically likely to do. This is because whether or not we actually do so, we nonetheless have a reason to do so, epistemically speaking.

How are we to account for this seemingly normative relation between the perception and the belief formation (or maintenance)? Understanding it in terms of a favouring relation would seem to be the natural way to cash the way this happens. We have a reason to form the belief (or maintain it) because of the fact that we perceive something under normal circumstances *favours* that particular course of epistemic behaviour.

What kind of favouring relation is it that such reasons entail? Intuitively, such reasons might seem more like moral and prudential reasons than, say, institutional reasons. This is a view that Cuneo gives voice to:

[E]pistemic reasons are authoritative inasmuch as the decisiveness of some such reasons for an agent is not a function of whether she wants to act in an epistemically commendable fashion, or belong to a certain social group, or has entered into certain agreements with others, and so forth. (2007: 59)

What Cuneo seems to get at is that (at least some) epistemic reasons have *strong normative force*, and that this cannot be understood as a function of some social or conventional set of institutional norms or rules.
As discussed previously, epistemic reasons seem to involve favoring, at least at first glance. And while this brought out by an example involving perception, such as COFFEE CUP, it can be helpful to appeal to what Cuneo points out as especially strong epistemic norms, namely the norms of logic. Let us therefore consider the following everyday situation that seems to involve an epistemic reason related to the norms of logic:

\[ \text{IF } P, \text{ THEN } Q \]
You believe that if the lights are on at John’s house, then John is home. You walk past his house and notice that the lights are on. You have no evidence that John is not home.

An attempt to extract the epistemic reason involved is the following.

**The conceptual structure of an epistemic reason**

<table>
<thead>
<tr>
<th>Fact</th>
<th>Circumstances</th>
<th>Favours</th>
<th>Agent</th>
<th>Doxastic behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>You believe that ( p ), and you believe that if ( p, \text{ then } q )</td>
<td>-</td>
<td>( \rightarrow )</td>
<td>You</td>
<td>Forming (or maintaining) the belief that ( q )</td>
</tr>
</tbody>
</table>

There is an epistemic reason for you to form (or maintain) the belief that \( q \).

We will later look at ways of trying to explain the structure of epistemic reasons without appealing to favouring relations at all, but for now, let us see whether the above construal seems plausible. And remember, even if it is the correct construal, we have not yet determined what kind of favouring relation that is involved, though we saw Cuneo indicating that it must at least have strong normative force.

In order to say something about the normative force of such reasons, let us first look at how we should understand the favouring relation involved. One way this might be done, is by claiming that all that is meant is that given that you believe \( p \) and also believe that if \( p, \text{ then } q \), believing that \( q \) should be done according to the norms of logic, or some similar correct-according-to relation. This is meant to be similar to the way that given the fact that placing your elbows on the table are disfavoured by the rules of etiquette, doing so is wrong according to the rules of etiquette. We allowed such according-to-relations to be reason-giving, although it is clear that they are not the provider of particularly weighty reasons.

\[\text{Cuneo categorizes such reasons as being } \text{categorical}, \text{ but as we have seen, that is not the best way to characterise this type of reasons. What is important rather, is their normative force.}\]
What makes a reason weighty? As we saw with moral reasons, weighty reasons are not grounded in according-to-relations, but instead needs a form relation of favouring relation which is capable of giving rise to strong normative force. Of course, it is unhelpful to define weighty in terms of strong normative force, unless more is said about what normative force is. In the preceding chapter we have found one way of cashing out this notion, and in fact, all the reasons with strong normative force that we have encountered so far have had such force in virtue of being irreducibly normative. We have therefore not yet found any other way for a reason to have strong normative force than for the favouring relation involved to be irreducibly normative.

Unless there are other options we have not discovered, the favouring relation involved in epistemic reasons must therefore be either irreducibly normative or a correct-according-to relation. Which of these best captures the favouring relation involved in epistemic reasons? Remember \textit{If } P, \textit{then } Q. Imagine that I were to tell you that since you believe that \textit{the lights are on}, and that \textit{if the lights are on, then John is home}, you should also believe that \textit{John is home}. Imagine further that you were to reply that you simply do not care whether you believe that \textit{John is home} or not, and, in fact, you would rather prefer to believe that \textit{John is not home}.

In such a situation, should we, as we often do with many of the rules of etiquette, shrug our shoulders and chalk it up to personal differences, or would you feel that there is something \textit{really} wrong with such a stance? It does seem like, unlike etiquette, you cannot simply opt out of the basic norms of logic. Before deciding this, let us make a clarification which could be though to interfere with our findings so far.

When considering what kind of favouring relations is involved in epistemic reasons, it is important to distinguish between different norms that might apply to an agent in any given case. For instance, it might be the case that an agent has strong non-epistemic reasons for believing \( p \), even if the agent has \textit{strong evidence against } \( p \). If, say, in order to succeed in some enterprise, one benefits from an unjustifiably high amount of self-confidence, it might be that, all things considered, the agent should believe without evidence.\footnote{See Feldman (1988: 235) for one such example.} In such cases, one might have certain non-epistemic reasons that override the epistemic ones, in the same way that some moral reasons can override hypothetical reasons. Take for instance a scenario where in order to save a drowning child nearby, you must risk tarnishing your newly bought shoes. You might care a whole lot about your shoes, but the moral requirement to save a drowning child would surely seem to trump those reasons.

That a type of reasons can be overridden by another type of reasons does not mean that the overridden type does not have strong normative force. Similarly one can hold that non-epistemic reasons (e.g. moral or prudential) might override epistemic ones, while at the same time holding that one cannot opt out of epistemic norms in the same way one can opt out of the rules of etiquette or a game.
To illustrate how this can clarify what kin of favouring relation is involved in epistemic reasons, let us look at an example from Peter Railton. He has suggested that we might not have any substantive reason to comply with logical “oughts” (Railton 1986). An agent does not have a reason to, say, sort out all his beliefs and make sure that there are no inconsistent beliefs in his belief set. As Railton puts it:

[It could be said that I logically ought not to believe both a proposition p and a proposition that implies not-p. However, it may not be the case that every rational agent will have an instrumental reason to purge all logical contradictions from his thought. It would require vast amounts of cogitation for anyone to test all of his existing beliefs for consistency, and to insure that every newly acquired belief preserves it. . .

[H]e may have no more reason to lay down his present concerns and wade in after [such a contradiction] than he has to leave his home in suburban New Jersey to hunt alligators in the Okefenokee on the off chance that he might one day find himself stranded and unarmed in the backwaters of southeast Georgia. (Railton 1986: 201–202)

In this quote, Railton seems to think of epistemic reasons as instrumental (i.e. hypothetical) reasons, and then it might of course be plausible that one does not have a reason to eliminate all contradictory beliefs. But as we have seen above, our conception of an epistemic reason does not seem to be that of a hypothetical reason, in that they are inescapable. A better gloss of the situation Railton discusses is therefore that there are epistemic reasons not to hold inconsistent beliefs and that these are not hypothetical. However, there might very well be hypothetical reasons that override the epistemic reasons.

That is, one should not, epistemically speaking, hold inconsistent beliefs and this gives agents a (non-hypothetical) reason to sort out their belief system. However, because of our cognitive constraints and so on, and because it would be massively taxing to sort through our beliefs, we have a hypothetical reasons not to sort out our beliefs in this way, which trumps the epistemic reason.

When considering epistemic reasons we should therefore make sure that there are no other norms interfering with our evaluation of the epistemic normative framework. Since we have operated on the assumption that epistemic reasons are not to be subsumed under hypothetical reasons, we must therefore be careful to separate out such non-epistemic reasons. This does not seem that hard, as we can ask whether you have a epistemic reason to avoid believing p and not-p even if believing this conjunction does not satisfy (or frustrate) any of your ends now or forever. Can you at will opt out of the epistemic norm not to believe contradictions, in the sense that you don’t really have a reason to follow it? Is there nothing really wrong with believing that something is completely blue and not blue, as long as this does not frustrate any of your ends? Is it the case that it is only wrong-according-to-the-norms-of-logic, but if you don’t particularly care about those norms, you are free to believe what you want without reproach.

Intuitions on this question might differ, but holding that from the epistemic point of view one has a decisive (i.e. weighty, normatively strong) reason to avoid such contradictions seems, at the very least,

\footnote{Again, if it turned out that epistemic reasons are hypothetical, then they would be irreducibly normative according to the arguments of the previous chapter.}

88
plausible. This can be taken to indicate that epistemic reasons do possess strong normative force. Furthermore, since the only way we currently have of cashing out this notion is irreducible normativity, epistemic reasons would therefore entail irreducible normativity.

Nonetheless, we should look at attempts that in various ways try to reduce epistemic reasons, either to epistemic support relations, hypothetical reasons or institutional reasons.

5.3 The possibility of reduction
I have now presented a line of reasoning intended to make it at least plausible that epistemic reasons are committed to irreducible normativity. But while there certainly are accounts of justification and of epistemic reasons which might best be characterized as done above, there are also both deontological and non-deontological accounts of justification which might be though to avoid the conception of epistemic reasons outlined above. For instance, there are accounts that attempt to reduce what it means for a belief to justified to something else. Might not such accounts be good candidates for giving an account of epistemic reasons that does not rely on their being irreducibly normative? In the next sections, I will look at such attempts to reduce epistemic reasons to something else.

5.3.1 Evidential support relations
One common way to account for the notion justification (either deontological or not) is to attempt to reduce it to something else. One candidate for what justification can be reduced to is some form of epistemic support relation. One view that attempts just such a reduction is evidentialism (e.g. Conee and Feldman 2004). A simple formulation of evidentialism is as follows:

\[
\text{evidentialism} \quad \text{Person } S \text{ is justified in believing proposition } p \text{ at time } t \text{ if and only if } S's \text{ evidence for } p \text{ at } t \text{ supports believing } p. \quad \text{(Mittag 2015)}
\]

According to evidentialism, an agent is justified in believing a proposition if and only if the agent’s evidence supports having the belief. An account of justification such as this engenders a number of questions: Which things can count as evidence? What is it for an agent to have evidence? What is the structure of evidential support? Despite this, I will limit my discussion to the question that is of primary interest in this chapter: What are we to understand by “support”? What does it mean for a body of evidence to support a belief?

In answering how we are to understand the nature of the evidential support relation, let us first assume a deontological understanding of evidentialism, on which justification is supposed to tell us something, if only indirectly, about what we should believe. If an epistemic support relation, such as the

\[136\text{ See the essays in Conee and Feldman (2004) for suggested answers to these questions.}\]
evidential one, is supposed to tell us what we should believe, it follows from Mackie’s platitude that it also tells us what we have a reason to believe.

If the epistemic support relations are supposed to give rise to epistemic reasons, it would seem that such accounts have simply pushed the pressing question one step further back. Instead of asking what epistemic justification is, and how it can be a guide to what we should believe, the question now becomes: In virtue of what does evidential support relations (which grounds epistemic justification) provide a reason to believe a proposition? To see that this leaves us in exactly the same spot as before, let us sketch an illustration of the relation that seems to hold between evidential support and a reason to believe a proposition in the Coffee cup scenario described above.

The conceptual structure of an epistemic support relation (deontological)

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Circumstances</th>
<th>Supports</th>
<th>Agent</th>
<th>Doxastic behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>You see a coffee cup</td>
<td>Awake, attentive etc.</td>
<td></td>
<td>You</td>
<td>Forming (or maintaining) belief in coffee cup.</td>
</tr>
</tbody>
</table>

An epistemic reason

You have a reason to form (or maintain) the belief that there is a coffee cup (on the side of the desk).

supported by evidence – it should be justified. If you believe something that is not supported by your evidence, then you are not justified in believing it. One way this is normatively guiding, is by holding that lack of evidence in support of a belief, disfavours that belief. This shows that even if a deontologically minded epistemologist were to reduce epistemic justification to some form of epistemic support relation, this support relation would still be irreducibly normative because it evidence is supposed to favour (or disfavour) agents adopting, maintaining or eliminating certain beliefs. As long as epistemic support relations like the evidential relation are taken to be normative in this sense, in that it gives rise to epistemic reasons, it seems that it must be committed to irreducible normativity.

Of course there is also the possibility of holding the normative favouring relation between the relata to only be a correct-according-to relation, where some belief is correct-according-to-the-evidence, but where this does not provide a weighty reason to form or maintain the belief. Such accounts therefore collapse with the general account of epistemic reasons given above. This would seem implausible for the same reasons as mentioned when we considered whether epistemic reasons have strong normative force. We will in any case return to this alternative later.

The other option is, of course, that we should not understand evidential support relations as a normative relation in this sense. I will pursue this argumentative strategy next.
5.3.2 Non-normative accounts of epistemic support relations
Let us now consider the possibility that we should adopt some form of non-deontological notion of justification, and that we should understand epistemic support relations as non-normative. Many have been sceptical to the possibility of understanding such relations without appealing to normativity in some way, but as I will argue, there does not seem to be a principled reason to such objections.

The general form of non-deontological accounts of justification might be construed as follows.

\[
\text{NON-DEONTOLOGICAL JUSTIFICATION} \quad S \text{ is justified in believing that } p \text{ if and only if } S \text{ believes that } p \text{ on a basis that properly probabilifies } S\text{'s belief that } p. \\
(\text{Steup 2005: §2.1})
\]

This then, says something about when an agent has a belief that is justified, where “justified” here only means that the agent believes that \( p \) on a basis that is sufficiently truth-conducive, where what truth-conducive means is cashed out differently by different views. For instance, one could say that it is simply a matter of probability, such that if the body of evidence the agent has available makes it, say, \textit{more likely than not} that the belief is true, \textit{then} the evidence supports it. This relation, in itself, does not seem metaphysically objectionable.\(^{137}\)

In order to get a better grip on what such a non-deontological account of justification would look like, let’s look at a specific example:

\[
\text{RELIABILISM} \quad \text{If } S\text{'s believing } p \text{ at } t \text{ results from a reliable cognitive belief-forming process (or set of processes), then } S\text{'s belief in } p \text{ at } t \text{ is justified. (Goldman 2015)}
\]

This then, is another form of epistemic support relation, that of \textit{reliability}.\(^{138}\) It might seem difficult to grasp how such an epistemic support relation as reliability, and the consequent justification it confers, is supposed to be, provide or indicate \textit{a reason to believe} something. One might think that this could be done by appealing to truth, since if one takes true beliefs to be an epistemic goal, as seems relatively plausible, then it would be instrumental to that goal to have justified beliefs – after all, why care about justified beliefs if they are not in some way meant to secure true beliefs.

\(^{137}\) There might be specific problems connected with the theory of probability, but in that case, one could use another notion of truth-conduciveness such as evidence, or as we shall see, reliability.

\(^{138}\) But to some extent, so does the original evidentialists (Conee and Feldman 2004). To them, the evidential support relation first and foremost obtain between a doxastic state \( D \), a proposition \( p \), and some body of evidence \( E \) (propositional justification), and an agent believing in \( p \) is then required to possess \( E \) for (doxastic) justification in believing that \( p \). An evidentialist can say: \( S\)’s belief in proposition \( p \) at \( t \) is justified iff \( S \) has evidence \( E \) at \( t \) that supports \( S\)’s belief in \( p \). Notice first of all that this principle talks about \textit{a belief} being justified, and not about an agent being justified \textit{in believing} something (cf. Littlejohn 2012: 5–6).
The reliabilist can therefore claim that the epistemic support relation in question have some connection to truth, and that it is in virtue of this that that such support relations provide us with reasons to believe, maintain or eliminate beliefs. After all, if we should believe anything, epistemically speaking, we should surely believe the truth. Epistemic support relations, such as reliability, probability or evidence, make the relevant belief likely to be true. They are, therefore, truth-conducive, and it is in virtue of this that they confer justification on our beliefs. The upshot of this picture is that the underlying support relations are (or at least can be) truth-conducive, and that is why they provide an agent who with reasons. For instance, if agent believes that \( p \) where \( p \) is not justified, it seems that it might be less likely that \( p \) is true. Therefore, if you have a belief \( p \) that is not justified, that is, it is not epistemically supported, this provides you with a reason not to believe it.

It is not difficult to see where this line of thinking leads; some feature, whether it is truth, truth-conduciveness or something else, seems to be, provide, or indicate reason for believing a proposition, partly in virtue of providing justification. This, however, takes us straight back to the previous analysis of epistemic reasons, where some feature, such as some other belief(s) an agent has, or that a proposition is true, or some other thing, favours an agent believing a particular proposition. If we were to take truth or truth-conduciveness to engender reasons to believe, it would therefore best be captured according to the already spelled out account above. It therefore seems hard to imagine in what sense an account of epistemic justification or epistemic support relations could both be truly non-deontological and hold that there is a normative relation involved in the epistemic support relation.

At this point, the proponent of non-deontological accounts of justification may simply claim that, well, maybe we should shed the reason-giving picture altogether, and give up the idea of epistemic support relations (or the consequent justification) giving rise to reasons for belief. In short, they might hold that the epistemic support relations appealed to by non-deontological accounts might tell us something about truth-conduciveness, and they might help us analyse knowledge, but they do not tell us what we have reason believe.

However, by going in this direction, it becomes clear that there does not seem to be any epistemic reasons. This non-deontological line of reasoning, then, seems to lead to an epistemic error theory all on its own. It seems to end up with the claim that tough we ordinarily believe there are epistemic reasons to believe, there really aren’t any. In an attempt to avoid this conclusion, let us look at some other possibilities for reducing epistemic reason.

5.3.3 Epistemic reasons as hypothetical reasons
As I mentioned earlier, some have argued that epistemic reasons are simply hypothetical reasons. I have argued above that this is not a plausible way to understand epistemic reasons, since they seem to obtain
independently of an agent’s end, and also seem to be committed to irreducible normativity. Furthermore, even if it were possible to reduce epistemic reasons to hypothetical reasons they would still, according to the analysis of the last chapter, be committed to irreducible normativity. As I argued there, all normative hypothetical reasons are committed to irreducible normativity. This then, seems like a doubly unattractive solution.

However, in the previous chapter, we saw that there was also another type of hypothetical reasons. These were admittedly not reasons for action as they did not say anything about what an agent should do. In this respect, they are more like motivational reasons, not in that they explain why we did what we did, but in that they are only saying something empirical about certain relations that, in fact, hold.

Even so, it might be that there is a comparable class of epistemic reasons. Olson (2014: 158–159) seems to argue that there is a legitimate use of such epistemic reasons, where these are simply claims to the effect that for some agent to believe that $p$ will, as a matter of fact, satisfy some end of that agent. Again, and for exactly the same reasons as was presented when I discussed empirical hypothetical reasons, such reasons would not be normative, in the sense of being prescriptive.

So, while this might legitimize talk of epistemic reasons which are not committed to irreducible normativity, it is important to notice that these sorts of epistemic reasons are the equivalents of empirical hypothetical reasons and motivational reasons, in terms of their non-normative nature.

5.3.4 Epistemic reasons as institutional reasons

Might it also be the case that there is a legitimate use of “reason” where epistemic reasons can be understood as institutional reasons? That is, could it be the case that some social, conventional or institutional roles or rules give rise to epistemic reasons of an institutional kind? It seems like it might. For instance, Olson argues that certain social institutions might give rise to epistemic reasons of an institutional sort.

One might … hold that since the goal of many intellectual endeavours, e.g. metaethics and metaepistemology, is to get at the truth, it is correct for people engaged in such activities to have true beliefs on the subject matter. (2014: 59)

Again, institutional reasons are fickle creatures, and they do not provide what Joyce calls “real reasons” (2001). However, they are normative, in the sense of being prescriptive. So if there is a legitimate use of epistemic reason talk which allows for them to be reducible to institutional reasons, that means that there are some normative epistemic reasons, although they are only institutional and therefore possess only weak normative force. Furthermore, as I have argued above, it is not seem like all epistemic reasons are of this type.
5.4 Summary
In this chapter, we have seen that in targeting irreducible normativity, moral error theory generalises to also target most epistemic reasons, though perhaps not all of them, and not even all normative epistemic reasons. We have found that epistemic reasons can be separated into at least three classes.

First, there is the paradigmatic form of epistemic reasons, where such reasons answer to certain platitudes about authority, such that they are prescriptive, inescapable and they arguably possess strong normative force. Such epistemic reasons are committed to irreducible normativity. Second, there is an interpretation of epistemic reasons where they are purely empirical claims, in parallel with empirical hypothetical reasons. This sense is indeed unproblematic, although, as we have seen, such reasons are not normative reasons – they don’t tell us what we should do. Third, there also seems to be epistemic reasons in an institutional sense, where certain norms and rules that govern enterprises such as science or research may be said to give rise to institutional reasons to comply with these norms and rules.

We can therefore conclude that some epistemic reasons do entail irreducibly normative favouring relations, while some do not.

6 Challenges and prospects for metanormative error theory

[O]f all the non-robust-realist metaethical and metanormative options, it is a global metanormative error theory that, in a way, I find most respectable. Such error-theorists do not kid themselves about the commitments of normative discourse, and they proceed to boldly follow the argument from these commitments and the belief that they are unsatisfied to its natural, if extreme, conclusion.

Enoch (2011: 117)

6.1 From moral to metanormative error theory
In the previous five chapters, we have witnessed why the most prominent arguments for moral error theory lead to a generalized, metanormative error theory. We have also seen two instances of what such a global, metanormative error theory targets, namely all practical and epistemic reasons with strong normative force. In this chapter, I will evaluate whether the move from a localized moral error theory to a metanormative error theory robs error theory of its plausibility, and whether it substantially changes the evaluation from Chapter 3, where we saw that a local, moral error theory was found to be highly plausible.

A metanormative error theory will face all the objections that the domain-specific error theories face, and so I start by looking at what these objections amount to. I then go on to say something about the relative plausibility of the views we have considered.
6.1.1 Practical error theory: consequences and objections

In Chapter 4, we saw that all normative hypothetical reasons will be targeted by a metanormative error theory. We also discussed empirical hypothetical reasons, but I argued that such reasons are not prescriptive in that they do not tell us what we should do and, consequently, they are not normative. The fact that all hypothetical reasons with strong normative force are committed to irreducible normativity further strengthens the assumption that it is this irreducible normativity that gives rise to strong normative force.

The dominant view of practical rationality, instrumentalism, holds that all practical reasons are normative hypothetical reasons. If instrumentalism is true, all practical reasons are therefore targeted by a metanormative error theory. Even if instrumentalism is not true, and there are other forms of practical reasons such as, say, categorical practical reasons, these too are very likely to be committed to irreducible normativity, though I will not defend this claim here.

Let us for the moment assume that instrumentalism is true and that all practical reasons are hypothetical reasons. Practical judgements, such as “I ought to get a glass of water”, entail that there are reasons to get a glass of water, and, as we have seen, in order to fulfil the role they play in ordinary discourse, such reasons would have to be irreducibly normative. Since the practical error theorist is committed to there being no irreducibly normative reasons, all our judgements about what we have reason to do would be untrue. Looking back to our definition of an error theory in Chapter 1, we can see that this would constitute an error theory about practical reasons.

**PRACTICAL ERROR THEORY** All practical judgements (i.e. judgements about what we have a normative reason to do) are untrue.\(^{139}\)

Such a view is certainly radical, though not unheard of. Elijah Millgram describes what he terms nihilism about practical reasoning as follows:

Nihilism about practical reasoning is the view that there are no legitimate forms of practical inference, and that consequently there is no such thing as practical reasoning: appearances notwithstanding, there is no mental activity that counts as figuring out what to do. (2001b: 3)

This description seems to be a little confused, as the error theorist (i.e. nihilist) does not have to deny that there is mental activity that counts as trying to figure out what to do – such mental activity is surely ubiquitous. What the error theorist does deny is that anyone gets it right, in the sense of coming up with an answer to the question of what actions one’s practical reasons favour. There might therefore very well be something worth calling practical reasoning, it’s just that such reasoning produces no truths about what one

\(^{139}\) There might arise problems of formulation for the forms of error theory I present in this chapter too, just as we saw with respect to moral error theory in Chapter 1. I will not spend time discussing these. In any case, applying the solution suggested there, and restricting these other error theories to positive practical, epistemic and normative judgements, should deal with most such problems.
Trying to figure out what to do, in the sense of what one should do, is something most people do all the time. More to the point is the following:

Everyone acknowledges that there are considerations that are presented as reasons for action but that, on second glance, aren’t really reasons after all . . . the nihilist is someone who thinks that all would-be reasons for action are like those. (Millgram 2001b: 3)

One might think that this alone is enough to be the death knell of metanormative error theory. But let us look closer at some of the objections against this view to see if they hold up.

6.1.1.1 The self-undermining objection and the inconsistent practitioner objection
As we shall see, certain types of objections are common against moral, practical, and epistemic error theory. In this section, I will present two of them. The first, which I will name the self-undermining objection, holds that error theories are unstable in that they, in some way or other, undermine themselves. We have already seen this objection with respect to moral error theory, in the form of the problem of formulation, and we will see that it also shows up with respect to both practical and epistemic error theory. The second common objection, which I will name the inconsistent practitioner objection, holds that an error theorist is guilty of some form of inconsistency with respect to theory and practice.

First, we will look at an instance of the self-undermining objection against practical error theory. The error theorist about practical reasoning might be thought to present her view as a view of practical reasoning. As such, it could be argued that she attempts to answer the question of which forms of practical reasoning are legitimate. In other words, the practical error theorist supposedly attempts to answer the question of how we should reason with respect to practical matters.

[S]uppose that the question of what the legitimate forms of practical reasoning are is itself a practical question (the question of how to think rationally); then if nihilism were true, there would be no good arguments for it. (Millgram 2001b: 3)

First of all, as we saw in Chapter 5, it seems like there are two questions in play here. The first is about which way of thinking would in some sense satisfy our ends, where this is a question of instrumental rationality and hypothetical reasons. Second, there is the question of how we should think from an epistemic point of view (from the vantage point of truth, etc.). The first of these surely is a practical question, and if the practical error theorist makes pronouncements about how we ought to think to best satisfy our ends, that would certainly be self-undermining. However, the practical error theorist does not tell anyone how they should think in order to satisfy their ends. Rather, she merely presents evidence for the view that there are no practical reasons. From this alone, there does not follow any practical imperatives of the form “Think like this!” or “Believe this!”141 While the error theorist presents the evidence that indicates the truth of her view,

---

140 There are important similarities between this point and the discussion of practical deliberation in taken up in section 6.3.1.
141 I discuss this point in more detail in the discussion of the objections to epistemic error theory in section 6.1.2.
she does not claim that it is a (irreducibly normative) reason to believe it, either on epistemic or instrumental grounds.

What about the claim that because the practical error theorist does not believe there to be any practical reasons, she cannot come up with any good arguments for the view? As I will explain in more detail in the objections to epistemic error theory, this rests on conflating epistemic support relations, such as an evidential relation, and reasons to believe. Evidential relations, as we previously have seen, can be entirely non-normative, in that something, e, constitutes evidence for a proposition, p, without thereby giving rise to reasons to believe that p. The error theorist can therefore present evidence which indicates the truth of her view. And, if one allows that one can construe arguments as a set of propositions, where some (i.e. the premises) indicate the truth of another (i.e. the conclusion), then she can full-well present arguments to the effect that her view is true (Olson 2014: 162–163).

Furthermore, if the arguments of the practical error theorist are good (in the sense of probably true), even the metanormative error theorist can claim that since truth is a goal inherent in academic research, researchers should accept epistemic error theory since this would be correct according to academic standards. This is an instance of the type of institutional epistemic reasons we looked at in Chapter 5. Such reasons are entirely legitimate to appeal to even for the metanormative error theorist.

In short, then, the practical error theorist can present at least one form of argument on behalf of moral error theory, namely arguments to the effect that the view is true. Of course, what the error theorist cannot claim is that this provides irreducibly normative reasons for believing the view.

Let us look at another, more practical argument against practical error theory, this one from Thomas Scanlon:

Genuine skepticism about reasons – skepticism about whether anything ever “counts in favour of” anything else in the sense typical of reasons . . . would be a very difficult position to hold. Perhaps one could hold such skepticism just about reasons for action, holding that although various states in fact move us to act, there is no sense to the question of when we have good reason for these actions. But even this strikes me as awkward and unstable. To hold it consistently one would need to regard all one’s actions as things that merely happen, and to abstain from taking at face value any thought about what could be said for or against performing them. (Scanlon 1998: 19–20)

I think Scanlon is exactly right that the way to cash out practical error theory is by holding that while various mental processes and states, including attempts at “figuring out” what to do, in fact lead us to act, there is no favouring of the irreducibly normative sort involved. Contra Scanlon, I would say that there is good sense to be made of when we have reasons for acting: never. At least not in the sense of having irreducibly normative reasons for acting.

---

142 As Olson (2014:163) points out, in order to determine whether a given argument is “strong” or “good” it might be necessary to invoke certain norms having to do with what counts as either a strong or a good argument. This, however, seems possible to do in terms of institutional reasons, where the strength of an argument might simplistically be thought to be closely related to how probable its premises makes its conclusion.
What about Scanlon’s claim that it would, as a matter of fact, be difficult to practice practical error theory in everyday life? Scanlon seems to think that it will be difficult to align one’s cognitive processes with the theoretical view of practical reasoning that such an error theory involves. That may be so, but consider visual illusions such as the Müller-Lyer illusion. You know that the lines are of equal length, but nonetheless, you cannot avoid seeing them as different in length. While I will not argue for this at length here, it seems entirely possible to claim to know that there are no practical reasons, in the sense of irreducibly normative reasons, while sometimes feeling or temporarily believing that there are. This similar to a vision scientist being fooled for a second by a suitably inventive version of the Müller-Lyer illusion.

In short, then, practical error theory seems radical, as it implies that there are no (irreducibly normative) reasons for action. As one writer points out, the view “suffers from a shortage of contemporary defenders” (Millgram 2001b: 3), but as I have noted earlier, error theorists who embrace the general form of the argument for error theory presented in Chapters 2 and 3 seem committed to this view. It therefore seems worthwhile to explore the prospects and challenges of practical error theory in more detail than has been done so far. This is not the place for such an undertaking, but hopefully it will follow as a natural development of error theorists (and others) realizing the inter-domain similarities regarding the normative.

In summary, I do not think that practical error theory deserves to be as easily dismissed as it often is. This trend substantiates Daly and Liggins’s claim that, “When error theories are mentioned, they are commonly dismissed out of hand” (2010: 210). As far as I can tell, there is no plausible version of either the self-undermining objection or the practical inconsistency objection that serves as the knock-down argument they both often are taken to be.

6.1.1.2 Debunking explanations and non-moral reasons

In order for the argument from queerness that we discussed in Chapter 3 to successfully generalize to other domains than the moral, it would be necessary to provide a debunking explanation of similar plausibility as the one presented for irreducible normativity in the moral case. I will not present such debunking explanations here, but say why I think it seems reasonable to expect such explanations to be forthcoming.

In Chapter 3, we looked into the general structure of accounts that try to explain how evolutionary, social and cultural pressures and the like have influenced both the way in which we make moral judgements and also the contents of those judgements. In the case of moral reasons, there have been put forth relatively detailed theories about the mechanisms producing and shaping such judgements, and what the benefits of such evolutionary adaptations would consist in. The question is therefore whether we should expect there to be similar accounts of why we feel the pull of practical and epistemic reasons in the way we do. This is not

---

143 Olson (2014) embraces the move from moral to metanormative error theory, and also accepts that (non-empirical) hypothetical reasons for action are committed to irreducible normativity, but he does not mention the implications this has for practical reasons or practical rationality more generally.
something I will venture into here, as it would require a large and interdisciplinary undertaking, but I will only register my optimism on such a project’s behalf by providing some speculative comments.

For agents to believe that they really have a reason to act on their ends, or to believe what they take the evidence to indicate as being true, seems, to put it mildly, evolutionary beneficial. Furthermore, that this is so appears to the agent as self-evidently true, and in need of no further evidence. In fact, practical and epistemic judgements seem to give rise to an even stronger appearance of self-evidence than moral judgements, which themselves have a long history of being claimed to be self-evidently true. What would be the consequences for an organism with a cognitive capacity as great as ours, who did not believe that they had powerful reasons to satisfy their own ends or to believe what is true? It seems reasonable to suppose that such a species would be at a significant disadvantage in the struggle for survival. Of course, it is conceivable that such a species could simply act on their desires, while not taking them to function in a normatively guiding or justifying way. However, as we saw in the moral case in Chapter 3, believing in irreducible normativity can bring its own set of benefits.

In short, I think there is no reason to deny that such a debunking explanation could be forthcoming with respect to other domains than the moral.

6.1.2 Epistemic error theory: consequences and objections

Let us now move on to consider the consequences that flow from our investigation of epistemic reasons in Chapter 5. There we saw that all epistemic reasons with strong normative force are committed to irreducible normativity. This again supports the view that it is such irreducible normativity that provides reasons with strong normative force. That all such epistemic reasons involve irreducible normativity leads to epistemic error theory, which holds that all (irreducibly normative) epistemic judgements, i.e. judgements concerning what we should believe, refrain from believing or maintain belief in, are untrue.

**EPISTEMIC ERROR THEORY**

| All (irreducibly normative) epistemic judgements are untrue. |

This is perhaps an even more radical view than either moral or practical error theory. As we saw, practical error theory is often dismissed as being self-refuting. This is even more often the case with epistemic error theory, as we will see below. In answering the following objections to epistemic error theory, I will in large part rely on the responses put forth by Olson (2014: Chapter 8).

---

144 Compare the discussion in Chapter 4 of the insistence of instrumentalists that something like desire-satisfaction just is what we have reason to pursue.
6.1.2.1 Cuneo’s three undesirable consequences of epistemic error theory

The similarities between moral and epistemic reasons have not gone unnoticed, and they have especially been employed by moral realists who try to use them in companion in guilt-style arguments against error theory and more broadly, moral anti-realism in general.

In The Normative Web (2007), Terence Cuneo explores the similarities and analogies between moral facts, reasons, and norms and their epistemic counterparts. His core argument is that they are sufficiently similar so that any form of antirealism about one of the two domains will also entail antirealism about the other. And since antirealism about the epistemic domain (i.e. about epistemic reasons) is highly implausible, so is moral antirealism.

Cuneo argues that epistemic reasons are committed to irreducible normativity, and his goal is to show that whatever the costs of such commitments, the alternatives are much worse. In the course of arguing for the existence of epistemic facts, he therefore wishes to show that the alternatives to epistemic realism have deeply problematic commitments, and as for epistemic error theory, Cuneo claims it is “deeply unattractive because it implies what I call ‘the three undesirable results’” (2007: 117). Cuneo thinks each of these three consequences is individually sufficient for rejecting epistemic error theory. In the following, I will consider these three results, and one line of reply to each of them.

The first undesirable result: The self-undermining objection

Cuneo’s first undesirable consequence of epistemic error theory is a form of the self-undermining objection introduced previously, as it seeks to show that such error theories are self-undermining, in either of two ways. Cuneo presents this objection as a dilemma that faces any epistemic error theorist: Either an epistemic error theorist must claim that there are epistemic reasons to believe the view she is championing, or she must claim that there are no such reasons in favour of believing the view.

If the epistemic error theorist claims that there are epistemic reasons to believe her view, that is, reasons to believe it “from the perspective of representing reality aright, gaining true belief, knowledge, understanding, or the like” (Cuneo 2007: 117), she is impaled on the first horn of the dilemma. Since epistemic error theory claims, by definition, that there are no such epistemic reasons, such a claim would make the position self-defeating.

If the epistemic error theorist instead opts for the second horn of the dilemma, choosing to claim that there are no epistemic reasons to believe the error theory then a rational person can safely ignore the view, as he is under no rational pressure to take heed of it. Either way then, the theory undermines its own purpose, either by being self-defeating or by giving an opponent of the view no reason to believe it, thereby making it “polemically toothless”.

145 Cuneo calls the view epistemic nihilism, but calling it epistemic error theory seems just as, if not more, apt.
As we have seen, the epistemic error theorist’s thesis is that there are no irreducibly normative epistemic reasons, and consequently that no judgments about such reasons are true, so it will obviously not be possible for the error theorist to point to such reasons in favour of the view. This, however, does not mean that the error theorist cannot produce anything that might make an opponent seriously consider the view.

As we saw in the first objection to practical error theory, what is needed in order to answer the self-undermining objection is to draw on the distinction that was introduced in the previous chapter. We saw that it is possible to separate epistemic support relations from the deontological and normative component which is involved in many theories of justification. In short, then, it is possible to hold that epistemic support relations, such as evidential or reliability relations, hold while denying that such relations give any agent a reason to believe a particular proposition, in the sense of an irreducibly normative. One simple way of cashing out this distinction is by employing a notion of evidence that is entirely non-normative, and which does not have any links to justification in the deontological sense. On such an account, \( e \) can be evidence for \( p \) in virtue of \( e \) making it highly probable that \( p \). Olson summarizes the strategy thus:

The error theorist avoids the first horn of the dilemma by distinguishing between arguments to the effect that \( p \) is true, and arguments to the effect that there are reasons to believe that \( p \). (Olson 2014)

The error theorist can therefore claim to be presenting evidence or even arguments on behalf of epistemic error theory, in the sense of presenting evidence or arguments to the effect that the view is true. Thus it is possible for the error theorist to claim that there is good evidence that the view she is championing is true, although she cannot claim that her opponents should believe it for that reason, at least not unless this is only meant in the sense that academic research has a norm for seeking truth, institutionally speaking. In this way, epistemic error theorists can align themselves with non-deontological conceptions of justification, which do not necessarily engender irreducibly normative epistemic reasons.

The error theorist therefore agrees that there are no irreducibly normative epistemic reasons to believe the view, and this exposes the error theorist to the second horn of the dilemma. By accepting that the error theorist can provide no reason to believe the theory, the position might be thought to become polemically toothless. But this is a consequence the error theorist can accept,

since error theorists are not in the business of offering arguments about what would be rational to believe or about what there is epistemic reason to believe. (Olson 2014: 158)

Rather, as in the case of practical error theory, the error theorist is only claiming that the theory holds a high probability of being true, or that it is plausible in some other (non-normative) sense. Olson (2014: 158) further argues that while the error theorist is polemically toothless in the sense of not even participating in debates over what it is rational to do, this is not the case across the board. The error theorist can claim to be engaging in discussions over which metaethical or metanormative views are true, although remaining silent about which of them we should believe. This, then, does not appear to be a weighty objection. Really, it is
just a way of underlining the commitments of an epistemic error theory, and does not seem to add any additional challenges for it.

The second undesirable result: The objection from lack of epistemic merits

According to Cuneo, if one were to accept that there are no epistemic facts or reasons of the type discussed in the last chapter, it follows that

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

(2007: 119)

This is the basis of the second undesirable result. However, as Olson (2014: 164–5) points out, and as we saw in Chapter 5, there is a form of normative epistemic reasons which the error theorist can make use of, namely the epistemic form of institutional reasons. As such, it seems open to the error theorist to hold a belief to be faulty or praiseworthy according to some set of rules or norms. For instance, Olson suggests there might be a notion of a “responsible believer”, and that certain things are compatible with being such a believer, while some are not. Among the things which are not compatible is to believe in things for which you yourself do not think there is good evidence.

Cuneo further argues that the error theorist is lead to accept the following “Moorean paradox” (2007: 43):

\[ EMP \quad \text{Epistemic error theory is true, but there is no reason to believe it.} \]

Cuneo points out that while such a proposition might seem paradoxical, that is not how it is interpreted by the error theorist. Indeed, the error theorist is committed to holding that there is nothing paradoxical about it, since she is committed to both of the claims involved.

However, an epistemic error theorist might have the resources to show that there is something strange about such a claim. First of all, we can point out that a person who subscribes to this proposition either takes herself to have sufficient evidence to the effect that the view is true, or she does not. If she does, the paradoxical nature of the proposition is only apparent, as the person who subscribes to EMP is in essence saying:

\[ EMP^* \quad \text{Epistemic error theory is true, and I believe there to be good evidence to the effect that error theory is true, and I believe the view on those grounds (but I do not believe that there are irreducibly normative reasons for believing it).} \]
However, the most natural interpretation of EMP, and which might give it an air of paradox, is that it seems to expresses a second possibility, that the subscriber does not take there to be sufficient evidence to the effect that epistemic error theory is true. We therefore get:

\[ EMP^{**} \quad \text{Epistemic error theory is true, but I do not believe there to be sufficient evidence to the effect that error theory is true, and so I don’t believe the view on the basis of evidence (and I do not believe that there are irreducibly normative reasons for believing it).} \]

In this second case, the subscriber would run afoul of the requirements of being a responsible believer, which includes believing something without sufficient evidence (Olson 2014: 166). Of course, as the pressure to be a responsible believer is no greater than the pressure to comply with any other institutional rules, the error theorist is not in a position to claim that the subscriber to EMP** should change her mind. Nonetheless, it seems like the error theorist has the resources necessary to explain why EMP has “an air of paradox” while, at a closer look, not really being paradoxical.

To return to the point mentioned earlier, Cuneo takes epistemic error theory to imply a “sweeping form of epistemological skepticism” (2007: 119), in the sense that without epistemic merits such as justification and the like, we cannot have any justified beliefs. But is this necessarily so? As we have seen earlier, contemporary epistemology operates with multiple distinct notions of justification. One of these, the deontological and irreducibly normative kind of justification, error theorists must hold to be non-instantiated. Even so, it is open to the error theorist to claim that our beliefs can be justified according to the non-deontological conception of justification, as long as it is not otherwise committed to irreducible normativity. From this it follows that while the error theorist cannot claim that one has (irreducibly normative) epistemic justification to believe a proposition, there might be other forms of justification and reducibly normative epistemic reasons that obtains.

This second objection to epistemic error theory again seems to highlight the consequences of the view, but without it being clear that it adds anything objectionable to the theory beyond that.

The third undesirable result: The self-undermining objection revisited

In his third objection to epistemic error theory, Cuneo goes on to argue that it is an undesirable consequence of epistemic error theory that it makes arguments impossible. The reasons for this are as follows. Cuneo begins by taking the paradigmatic form of arguments to be such that “the premises are offered in support of

---

146 The requirements for being a responsible believer are likely far more complex, but I only mean to illustrate the general point.

147 This is also argued by Richard Rowland (2013), who argues that moral error theory entails that there are no epistemic reasons and that if there are no epistemic reasons, then no one knows anything. However, this argument presupposes that the kind of epistemic justification needed for knowledge is of an irreducibly normative kind. As we have seen, there are alternative notion of epistemic justification. In order to make the argument go through it would therefore have to be established that knowledge requires an irreducibly normative form of epistemic justification.
its conclusion in the sense of providing evidential support for its conclusion” (2007: 121). He further goes on to claim that a statement being offered as evidential support *just is* for that statement to be offered *as a reason* for accepting that conclusion. Therefore, if there are no reasons, in the relevant epistemic sense, there can be no evidential support. And, without evidential support, there can be no arguments of the paradigmatic form. From the fact that there can be no such arguments, it follows that there can be no non-question-begging arguments either for or against epistemic error theory, since such arguments would presuppose the existence of just the type of reasons which the epistemic error theorist denies.

If the error theorist was to go along with this, she would be committed to the claim that there could not be an argument for anything. This gives us what Cuneo takes to be the third undesirable result of epistemic error theory.

At this point, it should be clear how the error theorist should answer this objection, namely by answering along the lines of the reply to the first undesirable result and the first objection we looked at for practical error theory. As mentioned, Cuneo thinks of evidential support relations as inherently normative, and that one cannot separate out a non-normative notion of evidence or other forms of epistemic support.\(^{148}\) However, I hope to have shown that this is very much an option, and that there is a perfectly acceptable notion of evidence which is simply some form of probabilization.

With such a non-normative notion of evidence, it is possible to claim that epistemic error theory can very well account for the existence of arguments, both in general and for and against the theory.

Again, this objection does not seem to provide any independent evidence to reject epistemic error theory, but only serves to highlight its commitments. If we therefore buy epistemic error theory on other grounds, and are aware of its commitments, Cuneo’s three “undesirable results” should not be particularly alarming.

6.1.2.2 Epistemic error theory and the normativity of belief

Epistemic error theory holds that there are no irreducibly normative reasons to believe a proposition. If there was a way to show that reasons to believe were in some way indispensable, or at least that giving them up was theoretically costly, that would help secure the claim that there are, in fact, such reasons. One way in which this has been argued for is through the claim that it is an essential part of what it is to *ascribe a belief* to an agent that one also imputes to that agent certain epistemic reasons for belief.

The idea is therefore that it is *constitutive* of beliefs ascriptions that they impute an agent with reasons to believe certain propositions (and to not believe others). If the reasons that are additionally ascribed

\(^{148}\) He is not alone. In “What is ‘Naturalized Epistemology’?”, Jaegwon Kim (1988: 390–91) writes: “In any event, the concept of evidence is inseparable from that of justification. When we talk of ‘evidence’ in an epistemological sense we are talking about justification: one thing is ‘evidence’ for another just in case the first tends to enhance the reasonableness or justification of the second. … A strictly nonnormative concept of evidence is not our concept of evidence; it is something that we do not understand.” While the error theorist might admit that in ordinary discourse there is a *pragmatic* assumption that evidence just *is* a reason to believe, there does not seem to be any particular problem involved in separating these two notions, as we saw in Chapter 5.
to the agent are *irreducibly normative reasons*, then it seems we cannot get rid of talk of irreducible normative reasons without also getting rid of ascriptions of belief. That we can no longer legitimately ascribe beliefs would be a large cost for the error theorist.

To see why some have claimed that belief ascriptions impute reasons to believe, consider an ascription such as “A believes that $p$”. What does this say about A? It is relatively uncontroversial that it says something along the lines of *A takes it to be the case that $p$*. That however, seems also to be true of an agent who assumes or supposes that $p$. What is the difference between these propositional attitudes toward $p$?

Those who believe that belief-ascriptions are constitutively normative take belief-ascriptions along the lines of “A believes that $p$” to not only say about A that he takes it to be the case that $p$, but also that, in virtue of it being a *belief*,

A is subject to certain norms to the effect that there is reason for A, for example, not to believe simultaneously that not-$p$; that there is reason for A not to disbelieve anything that she believes follows logically from $p$; that there is reason for A no longer to believe that $p$ in the face of strong evidence that not-$p$; and so on. (Olson 2014: 167)

Olson suggest that the simplest answer on behalf of an epistemic error theorist is to *grant* the claim that belief ascriptions do implicitly ascribe reasons to believe in the ways outlined in the quote above. *However*, the error theorist cannot, of course, grant that the kind of norms that are constitutive of belief ascriptions engender *irreducibly normative* epistemic reasons. Rather, the error theorist can hold that such norms simply provide the conditions for whether something *is* a belief or not. This is familiar from other sets of norms and rules. Take the rules of chess. To a certain extent, they specify what you must do *in order to be playing chess at all*. In other words, such rules are *constitutive* of the practice they govern. To see this, imagine that after having set up the chessboard correctly, you and your opponent start moving the pieces around willy-nilly with no regard for the actual rules of the game. Then it would seem that you are no longer playing chess at all, though there is probably not sharp lines here.

Olson (2014: 168) suggests we see belief in a similar fashion, where we take the norms outlined above to be either constitutive or regulative, such that they specify either what you must do *if you are to be considered as conforming to the role in question*. The norms of belief outlined above would therefore either specify what a believer must adhere to in order to be considered a believer at all. Or, alternatively, some of them only specify what a believer must adhere to if he is to be considered a competent or a responsible believer.

As should be clear, this answer tries to cash out the epistemic reasons involved in belief ascription as institutional epistemic reasons. These, as we have seen, need not worry the error theorist.
6.1.3 Metanormative error theory: consequences and objections

As we have seen, the most prominent and, arguably, the most plausible argument on behalf of moral error theory leads to a global, metanormative error theory. Such a theory holds that all judgements about irreducible normativity, in the sense of judging something to irreducibly favour something else, are untrue. We therefore get the following definition of metanormative error theory:

**Metanormative Error Theory**

All (irreducibly) normative judgements are untrue.

We can now see how any metanormative error theorist is a moral, practical and epistemic error theorist by implication, at least concerning the judgements about irreducible normativity involved in these discourses.

Having now looked at some individual objections that have targeted one or the other of these local error theories, I will now discuss one argument directed directly at the metanormative error theorist.

6.1.3.1 The argument from the practical indispensability of deliberation

As we saw in Chapter 3, Enoch (2011) has presented an argument meant to show that normative properties are indispensable. Often indispensability arguments spring out of explanatory indispensability, where the thing that is claimed to be indispensable provides some form of explanatory value. Enoch is clear that his argument from deliberative indispensability does not provide any such explanatory value. However, as we saw in Chapter 3, there could conceivably be other forms of indispensability. And, as I will return to below, it seems we must have some principled way of determining which forms of indispensability arguments licence ontological conclusions about what exists, and which do not.

Furthermore, as I mentioned in Chapter 3, this argument does not target moral properties in and of themselves, and as long as an error theory is restricted to only the moral domain, it can bypass Enoch’s indispensability argument. However, since we have seen that many of the most prominent forms of moral error theory do generalize to the entire irreducibly normative domain, it becomes pressing to take a closer look at the argument.

The argument, in short, goes like this. Deliberation about what we should do is ubiquitous in daily life. We weigh and consider whether we should do *this*, or rather do *that* instead, and we often consider many things when doing so, such as what we want to do, what we should do, what we are expected to do, and so on. However, what underlies such deliberations, according to Enoch – what we are trying to discover – is what we should do. The “should” involved is irreducibly normative and, Enoch believes, in deliberating, we are committed either implicitly or explicitly to the existence of irreducibly normative reasons.\(^{149}\) This is because, in deliberating, we are doing the practical equivalent of attempting to discover the answer to a factual question: What do we have most (irreducibly normative) reason to do.

\(^{149}\) The commitment can be of the form that one implicitly is committed to their existence without being able to explicitly recognize this. Enoch’s sense of commitment is even compatible with the agent explicitly denying a commitment to such reasons (Enoch 2011: 74; cf. Olson 2014: 172–173).
Now, it would be possible to take objection to the way Enoch spells out the everyday notion of deliberation. He holds it to be a question of irreducibly normative reason, but it seems that there is also something to be said for the idea that, at least in many instances, deliberation is just as much a question of what we want or desire. Olson (2014: 173–174) argues that deliberation might often be understood in this way, and that such a picture might not be in conflict of our first-person experience of deliberation. However, to such a reply Enoch could point out that hypothetical reasons, at least when understood as reasons for action and not simply empirical claims, are themselves committed to irreducible normativity. If the idea is that deliberation can be grounded in what would serve the ends of the agent, it does not necessarily remove the irreducibly normative element from deliberation. If the agent wonders “What should I do?” in the sense of “Given my ends, what do I have most reason to do?”, the commitment is still intact.

Given this, let us grant Enoch that in deliberating, we are committed to the existence of irreducibly normative reasons. A more direct reply, which Enoch himself considers, is the following. If one in deliberating is trying to find out what one has irreducibly normative reason to do, then this is not unlike the scenarios we encountered in the previous chapters, where we have seen just how common the search for irreducibly normative reasons in fact is. Ethicists do it, theoreticians of practical reasons do it, and the epistemologist often does it.

It might therefore seem like the error theorist should simply reply the same way to Enoch’s objection as we did then: True, it is often the case that people assume that there are irreducibly normative reasons, and this is essential to many of our practices. However, as I argued in Chapter 3, it is highly plausible that there are no such reasons. There are a number of objectionable aspects to them, and even though they are central in our pre-theoretical thought and talk, it seems like this cannot justify believing in them when we seem to have a more plausible account which can both explain all our practices, and does not need to postulate the entities in question. The upshot is that we do engage in these practices, so they are obviously possible. Despite this, we never succeed in finding out what we should do or believe, since there are no irreducible reasons. As such, all these practices and discourses are shot through with error. This reply seems to have been sufficient in the case of moral, practical and epistemic reasons – so what is so different about deliberation?

Enoch argues, roughly, that there are reasons to think that if irreducible normative properties are indispensable for (successful) deliberation, and (successful) deliberation really is something we want to allow for the possibility of, then it seems that we are allowed to take on an ontological commitment to irreducibly normative reasons on the grounds of its indispensability to deliberation.

It is common to allow that explanatory indispensability licenses ontological commitment, and Enoch argues that, unless there is principled way “of drawing the distinction between types of indispensability that can ground ontological commitment, and those that do not”, we have no reason to think that indispensability for (successful) deliberation cannot earn normative properties ontological rights (2011: 66–67). In response,
Olson (2014: 176) has suggested that there is a natural way to draw just such a distinction: only those indispensability arguments that are truth-tracking license ontological commitments. Arguments from explanatory indispensability work because the world is “explanation friendly”, that is, the world is such that the best explanations are also the true ones.\footnote{One might be sceptical of such explanatory indispensability arguments too, but that is not what is at issue here, since both the error theorist and Enoch accepts them.}

We have good reason to think that the world is explanation friendly, because inferences to the best explanation are successful; they work, both in science and everyday life, and the best explanation of this success seems to be that inferences to the best explanation are truth-tracking. Do we have good reason to suppose that the world is “deliberation friendly” in the sense that would make arguments from deliberative indispensability truth tracking? There is no parallel to the success of inferences to the best explanation for deliberative indispensability. Moreover, the argument presented in Chapter 3 seems to give us good reason to think that practical deliberation, if it is committed to irreducible normativity, is not truth-tracking, because irreducible normativity is metaphysically queer. There we looked at an argument against there being irreducibly normative reasons, and found it to be highly plausible. There, we accounted for the costs involved in rendering some common and strongly held beliefs false, and since we are not only rendering moral judgements wrong, but also all instances of deliberation, this is a cost that must be taken into account. But as far as I can tell, it does not seem like the calculation of plausibility points should come out that differently. It therefore seems as if the argument from queerness still has bite, and should serve to undermine our faith in the world being “deliberation-friendly”.

But as we saw in Chapter 3, Enoch thinks we should have a very high confidence in the truth of basic moral claims, and he thinks likewise of successful deliberation:

It is still possible, of course, that deliberation is illusory, that it essentially relies on a false belief in normative truths. But first, we would need a very strong argument to believe that. (Enoch 2011: 79)

A very strong argument against irreducibly normative reasons is what I attempted to provide in Chapter 3, in the form of an argument from queerness. At this point then, the question is if when all is said and done, we should believe in the existence of irreducibly normative properties. With respect to that question, all the indispensability argument does is to point out another thing that would be targeted by a metanormative error theory. That such a metanormative error theory would do away with much that we hold dear can of course be added to the costs of the view, as we also saw in the case of the Simple Moorean Argument in Chapter 3. This, however, is no knockdown argument against a metanormative error theory. And as such, the possibility that deliberation is illusory is simply one more thing that has to be factored into the “all things considered” judgement that theory choice involves.
6.2 Counting the costs, weighing the benefits
We have now looked at a number of objections that have been levelled at various specific error theories, and also at a general metanormative view. We have seen that the error theorist has the resources to answer all these objections. The objections nonetheless serve to underline the costs and commitments that follow from embracing a metanormative error theory, although they do not seem to add any costs over and above the costs that are involved in rejecting irreducible normativity in the first place. That is, these arguments have not shown that there is some deep flaw in metanormative error theories, say, that they are incoherent, self-undermining or that a metanormative error theorist of necessity must be inconsistent in thought and action.

As metanormative error theory is not all that well explored, it might very well be that there are other objections to the view that I have not considered here. That will be one avenue for further research.

6.2.1 Metanormative error theory and metanormative non-naturalist realism
In the preceding chapters, I have tried to push the following narrative: When it comes to many important discourses – at the very least moral, practical and epistemic discourses – it seems that we are restricted in terms of the viable theoretical options available. These discourses certainly are, as Cuneo puts it, “realistic on the face of it” (2007: 11–12). And furthermore, the realistic commitments of these discourses seem robust, in that a naturalistic account that satisfies all the commitments seems unlikely to be forthcoming. For these reasons, when it comes to accounting for these discourses, it appears difficult to go in a direction other than either an error theory or a (robust) realist position. Furthermore, because of the similarities between these discourses, there is pressure to provide a unified, metanormative account of them. This, in turn, means that there is pressure to move from being either a non-naturalist or an error theorist about one of these discourses, to generalize the view to all of them.

There is also an internal pressure for such generalization in the case of moral error theory (and perhaps also from within non-naturalism). This is because moral error theory, in its most common forms, generalizes to the entire irreducibly normative domain. Again, this move from a local, metaethical theory to a global metanormative view means that we will have to do a recount of plausibility points in order to determine whether metanormative error theory is equally plausible to what we found moral error theory to be in Chapter 3.

Doing a full recount is more than I have the space for here as that would involve running an argument from queerness against each of the other domains and show that it generalizes successfully. This requires, among other things, that the debunking explanation is equally applicable to these other normative domains. Although we are concerned with the same phenomenon in all domains, namely irreducibly normative favouring relations, there are differences between these domains, and that would have to be accounted for. But if we assume – though I will not argue for it – that the argument from queerness is at least equally strong in the case of the other domains we have looked at, what are the results likely to be?
The costs of a metanormative error theory should by now be obvious: There are no irreducibly normative truths, and thus no irreducibly normative reasons for action, belief or moral behaviour. Furthermore, deliberation, in the sense sketched above, would be doomed to failure. This is quite an overturning of common sense opinion. Now, if there were other reasons for action, belief and moral behaviour which both had strong normative force and was not committed to irreducible normativity, this would arguably not be that big of a cost. The error theorist could then, like the naturalist, claim that while we might not get everything we had hoped for, we might still get all we need. The metanormative error theorist, however, provides no substitute. Of course, there are true claims about empirical hypothetical reasons, and various institutional reasons. But again, none of these reasons are claimed to have strong normative force, which has been the bread and butter of normative reasons, naturalistic or not.

The cause for this, as I have tried to show inductively in the preceding chapters, it that wherever we find strong normative force, we also find irreducible normativity. The upshot is therefore that there are no real reasons, in the sense of reasons with strong normative force according to metanormative error theory. This, on any way of counting, is an enormous cost.

As for the benefits of the view, they are, first, the ability to take our thought and talk at face value, and to provide a believable and adequate account of the commitments various discourses possesses. Additionally, the view creates no additional implausible commitments, epistemological, ontological, or otherwise. It is “the most minimal” metanormative position and, at least in terms of theoretical commitments, it is arguably the “null hypothesis against which other accounts of practical reasoning must be defended” (Millgram 2001: 3).\footnote{Millgram talks only about practical error theory, but the point generalizes. Defending a position on exactly where the burden of proof is located in a metaethical or metanormative debate is controversial and not something I will dwell on here.}

As for non-naturalist realism, it stands at the opposite end of the minimal-maximal spectrum. While error theorists and non-naturalists usually agree on the commitments of normative discourses, they disagree about whether they are satisfied. The non-naturalist thinks we must mould our ontology to fit those commitments, for instance because of their indispensability to the success of certain fundamental human practices. The costs and benefits of the non-naturalist picture are therefore the inverse of those of metanormative error theory.

In short, these views, while descriptively adequate, are both enormously costly to adopt. As for relative plausibility, this becomes harder to measure when we consider the metanormative and not only the metaethical versions of these views. I will therefore restrict myself to pointing out that these views seem fairly comparable in terms of plausibility – it is not clear that one is significantly preferable to the other, though I think the ultimate outcome will be a result of whether one places the highest credence in pre-
theoretical opinions or in the argument from queerness. I incline towards the latter, and therefore see a metanormative error theory as marginally more plausible than metanormative non-naturalist realism.

6.2.2 Other views
The picture that emerges is therefore that of two descriptively adequate, but enormously costly views where it is surprisingly difficult to assign a significant advantage to either view. As we have seen, much hinges on the prior credence one gives to our ordinary pre-theoretical experience of the world as a place that is deliberation-friendly and which is imbued with substantial reasons. The non-naturalist places such a high credence in these things that it raises the bar for what it takes to disprove them very high – arguably, too high.

At this point the metanormative theorist might begin scouting for less costly alternatives, such as various forms of metanormative non-cognitivism or naturalism (in the sense of a metanormative equivalent of moral naturalism). Such views try to lessen the commitments of ordinary discourses in order to more easily fulfil them. If this could be made to work, they could quickly get the upper hand on both metanormative error theory and non-naturalist realism. However, and although I have not discussed such views at any length, it seems clear that they provide a less adequate description of ordinary discourse. This might not be seen as a problem, and the non-cognitivist or moral naturalist might even label their views revisionary. This, however, runs the risk of leaving ordinary discourse in no less error than the error theorist does. After all, if ordinary discourse gets the actual nature of some phenomenon entirely wrong, their judgements about the relevant phenomenon might all be untrue, leaving the discourse in question factually defective.

Since I have not discussed such views at any length, I can only note that though I am not optimistic about the capability of such views providing a descriptively adequate account of moral discourse, I agree with Enoch that we cannot exclude the possibility that there will one day be a position “whose theoretical value will be worth its cost” (Enoch 2007: 109).

6.2.3 Conclusion
Moral error theory – the metaethical view – has long had an image problem. As Lillehammer points out, before (and sometimes also after) the publication of Mackie’s *Ethics: Inventing Right and Wrong*, error theoretic views were “rarely seen on the big philosophical stage, making only brief appearances, and then mainly in caricature” (Lillehammer 2010). In the decades since Mackie’s book, this view has slowly gained both popularity and respect, and is now seen by even some non-naturalist realists as both a plausible view and one that demands to be taken seriously. In Chapter 2 and 3, I tried to support such a view, by showing that an error theory confined to the moral domain seems highly plausible.

Along with the increased appreciation of the similarities between different normative domains, there has been a dawning realization that the motivation and arguments on behalf of moral error theory cannot be
contained to the moral domain, but that they spread to the whole of the irreducibly normative, and that the view therefore transforms into a metanormative error theory. This metanormative view – metanormative error theory – has received something of the same reception as moral error theory did in its early years: It is usually taken to be self-undermining, impossible to hold in practice, or otherwise not able to even get off the ground. My main goal in the preceding has been to show that such a view deserves to be taken seriously, partly because it shares many of the virtues of moral error theory; although it also inherits its vices. To my eyes, it appears to be one of the most formidable and attractive metanormative views currently available, and if nothing else, it deserves serious scrutiny and not casual dismissal.

Appendix: Moral naturalism and the conceptual claim

Moral naturalists hold that there exist “objective moral facts and properties and [that] these moral facts and properties are natural facts and properties” (Lenman 2006). Error theorists usually hold that all such naturalist theories of morality fail to give a satisfying account of moral discourse. How could one go about arguing for such a strong claim? Any account that seeks to adequately account for ordinary moral discourse, and especially one that seeks to save such discourse from systematic error, needs to respect certain platitudes about what the participants of that discourse take themselves to be doing. That is, there are certain criteria of adequacy for any of morality.

What are those criteria? In order to uncover them, a good place to begin is Charles Stevenson’s suggestion for a naturalist theory: “Something is morally good if and only if it is pink with yellow trimmings” (1937: 14, cited in Joyce 2006: ). What his playful quip illustrates is that moral naturalism has to balance two considerations that pull in opposite directions. First, it has to make sure that the commitments it attributes to ordinary moral discourse are naturalistically kosher (as are yellow things with pink trimmings). If they are not, it fails to be a form of naturalism. This consideration puts pressure on the moral naturalist to not employ conceptual frameworks which engender non-naturalistic features. For this reason, they cannot avail themselves of the favouring relation framework as worked out above. On the other hand, such theories must manage to respect platitudes about morality and moral discourse. If moral discourse, on the moral naturalist’s interpretation, looks nothing like what the participants take it to be, as is the case in Stevenson’s quip, then either those participants are thoroughly in error about what they are doing, or the moral naturalism fails to be a theory of morality (as Stevenson’s theory clearly does). In short, if the moral naturalist wishes to

---

152 One strategy I will not pursue is the so-called Open Question Argument against moral naturalism, mostly because I think it is more of a surmountable challenge than a knockdown argument (cf. Feldman 2005; Smith 1994: Chapter 2).
avoid error theory or irrelevance, she must balance the requirements of naturalism with a respect for the plaititudes of ordinary moral discourse.

How are we to evaluate the prospects of this balancing act? As Joyce (2006: 191) points out, we do not necessarily demand a “perfect fit” between a proposed naturalist theory and our pre-theoretic understanding of our moral practice. It is therefore open-ended just how well a naturalist theory must manage to capture its target domain in order to be considered successful:

It would be a mistake to get hung up on [the number of individual platitudes a naturalist theory manages to capture], or to think that some general “golden ratio” of what constitutes enough can be supplied. Perhaps for some concepts there is a platitude that we consider of central importance, such that failure to underwrite this single criterion is sufficient to warrant rejection. Perhaps for some concepts we will be willing to have more relaxed standards than for other. (Joyce 2006: 191)

One way to try to thwart any attempt at giving an acceptable naturalist is therefore to single out one or more features of morality as core and non-negotiable, in the sense that they must be accounted for, and cannot be left out in a naturalistic analysis. I have already attempted to do just this, by claiming that irreducible normativity is a core and non-negotiable feature of morality because it is needed to undergird platitudes about the normative force of moral requirements.

In answer to this, the moral naturalist can go one of two ways: Either the naturalist can (i) claim that such normative force can be accounted for within a naturalistic framework, or (ii) deny that it really is a platitude about morality that it possesses the relevant form of normative force. I will explore these two strategies below.

**Option 1: Naturalizing strong normative force and inescapability**

One popular way of pursuing the first option is to try to find a naturalistically acceptable type of reasons that possesses the strong normative force we associate with moral reasons. One way naturalists have tried to locate such normative force in the natural world is by claiming that moral reasons are just a subset of practical reasons more generally. As we saw previously, hypothetical reasons, which are usually taken to be naturalistically acceptable, possesses such strong normative force. Gilbert Harman gives a concise and simple formulation of this view:

In this view, for an act to be wrong is for the agent to have sufficient reasons not to do it, where this is equivalent to saying that, if the agent were to reason correctly, the agent would end up deciding not to do the act. (1986: 66, cited in Joyce 2006: 194)

On this account then, what one has sufficient reason to do is what one would do if one reasoned correctly. What would ground an action being right is therefore that an action “instantiates the property is-such-that-[agent A]-would-want-to-do-it-if-he-were-to-reason-correctly” (Joyce 2006: 195). This then, could be a

---

153 The argumentative strategy in the two next sections is closely based on Joyce (2006).
candidate for a form of reason which is seemingly both naturalistically respectable (as long as the criteria for
correct reasoning does not involve any non-naturalistic properties) and which captures the normative force of
moral reasons. However, while we have now seen a candidate for a naturalistically acceptable reason with
sufficient normative force – which Joyce calls “practical authority” – we expect more from moral reasons.
We also expect them to be categorical in that they are not dependent on the ends of the agent, i.e. that they
are inescapable.

We cannot be content just to find a property that has practical authority [i.e. normative force] – arguably
we have located such a reason in “is-such-that-{agent A}-would-want-to-do-it-if-he-were-to-reason-
correctly”. We must also satisfy inescapability; we need a property that has authority over people
irrespective of their interests. But it is doubtful that any naturalizable account can deliver this. (Joyce
2006: 196)

The moral naturalist therefore needs to come up with something to undergird inescapable reasons with
strong normative force without appealing to irreducibly normative favouring relations. At this point, the
moral naturalist could try to ground such inescapable reasons by claiming that they “depend upon contingent
(even if deep) facts about the agent’s desires and interests” (Brink 1984: 114). Somehow, then, all agents
who reason correctly must have converging ends (i.e. desires or interests, etc.) which gives any agent reasons
that more or less align with what we take to be moral requirements.

What the moral naturalist evidently needs is a substantive and naturalizable account of “correct practical
reasoning” (or “practical rationality”) according to which any person, irrespective of her starting desires,
would through such reasoning converge on certain practical conclusions that are broadly in line with
what we would expect of moral requirements. (Joyce 2006: 196)

However, to argue that we should expect the ends of all rational agents to converge seems more like a
hopeful gesture, than a probability. However, it would take us too far afield to argue for this claim at any
length, so I’ll settle for having put forth a challenge, which, until the moral naturalist have overcome it, will
stop any such theory in its track.

One additional argumentative strategy that may be mounted against such naturalist views can be
based on my arguments from chapter four. Moral naturalists who choose the first option usually attempt to
show that one can ground moral reasons in practical reasons, i.e. reasons for action. However, in chapter four
we saw that there can be no practical reasons that both have strong normative force and that are
naturalistically respectable. If this claim holds, then even if the moral naturalist found some kind of
normative practical reason that could play the role of grounding moral reasons, that practical reason itself
would not be naturalistically acceptable, simply in virtue of being a normative reason with strong normative
force.

154 For an argument that such an account is unlikely to materialize, see Joyce (2006: 196–199).
Option 2: Giving up on strong normative force and inescapability

The second option open to the naturalist is to accept that there is no way to capture the strong normative force and inescapability of moral reasons within a naturalistic framework. That is, they could accept the possibility that their theory might not always provide us with a (non-institutional) reason to act according to moral requirements:

The utilitarian for example, may give up on trying to demonstrate a necessary connection between acts that maximize happiness and our having reason to perform such acts. (Joyce 2006: 200)

But what are we to say about a kind of theory, which Joyce terms “naturalism without clout”, which claims, on the one hand, that some action, \( \Phi \), is morally wrong, and that, on the other, we don’t necessarily have a (non-institutional) reason to comply with the moral requirement to not \( \Phi \). Depending on one’s personal inclinations, one might even only rarely have reason to comply with such requirements. Some moral naturalists have taken this option. Peter Railton, for instance, has defended a theory of this sort. Admittedly, he embraces the fact that the theory involves a revision of our ordinary understanding of morality, but claims that it is a tolerable revision:

Revisionism may reach a point where it becomes more perspicacious to say that a concept has been abandoned, rather than revised. No sharp line separates tolerable revisionism and outright abandonment, but if our naturalist wishes to make his case compelling, he must show that his account … is a rather clear case of tolerable revision, at worst (Railton 1989: 159, quoted in Miller 2013: 186)

Railton further hopes that his theory will constitute “a plausible synthesis of the empirical and the normative” (1986: 163, quoted in Miller 2013: 186). Let us quickly see whether Railton’s own view fits the bill. Railton defines moral rightness as “what is instrumentally rational from a social point of view” (1986: 200, quoted in Miller 2013: 196). Furthermore, he accepts only the existence of reasons that are grounded in an agent’s ends, that is, which are hypothetical. He also accepts the claim that we have no reasons to expect the aims of all rational agents to converge in such a way as to align them with what are commonly taken to be moral requirements (Railton 1986: 201). As a result, Railton ends up with a theory where we have to accept that we ought to do what is morally right, but it may very well be that we have no reason at all to do so.

To support this claim, Railton seeks to show that there are examples where ought’s do not give rise to reasons, i.e. that Mackie’s platitude is false. He does this by pointing to “epistemic ought’s”. Let us say that it is true that one ought not to have contradictory beliefs. According to Mackie’s platitude, it would therefore follow that one has a reason to not have contradictory beliefs, and, therefore, to root out such inconsistencies. Is that the case? It would seem not. It would be strange if what we should in fact do was to spend countless hours mapping out all our beliefs and rooting out all inconsistencies. Sometimes, of course, it will, for some reason or other, be important for us to have true, and therefore consistent beliefs. And in
such cases, where the “logical ought” aligns with the ends of the agent, there will be a reason for the agent to comply with the ought, in virtue of the normative force of hypothetical reasons.

First of all, Railton should probably adopt the account of institutional reasons outlined earlier, so that he could preserve Mackie’s platitude, by, for instance, taking both moral and epistemic ought’s to give rise to institutional reasons, that is, reasons with weak normative force. However, this would only take him so far, since, as we have seen, the claim that moral reasons are institutions seems implausible on the face of it. And while it is tempting to claim that such accounts of morality are simply a bad match with ordinary discourse, a moral naturalist of Railton’s stripe is likely to respond that it is a tolerable revision of our concept of morality, and that it “earn its place by facilitating the construction of worthwhile theories” (1989:157).

But is it a tolerable revision? And how can we determine if it is? In chapter 2, I argued that the commitment to irreducibly normative favouring relations are so central to our understanding of morality, and in particular of moral reasons, that without it, we would no longer be talking about morality. However, it is not easy to secure such a claim. Some error theorists have even given up on the idea of finding out whether something really is a non-negotiable part of a concept only by attempting to analyse it in the way I attempted in chapter 2 (e.g. Joyce 2006: 200).

So when the moral naturalist pursues the option ii), and simply claims that “no, strong normative force and inescapability are not central and non-negotiable commitments of moral discourse”, we have reached an impasse. In order to move the discussion forward we, therefore, need to look for an independent argument. One suggestion for an alternative way of settling the dispute is to look more specifically at the role the concept in question plays in a given discourse. If we want to find out whether something is a non-negotiable part of moral discourse we should therefore try to imagine that same discourse with the modified concept in use, and then evaluate whether the revised concept, with the problematic feature removed, could carry on playing a sufficiently similar functional role to the original concept (Joyce 2006: 201):

[T]he question we need to ask is whether moral discourse could carry on playing whatever role it does play if the connection between its prescriptions and the reasons people have to comply were merely a reliable contingent one. If not, then we have grounds for doubting that such a framework counts as moral at all. (Joyce 2006: 202)

So, what would happen if, like on Railton’s account, we would no longer take agents to automatically have a (non-institutional) reason to comply with moral requirements? We are supposed to imagine, then, that morality only sometimes provides us with a reason with strong normative force, and when it does, it is always because it aligns with an individual’s ends. On such an account, morality would in important respects become like etiquette. It would become like etiquette in that its requirements could often be flaunted with the same legitimacy that we flaunt the requirements of etiquette (Joyce 2006: 202–203). The requirements of
both would still be inescapable, in that everyone ought to follow the requirements of morality and etiquette. However, in those cases where a moral requirement does not align with one personal ends, one would be in a position to flaunt the requirement just as one is in a position to flaunt the requirements of etiquette when they do not align with one’s personal ends.

When a person decides to eat with their elbows on the table, it might very well be true that this has the property of being wrong-according-to-etiquette. As per Mackie’s platitude this fact is taken to give us a reason to not eat with our elbows on the table. However, such institutional reasons do not possess strong normative force. While stealing would, therefore, still have the property of being wrong-according-to-morality, this in itself would not provide any weightier reason for you to comply with it than the rule of etiquette about elbows provides you with. Of course, there could be many prudential reasons, reasons that stem from the agent’s ends, that would speak against stealing or putting elbows on the table, for that matter, but these would not be recognizably moral reasons. We would not normally take, say, fear of punishment as a moral reason not to steal.

To help bring out the consequences of a discourse revised in line with versions of moral naturalism which opts for the second option, we can add qualifiers before our judgments, just as we are likely to do in the case of etiquette (cf. Joyce 2006: 204). For instance, we are likely to say that “according to etiquette, you should not put your feet on the table while eating” or “according to the rules of etiquette, you have a reason not to put your feet on the table while eating”. It is not difficult to think of a context where we could follow up such statements with the addition that “but there’s really no reason to heed that – go ahead, put em’ up!”.

Since the imperatives and reasons of morality are, on the naturalist’s theory, logically equivalent to such institutional rules and reasons, we should be able to do the same in case of morality.

Imagine, therefore, that you encounter a killer who, after murdering a victim, goes through the following line of thought: “Of course, according to the rules of morality, killing is wrong. However, as there was nothing in my ends that gave me a reason not to act on my desire to go on a murderous rampage, I chose that. Again, sure, it is wrong according to morality, but, after all, that does not have any say over what I really ought to all things considered. I only did what it was practically right and rational to do!”

Furthermore, think about how we could respond to this. It seems like this form of naturalism opens for the response that “It was completely unacceptable to kill that innocent person according to the dictates of morality, but yes, you had every reason to do as you did, and no real reason to refrain.”

Moral discourse on this account seems nothing like the one we now know. It does, however, look a lot like many other discourses that we do know. One consequence of opting for the second option, is that the moral naturalist fails to carve out a separate discourse for moral judgements. On this view, moral discourse essentially collapses into other discourses where one weighs the costs and benefits of actions on the basis of

---

155 This example is modelled on the one found in Joyce (2006: 204).
how they would serve the ends of the agent, i.e. prudential discourse. However, as I showed in chapter 1, there is good reason to think that moral judgements are distinct from other kinds of normative judgements. When confronted with the question, “Why do we need a distinct moral discourse?”, the naturalist, as Joyce points out, has only one answer available:

“We don’t”. As a champion of morality – as a defender of moral realism, no less – the moral naturalist must feel extremely uncomfortable confessing that moral discourse is superfluous and moral thinking serve no distinct function, and in fact are superfluous to our social decision making. (Joyce 2006: 208)

While I will not claim to have come close to showing that there could not be a naturalist theory of morality, I do think I have shown that it, at least in the incarnations outlined above, seems less than attractive. Given the ease with which the favouring framework managed to account for ordinary moral discourse, it seems to be preferable to any account the moral naturalist have presented so far. While it remains possible that the naturalist will manage to provide a framework which overcomes the difficulties mentioned above, this should not detain us from pursuing the more promising accounts offered in this thesis.
References


Broome, J. 2015, “Reasons Fundamentalism and What Is Wrong with It”, talk at CSMN, November 2. 2015.


Gauthier, D. 1986, Morals by Agreement, Oxford University Press.


Gibbard, A. 2012, Meaning and Normativity, Oxford University Press.


Scanlon, T. 1998, What We Owe to Each Other, Harvard University Press.


