Normativity and Ethics

The Error in the Error Theory

Kipros Lofitis

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 2018

Abstract

Do moral norms invariably supply agents with reasons? Does the dedicated immoralist necessarily have adequate reason to act morally? Ordinary thought may have it that the answers are *yes* and *yes*, but some people think *no* and *no*. The error theorist, in particular, thinks so on the grounds that all reasons, weak or strong, are instrumental: necessarily favouring actions that are instrumental to the agent's own ends. For this reason, she is inclined to impute ordinary thought with error. This thesis examines the error theorist's grounds for thinking this, and finds none convincing. The resulting conclusion is that — for all we know — moral norms *do* invariably supply agents with reasons, and that the error theorist herself errs in thinking that they do not. The thesis closes by considering an alternative route to supporting the claim that moral norms do not invariably supply agents with reasons, which even though is rather attractive, is also in dire need of support.

Acknowledgements

First and foremost, I must thank my parents for their unconditional support throughout my life – and of course, this thesis would have not been possible without them. I want to also thank my sisters, Kalia and Maria, for supporting and encouraging me. I must express my gratitude to the University of Nottingham for giving me the opportunity to come here and be taught by people of the finest calibre. I am also deeply indebted to the people associated with the *Vice-Chancellor's Scholarship for Research Excellence* for providing me with significant funding for my studies.

I must thank Neil Sinclair for his exceptional support as my first supervisor throughout the last four years (and up to the last day before submission). Neil encouraged me to pursue a PhD back when I was doing an MA, and I shall always be indebted to him for that. I also thank my former second supervisor, Uri Leibowitz, for helping shape the direction and focus of my research right at the beginning. Andrew Fisher also deserves mention for providing me with valuable advice on substantive matters throughout.

It would be a serious omission not to mention my friends – the people who have provided me with stepping stones to move forward throughout the last four years. Special thanks to George Neokleous – a person of wisdom who has assisted me in numerous ways. For participating with me in meaningful conversations (many of which related to the questions raised in this thesis), I thank Sotiris Souroullas, Nikolas Gavriel, Stelios Kasikis, Dimitris Ziggas, and Nakis Ekkeshis. I am also grateful to my fellow student at the University of Sheffield, Dannish Kashmiri, for an intellectual partnership that will last for years. A valuable friend, Tony Tavelloudes, deserves mention for being willing to listen to my ideas from an early age. Finally, to the unsung heroes with whom I now share distance: you didn't have to treat me as good as you did, but you did, and I thank you.

Preface

I normally do whatever I have to do to fulfil my ends: goals, aspirations, concerns and the like. *But* only when there are no important moral considerations at play which call for an incompatible course of action. Suppose I want to eat your cake, and a promising way of doing so involves making you a false promise that I will walk with you on your way home. Insofar as lying to you involves harming you, this is a consideration that is relevant to what morality wants from me: morality wants me to not harm you, and therefore it wants me to not lie to you — despite it being true that doing so is a promising way to get what I want. I normally let these types of thoughts influence my acting, and this is what separates me from the dedicated immoralist — a figure which will be a reference point throughout this thesis (Shafer-Landau 2009). A dedicated immoralist is someone with no concern whatever for moral considerations, whose ends are invariably immoral, and who can pursue them without any material personal cost.

But why do I let these moral considerations influence my acting in the first place — why do I care about what morality would have me do? I can attest to this much: I do have this thought, and I have always had it, that despite what I want and where my personal interest lies, moral considerations give me reasons for action. What morality would have me do, that is, is what I have important reason to do, and for whatever reason, I am a being who is so fond of those things — *reasons* — that lets them influence his actions on a regular basis. In this thesis, I concentrate on this thought and attempt to crystallise it. In chapter 1, I work my way towards arriving at a clear formulation of it, which makes accurate reference to important notions that underwrite this thought of mine: reasons, norms, and the weight of reasons. I then ask: is there anyone who disagrees with this thought of mine, and if so, on what grounds? The error theorist disagrees and will be my opponent throughout this thesis. The error theorist thinks that the thought that moral norms invariably supply agents with reasons is mistaken on the grounds that all reasons, weak or strong, are instrumental: necessarily favouring actions that are instrumental to the agent's own ends. For this reason, she is inclined to impute ordinary thought with error. This thesis examines the error theorist's grounds for thinking this, and finds none convincing. The resulting conclusion is that – for all we know – moral norms *do* invariably supply agents with reasons, and that the error theorist herself errs in thinking that they do not. The thesis closes by considering an alternative route to supporting the claim that moral norms do not invariably supply agents with reasons, which even though is rather attractive, is also in dire need of support.

In examining the above thought, or at least casting considerable doubt on thinking that it is mistaken, this thesis raises a number of interesting issues. It should therefore be of great interest to a number of people. It should concern those who are interested in the very notion of reason. This thesis crystallizes the notion of a reason: carefully distinguishing between explanatory, motivating and (importantly) normative reasons. It also explains different types of normative reasons, and attempts to understand considerations pertaining to the strength or the *weight* of normative reasons.

Before examining the question of whether dedicated immoralists invariably have reason to act morally, this thesis touches upon a fairly neglected question: is the

thought that they have such reasons really part of ordinary thought? In arguing that it is, this thesis engages with what may be called an empirical investigation: based on uncontroversial, common instances of the behaviour of those engaging in moral discourse, it attempts to infer whether the thought underlies such behaviour. This part of the thesis (chapter 3) should be of interest to moral relativists who entertain the hypothesis that the so-called assumption of moral absolutism (see 3.3) is not something that ordinary people subscribe to (people interested in the work of Harman, Finlay, and Joyce).

In its substantive part, the thesis examines the error theorist's grounds for thinking that there are no moral reasons. In doing so, it raises an important question pertaining to what types of normative reasons there are. It examines the plausibility of practical instrumentalism — the view that there are only reasons favouring actions that are instrumental to the agent's (for whom they are reasons) ends, against the plausibility of its rival, non-instrumentalism: the view that reasons need not be thus instrumental. In doing so, it raises a methodological question, namely how do we examine such conflicting intuitions held by respectable thinkers. In answering this question, it calls attention to two criteria that any theory of reasons should be subject to: first, the extent to which it answers to our intuitions, and second, the extent to which it can be given a metaphysical grounding. Thus chapter 4 should be of interest to anyone interested in the initial plausibility of competing theories of reasons, just as to those interested in which methodological principles govern similar dialectics.

In examining potential support for practical instrumentalism, the thesis touches upon a traditional debate: between internalists and externalists about normative reasons. The debate centres on an alleged connection between normative reasons and the motivation of the agent (for whom something is a reason). Internalists posit a close connection between the two, thereby excluding the possibility of so-called external reasons: ones that need not be potentially motivating for the agent. A plausible connection between internalism and instrumentalism will be made, in light of which, internalism may well be thought of lending support to instrumentalism and hence addressing the question of moral reasons. Chapter 5 should be of interest to anyone interested in the debate between internalism and externalism, but also the connection between the former and instrumentalism.

Our journey towards locating support for instrumentalism continues in chapter 6, where we consider an interesting argument which purports to show that as a matter of conceptual necessity, all normative reasons are instrumental. The idea will be that the notion of an instrumental reason is conceptually basic in that it is required to make sense of any reason-claim. This is an argument which attempts to infer a substantive conclusion — pertaining to what reasons there are — from a conceptual premise — pertaining to what the concept of a reason is. The chapter should be of interest to those interested in the relevant distinction between the *conceptual* and the *substantive*, but also in the plausibility of inferences from conceptual premises to substantive conclusions.

The investigation into the plausibility of instrumentalism, and by extension the question of moral reasons, ends with a consideration of a constitutivist argument. We shall understand a constitutivist argument as an attempt to ground all normative reasons in what is what constitutive of agency. The interesting claim that we shall

consider will be this: aiming at the attainment of one's ends is the single constitutive aim of agency, and hence agents as such have only reason to aim at the attainment of their ends. In other words, there are only instrumental reasons. This is effectively an attempt to pick out instrumental reasons and to infer, based on considerations about agency, their normative supremacy. Given that our objection to this argument will be general – meaning, applicable to virtually any constitutive argument, chapter 7 should be of interest to those interested in the plausibility of constituvist strategies at large.

The thesis closes by sketching an alternative way that the error theorist may argue that there are no moral reasons: that moral norms do not invariably supply agents with reasons. In contrast to the standard error theorist, who bases her claim that there are no moral reasons on the premise that all reasons are instrumental, the alternative error theorist does so on the premise that there are no normative reasons (of a certain type) to begin with. Understanding the subtleties of this alternative view requires a distinction (drawn in chapter 1) between two types of normative reasons, namely *pro tanto* and *sufficient*. The resulting, alternative view that emerges will have it that moral norms invariably supply agents with pro tanto normative reasons, though they do not supply any sufficient normative reasons. This will be a striking result, given that on the standard error theorist's view all reasons — pro tanto or sufficient — are strictly instrumental. The closing chapter should be of interest to anyone interested in alternative ways of arguing for error theory. Moreover, the chapter lists some possible, interesting implications of thinking that there are no sufficient normative reasons as such. Notably, an implication may be that despite the truth of error theory, i.e. the alternative version, it may not be possible to justify selfish (or even immoral) behaviour.

Lastly, the thesis as a whole should concern those interested in the plausibility of moral error theory as a family of views. Different branches of moral error theory may concentrate on the plausibility of different ideas that may be part of ordinary thought. Depending on the thought that is at stake, we get different types of moral error theory (see 2.2.1, 'The many moral error theories'). As many writers have observed, it is an error theory about normativity or reasons that seems the most promising candidate to successfully impute ordinary thought with error (cf. Olson 2014, Garner 1990). The substantive part of this thesis (chapters 4-7) effectively attempts to test possibly the most promising type of moral error theory to the core.

Historically, ordinary thought has got some things wrong. We sometimes thought that the earth was flat, but as it turned out, it is not. And though it was the task of the topographer to determine the answer to the question of flat earth, it may be the task of the philosopher to determine the answer to the question of moral reasons: the possibility of reasons invariably supplied by moral norms.

Table of Contents

ABSTRACT II

ACKNOWLEDGEMENTS III

PREFACE IV

1. ETHICS AND THE NORMATIVE QUESTION 1

1.1 Ethics and Morality 1

1.2 Normativity 4
1.2.1 Normative frameworks 4
1.2.2 Reasons 5

1.3 Thesis Question 8

1.4 Conceptual Analysis 12
1.4.1 Terms and Concepts 12
1.4.2 Drawing on the Classical Theory of Concepts 13
1.4.3 Smith's Method For Conceptual Analysis 14
1.4.4 Addressing the Paradox of Analysis 16
1.4.5 Defective Analyses and Defective Concepts 17

1.5 Summary 18

2. MORAL ERROR THEORY 19

2.1 A Distinct Dimension of our Experience 19

2.2 Moral Error Theory: General and Specific 19
2.2.1 The Many Moral Error Theories 20

2.3 Ordinary Moral Behaviour 23
2.3.1 The Unconditionality of Moral Judgement 24
2.3.2 The Practical 'Oomph' of Moral Judgement 25

2.4 Ordinary Moral Thought 27
2.4.1 Moral Inescapability 28
2.4.2 Moral Bondage 29

2.5 An Argument for Moral Error Theory 33

2.6 Summary and Planning 35

3. IN DEFENCE OF THE CONCEPTUAL CLAIM 36

3.1 A Conceptual Question 36

3.2 The Conceptual and the Substantive 36

3.3 Finlay's Terminology **38**

3.4 Degrees of Eccentricity **38**

3.5 Methodology **40**
3.5.1 Problems with Introspection 40
3.5.2 The Evidence, and Inference to the Best Explanation 42

3.6 Weak Evidence **42**
3.6.1 Linguistic Evidence 42
3.6.2 Appraisal Evidence 43
3.6.3 Reactive Evidence 45

3.7 Strong Evidence **46**
3.7.1 Address Evidence 47
3.7.2 Expectation Evidence 48
3.7.3 Disputation Evidence 50

3.8 Finlay's First Move: Disputing the Pervasiveness of the Evidence **51**
3.8.1 Are Such Practises Characteristic of Moral Discourse? 52
3.8.2 Such Practises Are Odd 54

3.9 Finlay's Second Move: Inference to the Best Explanation **55**
3.9.1 The distinction Between the Pragmatic and the Semantic Point of Moral Practise 55

3.10 Conclusion **59**

4. PRACTICAL INSTRUMENTALISM 60

4.1 From the *Conceptual* to the *Substantive* **60**

4.2 Practical Instrumentalism **63**

4.3 Why Think that Instrumentalism Is True? **64**

4.4 Extensional Adequacy: Problems for Crude Instrumentalists **66**
4.4.1 Failure to Exclude ‘Bad’ Desires 67

4.5 The Neo-Humean (Version of) Instrumentalism **68**
4.5.1 Distinguishing Between Objective and Subjective Reasons 68
4.5.2 The Neo-Humean Instrumentalism 69
4.5.3 The Apparent Merits of the Resulting Theory 70

4.6 Extensional Adequacy Revisited: Problems for Neo-Humean Instrumentalists **71**
4.6.1 Smith's Account of the Idealised Counterpart 72
4.6.2 Schroeder's Way 74
4.6.3 The Dialectic As It Stands 79

4.7 Metaphysical Grounding **80**
4.7.1 The Instrumentalist's Claim of Superiority 81
4.7.2 The Non-Instrumentalist's Response 82

4.8 The Instrumentalist's Mistake: A Diagnosis **87**
4.8.1 Rationality: Narrow vs. Broad 87
4.8.2 The Broad Conception of Rationality Is that of Reasons 89
4.8.3 Making Use of our Resources 90

4.9 Explaining Away the Apparent Appeal of Instrumentalism **91**

4.10 Conclusion **91**

5. REASONS INTERNALISM 93

5.1 What Is Internalism, and Why Are We Interested? **93**

5.2 Introduction to Internalism **94**
5.2.1 Motivation, Deliberation and Deliberative Routes 94
5.2.2 Reasons According to Internalists 94

5.3 The Argument from Alienation **97**

5.4 The Argument from the Explanation of Action **100**

5.5 Two Direct Attacks at External Reasons **103**
5.5.1 The Epistemology of External Reason Beliefs 103
5.5.2 The Semantics of External Reason Statements 105

5.6 Arguments for Externalism **107**
5.6.1 The Argument from Examples and Counterexamples 107
5.6.2 The Distinction Between Normativity and Psychology 110

5.7 Conclusion **112**

6. A CONCEPTUAL ROUTE TO A SUBSTANTIVE CONCLUSION 113

6.1 Our Investigation into a Substantive Question **113**

6.2 Scepticism about the Justification of Morality **114**

6.3 Two Naive Attempts **116**
6.3.1 An Alleged Link Between Rationality (In Some Sense) and Reasons 116
6.3.2 Insisting on the Primacy of the Instrumentalist Conception of Rationality 118

6.4 Two Sophisticated Attempts **120**
6.4.1 Insisting on the Primacy of the Reasons of Instrumental Rationality 121
6.4.2 An Argument in Favour of the Alleged Primacy of the Reasons of Instrumental Rationality 124

6.5 Conclusion **131**

7. A CONSTITUTIVIST ARGUMENT 133

7.1 What Is Constitutivism, and Why Are We Interested? **133**

7.2 Constitutive Resources **135**
7.2.1 Constitutive Norms 135
7.2.2 Constitutive Aims 137
7.2.3 Action and Constitution 138

7.3 The Analogy with Chess, Revisited **139**

7.4 Towards our Constitutive Argument **141**

7.4.1 A Parallel with Believing 141
7.4.2 Our Conception of Action: The Constitutive Aim of Agency 143
7.4.3 A Constitutivist Argument for the Supremacy of Instrumental Reasons 145

7.5 Objections 146
7.5.1 Agency, Schmagency? 146
7.5.2 Inescapability Does Not Entail Normativity 149
7.5.3 Normative Arbitrariness 150

7.6 Conclusion 153

8. STANDARD AND ALTERNATIVE ERROR THEORIES 154

8.1 The Standard Route To Error Theory 154
8.1.1 Support For the Standard Error Theory 156

8.2. An Alternative Route to Error Theory 157
8.2.1 The Merits of the Alternative Error Theory 159
8.2.2 Support for P1* 160

8.3 Assessment of the Alternative Error Theory 162
8.3.1 Is the Alternative Theory Plausible? 162
8.3.2 Should the Alternative Theory be Troubling to the Moralist? 164

8.4 Aftermath 166

BIBLIOGRAPHY 168

1. Ethics and The Normative Question

1.1 Ethics and Morality

We care about things. We are interested in getting some of the things we want some of the time. I want the new iphone, some nice clothes for the summer, to visit my cousin in Greece, to spend a weekend in a Greek island, among other things. Perhaps I cannot have all those things; I may need to sacrifice some for the sake of others. Before I do so, I may want to think it through — for I want to make the most of my practical decisions. If I want the new iphone, and know that I can get it 30 pounds cheaper from the store that has just opened in the nearby town, I may have to consider getting it from there. But I first need to calculate the costs of getting there, and see whether they would exceed the price difference, thereby making the decision less advantageous than I initially thought. Like you, I am rational in a certain sense: I calculate my actions with respect to my goals — I then do those actions that are mostly advantageous to my goals. In short: I am — at least partly — a *means-ends* rational being.

But like me, you are *more* than that. For we are all — at least the people I have met — also interested in living an *ethical* life. An ethical life is one that is characterised by a positive net balance of *justified* actions. An ethical action is one that we do not just do, or recommend that others do, based on self-interest. It is rather one that is justified via a *particular* kind of justification: ethical justification. As Peter Singer memorably puts it:

> If I am to defend my conduct on ethical grounds, I cannot point only to the benefits it brings me. I must address myself to a larger audience (Singer 1979: 10).

Thus if I aspire to ethically justify (say) abortion, I cannot just say: abortion is justified because it enables *me* to deal better with premarital relationships. To say 'enables *us*' may be a start. But even this will not do; to say that abortion is ethically justified because it suits the interests of my group (no matter how large the group is) will still not do. To arrive at an ethical justification, we need to inject some ideology. Consider this attempt: abortion is ethically justified because, even though foetuses have a *right* to life, women too have a *right* to decide what to do with their body (cf. Thomson 1971). This attempt features both a *we* that is big enough to apply to anyone finding herself in the relevant circumstances, and talk in terms of *rights* (or what one *deserves*, is *entitled* to or *should* get) — normative notions that transcend the unenlightened talk of mere *wants* and *interests* (more on normativity in 1.2).

Insofar as others (individuals or groups of people) are interested in justifying themselves and their actions to a wider audience, they can count as ethical. As Singer puts it, the mere attempt at justification 'is sufficient to bring the person's conduct within the domain of the ethical as opposed to the non-ethical'. We may regard this a *minimal* sense of *ethical* (cf. Garner 1994: 18). Ethical beings, in this sense, contrast with wild beasts and human animals: those with no appreciation whatever for a way of life shaped by ethical standards. A lion will eat a pregnant deer in order to get some

food, a contract killer will kill for money. Neither will be interested in justifying their actions.

But slave-owners in the past were keen to offer arguments for the justification of their actions. Some are likely to have thought that slavery actually benefitted the enslaved as the latter were unfit for other types of work, and others pointed to a certain interpretation of the bible on which slavery was approved by God. If asked, they would not just say: we do it because it benefits *us*, deal with it. (It is a legitimate question whether they also *felt* justified.) They regarded themselves as ethical beings, and many thought they were occupying moral high ground. Slave-owners, neo-nazis and duck-hunters: when asked for justification, they may all have a story to tell. In justifying their actions, they generate '*an* ethics' of their own — so to speak.

But are they *right*: *really* justified in their actions? On the face of it, some seem exceptionally deluded. No matter how well-thought-out a slave-owner's justification may be, to own a slave and mistreat her is still to own and mistreat a fellow human being, and hence an attempt to justify the latter seems doomed right from the start. Slave-owners may attempt to justify their actions by reference to actual facts; e.g. that extra workers are necessary for the development of their culture. This may be *a* or *some kind of* justification, in the sense that it explains why they do it, but it does not seem to suffice for a good enough justification. The slave-owners may have an ethics of their own, but it so naturally strikes us as the *wrong* ethics — a substantively *mistaken* ethical code. Intuitively, it seems possible that someone (or a group of people) may be *wrong* in thinking that their actions are justified. That is, they may have a justification for their actions — but one that fails to do the job for which it was hired: to convince a clear-headed thinker about the justification of actions.

The above type of criticism seems possible because common sense allows for a distinction between *some* ethics and the *right* ethics. Following others, we may reserve the term *ethics* or *ethical code* to refer to any old set of principles purporting to guide and justify actions, e.g. as in the slave-owners *ethics*, and the term *morality* to refer to the *right* or *correct* one (cf. Williams 1985: 6). To say that a group of people have an ethical code is only to say that they have a set of recommended and — on their view — justified ways of acting and being (Garner 1994: 17). To say that their actions align with morality, or that they are moral, is to say something more: that they not only live by some ethical standards, but that they live by the right or correct ones.[1] Slave-owners and nazis have ethics but not morality, they are thus to this extent mistaken, and duck-hunters — I believe — are too.[2] For we must have our own reservations as to whether 'our' morality — what we judge to be the right ethics — really is morality.

When Nietzsche famously called for 'a reevaluation of values', he was expressing an ethical attitude and a firm conviction: that the conventional ethical code

[1] This is not to claim that ordinary language perfectly aligns with this terminology. For example, we may sometimes call an alien ethical code *a* morality, but insist that it is not the *real* morality (cf. Joyce 2011: 528).

[2] We need to be careful to allow for two main possibilities. First, a group of people maintaining an ethical code that they do not judge to be the right ones - one that they merely favour or just suits them. Second, a group of people maintaining an ethical code that they also judge to be the right one: morality.

of his time did not represent morality — and thus should be replaced by a new one.[3] But perhaps Nietzsche and duck-hunters are not mistaken, maybe I am wrong and maybe they are right. In this thesis, I shall remain neutral on what these answers are: e.g. whether duck-hunting is really justified or not. These are normative matters deserving special attention, a lot more than an essay in metaethics can give them.

At the same time, I shall take it for granted that common sense is right in thinking that there are right or correct ethical codes: right or correct answers to ethical questions, and hence a *morality* in our sense. In particular, I shall assume the truth of a minimal form of moral realism: that there are facts about the moral status of situations — a moral reality that people are trying to represent when they issue judgements about what is (ethically or morally) right and wrong (Shafer-Landau 2003: 13). For example, I assume that duck-hunters are *either* right or wrong in thinking that duck-hunting is morally justified: that the question whether they are right or wrong both makes sense and has a correct answer. The realism that I assume is minimal enough as to remain neutral on the question that will concern us in this thesis (more on this shortly), but substantial enough as to have it that the sense in which duck-hunters are wrong (if they are) is *objective*, as opposed to *subjective* or *relative*.

If duck-hunting is wrong in either the subjective or the relativist sense, then there is no *single* answer to the question whether it is right or wrong. On subjectivism, the moral status of an act is fixed by subjective elements such as desires, feelings, purposes or points of views (Garner 1994: 59). Thus it may be right for duck-hunters, who happen to have an emotional attachment to the act, but wrong for others who happen to like ducks. Similarly, on relativism, the moral status of an act is fixed by conventions which may be peculiar to particular groups of people. Thus duck-hunting may be right *relative* to the duck-hunters' ways of doing things, and wrong *relative* to other peoples' ways. On this view, it is just odd to say that duck-hunters are wrong — *period* to hunt ducks (Harman 1975: 7).

By contrast, if duck-hunting is wrong in the objective sense, it is wrong in a similar sense that *1+1=3* is: independently of how we — or anyone else — happens to think or feel about it. In holding this, I take it that there is such thing as *ethical* reasoning: a correct, clear-headed way of thinking about ethics which promises to lead to substantive ethical truths: morality. And that we must keep thinking straight, exemplifying ethical reasoning by trying to remain unaffected by 'the distorting effects of self-interest'[4], in order to arrive at such truths.

In thinking the above, I oppose myself to moral subjectivists and relativists, and in assuming it, I am certainly begging some questions against them — important questions that many people (myself included) are interested in. But I shall do so because my concern here is to remain neutral on the *normative question*: granted that there are moral facts, even objective ones, about actions, *why should we do as moral facts say*? In an analogous sense: granted, there are facts about polite behaviour, why should we be polite? Relatedly, why should we be interested in a positive net balance

[3] (Nietzsche 1911)

[4] This is a name given by David Enoch to the phenomenon of letting our interests influence our thinking on ethical matters (Enoch 2009). I may have a vested interest in thinking that X act is justified, because it helps me deal with Y, and as a result I am (mistakenly) inclined to think that X is justified as a practise.

of morally justified actions over unjustified ones? In 1.2, I explain how I understand normativity, before I return and elaborate on the thesis-question in 1.3.

1.2 Normativity

We often agree on what *happened*, but strongly disagree on what *should* have happened. In the midst of a crisis, the president calmly walked through the press room, sat in front of the camera, and announced to the nation that the country is going straight to hell. This is what happened; but it may not be what should have happened. You think that the president *should* have said something more positive to lift the people's morale, but I think he did the *right* thing in expressing his honest view. Your brother points out that he *should* at least have worn a tie with the nation's colours during the press conference.

We may start by vaguely distinguishing between two dimensions of our experience: the descriptive and the normative. Insofar as we approach situations by thinking in terms of *shoulds*, we engage with the normative dimension. Indeed, our minds are host to a large number of thoughts that purport to tell us what we (and others in our circumstances) *should* do. Even though I feel far more comfortable in shorts, I think I should wear a costume at your wedding; even though I am tired, I think I should drive you to the airport; and even though I do not know everyone I meet, I think I should be polite to them anyway. We may refer to each of those thoughts as an instance of normativity.

We must note that normative language may not always be straightforwardly about *shoulds*. For example, we often call an action (murder) morally *wrong*, a state of affairs (slavery) morally *bad* or *impermissible,* a person (Hitler) morally *atrocious*. When we do so, we seem to be *describing* something or someone. But even though normative judgements may have descriptive content, normative thought seems to be fraught with *oughts.* That is, such judgements do not *purely* or *merely* describe things: they also imply something about what *should* be the case (Robertson 2009: 3). Normative judgements are certainly supposed to have straightforward practical implications. If that were not the case, learning that (say) murder is wrong would be like learning what time it is: its practical relevance would depend on other things (Garner 1990: 139).

Shoulds that get repeated for long enough may become norms: established (in some way and to some degree) ways of behaving. And norms that become obsolete or forgotten may be regarded as *shoulds* that no one cares about. Checking your mobile phone while talking to others used to be considered offensive behaviour: we used to think that one *should* not do so. Thus it used to be a norm of polite behaviour that one should not do so. But nowadays it seems to have become somewhat acceptable to do so: hence the relevant norm seems to have vanished, dissolved into a *should* that no one cares about.

1.2.1 Normative frameworks

Different norms may hang together, or at least be regarded as a unified whole. Examples include the norms of chess, football and any other activity that consists of a certain number of norms. When this happens, we may refer to the totality of such

norms as a *normative framework*: a set of norms that tell us what to do. In this thesis, we are interested in morality — and morality may be regarded as just another normative framework: a set of norms. These norms are derived from fundamental moral facts — such as the fact that killing is morally wrong. For any act that is morally wrong, there may be a corresponding norm advising against it, and for any act that is morally right, there may be one advising in favour of it.

Normative frameworks may occasionally require that we do things, and these things may sometimes conflict with what we *intend* to do. Henry is about to enter a night club but does not want to wait in the long queue to get in, so he skips the queue and goes right in front of the doorperson. The doorperson tells Henry to go back in the queue, but Henry knows the owners of the place too well and so he has three options: to either wait in the queue like everyone else, be polite to the doorperson and get in through the backdoor, or call the doorperson a nasty name and get in through the backdoor. What Henry really wants is to go with the third option. Why should he not do it? We may give different, sensible answers. But consider one such answer: Because calling the doorperson a nasty name is impolite and skipping the queue is unfair to the other people waiting. In other words, because doing that offends against the norms of etiquette and those of morality — comes into conflict with what these frameworks require.

Moreover, different normative frameworks may come into conflict with each other. Etiquette may require that I do not talk with my mouth full, whereas morality may require that I do so in order to warn you that the food contains allergens. When this happens, we may ask: which framework should we favour in following, and which should we neglect? But why think that we should do as *either* framework requires?

Merely citing the presence of a norm — i.e. to say that you have to do *this* (touch your nose), because norm N (*Touch Nose Move*) requires that you do this — often does not suffice to make us follow it (McPherson, forthcoming). And why should it? When we ask why should we do as the norm says, we are interested in something more than the fact that a norm obtains: we open ourselves to a discussion of *reasons* (Joyce 2000: 463).

1.2.2 Reasons

We, humans, take our acting very seriously. We tend to be less spontaneous than other animals in that we often premeditate before acting, and act for *reasons*. As a result, our behaviour can be often given a *reasonable* explanation. Most importantly, we are often interested in performing those actions that are supported by *reasons*. We may note that 'reason' is a polysemous term, and the sense in which we act for reasons, that which our behaviour can be given a reasonable explanation, and the sense in which we are interested in actions that are supported by reasons, are not the same (Audi 2002: 237).

First, there is the notion of an *explanatory* reason: a reason *why* someone does something. We may use this notion in explaining why Donald Trump withdrew from the Paris agreement: the reason he withdrew is because he does not believe in climate change. Second, there is the notion of a *motivating* reason: a reason *for which* someone does something. A motivating reason is, by definition, one that motivates the agent for

whom it is a reason (Schroeder 2007: 12). The notion of a motivating reason may also feature in the explanation of the previous example: the reason for which Donald Trump withdrew is because he does not believe in global warming. In other words, Trump's lack of belief in global warming proves to be motivating (i.e. it provides him with a motivating reason) and also happens to partly explain why he did it (i.e. it provides us with an explanatory reason). The two notions are distinct, however, as we can envision examples in which they fail to overlap. We may explain why a mindless zombie does things in virtue of certain buttons being pushed, without having to postulate a motivating reason that it has. These two senses of 'reason' will not concern me here — at least not primarily, and I shall use the term to refer to reasons of the next kind.

Third, there is the *normative* sense of a reason: the notion of a *normative reason for action*. This is a reason there *is* for doing something: a consideration that justifies or supports a prospective course of action. Normative reasons point to the right or justified course of action, and they often do so regardless of how agents tend to behave or how they are like.[5] They are normative in that in supporting prospective courses of action, they make it the case that agents *should* (or *ought*) to perform those actions. This is close to the sense of 'reason' that we are interested in when we ask why we *should* do as this or that norm says, but not quite accurate, as even this normative sense of a reason can be ambiguous.

To see why, let us imagine our conversation with Henry of the previous example. We tell Henry that he should not call the doorperson a nasty name. He asks why. We cite a norm of etiquette: *one should not call another a nasty name*. Henry retorts that this norm is nothing to him: why should I do as the norm says? To this he adds: What reason do I have to do as the norm says? We are quick to point out that the fact that the norm obtains is a reason for him to not call the doorperson a nasty name. Henry is baffled, he replies: *that* fact may be *a* reason for me not to call the doorperson a nasty name, but the fact that I would take immense pleasure from doing so is a much *better* reason for me to do so!

Now, we may not agree with everything that Henry says, but there is one thought we may wish to take with us: that there can be (normative) reasons that are in a sense *better* than others, and hence deserve more attention when we deliberate about what to do. The intuitive observation can be also cashed out in the following way: not all reasons are good reasons. Consider the case of a pyromaniac who is deliberating about whether to set her own house on fire. As reasons in favour of doing it, she cites the fact that doing so would satisfy her pyromania, the fact that other pyromaniacs finding out about the incident will feel good about it, and the fact that burning the house down will open up space for the construction of a more modern one. She then notes the presence of another class of reasons, this time against doing it: the fact that she will become homeless, the fact that doing it will endanger the safety of those living nearby, and the fact that satisfying pyromaniac desires serves to foster them.

[5] This is at least an intuitive thought that we may preserve at the outset, even though some people (instrumentalists and internalists about reasons) may beg to differ. If you left your wallet at my house, and you are not sure exactly how much money you left inside, I can take a few notes without you noticing it. Given how much I want the money and that I can get away with stealing a few, it is a safe bet to suppose that I am going to do so. I may be motivated to do it, and my doing it may make perfect sense given what we know about me, but I may still have normative reason to refrain from doing so.

Intuitively, there seems to be *better* reason (or reasons) for her to refrain from setting her house on fire.

We may begin tying up the above observations by noting that in its *minimal* sense, a reason is a *favourer*: a consideration that counts in favour of a prospective course of action, while remaining neutral on whether that action is supported by the best reason (cf. Scanlon 1998: 17). Reasons X,Y,Z may be in favour of Φing, while reasons A,B,C may be in favour of not Φing. The question of which of the two actions, Φing or not Φing, is *most* supported by reasons is — for all we know — left open. We may call each of these reasons a *pro tanto* (or *defeasible*) reason: a consideration that justifies an action but which may be overruled by better ones. There may be a pro tanto reason for me to go to my favourite cafe this evening (in virtue of my wanting to), and a pro tanto reason for me not to go to my favourite cafe this evening (in virtue of my having promised to stay home with you), and yet a conflicting pro tanto reason for me to go to church this evening (in virtue of my being a christian). *Pro tanto* means: to such an extent, or as far as this consideration (or fact) is concerned. And some pro tanto reasons may be *better* than others.[6]

Next, we must grant the possibility that pro tanto reasons can — at least sometimes — *add up:* combine to generate better reasons (Schroeder 2007: 126). For example, I have pro tanto reason X to brush my teeth (in virtue of the fact that doing so will keep them fresh), and yet another pro tanto reason Y to do so (in virtue of the fact that doing so will spare me the pain of visiting the dentist). It seems that when *combined*, X and Y give me an ever better reason (than each in isolation) to brush my teeth. Moreover, and relatedly, reasons may come into conflict with each other (Robertson 2009: 4). Suppose I have yet another pro tanto reason Z to not brush my teeth (in virtue of the fact that the taste of the toothpaste is awful). The combination of X and Y comes into conflict with Z, as the latter favours a conflicting action.

To make justice to the above observation, that reasons can add up, we may drop the rather ordinary characterisation *better* or *good*, as in a good practical reason, for the most accurate: *weightier* or *weighty*. We must also note that the weight of reasons need not be a matter of quantity. That is, it may be possible that I have 10 distinct pro tanto reasons to Φ, and only one to not Φ, but the latter is weightier (or a better) reason than the combination of the conflicting 10. One weighty reason may outweigh a large number of lightweight ones.

It is noteworthy that, in cases where the is no weighty reason in favour of an action, we sometimes overstate how little that action is supported by reasons. Upon hearing about a very sad incident, we say: there was *absolutely no* reason *at all* for him to commit suicide. We want to say that the unfortunate act should have been avoided, and in doing so we seem to be saying something that is not true. For supposedly, the deceased may have had a *pro tanto* reason to commit suicide in virtue of the fact that doing so would spare him the unpleasantness of what he was going through, and yet another in virtue of the fact that doing so would give him his 15 minutes of fame. At

[6] It is important to note that I shall remain neutral on which particular considerations can count as pro tanto reasons, given that my interest here is with sufficient normative reasons. For it may be that on some people's view, a silly consideration (e.g. the fact that eating my car will provide you with the recommended daily allowance of iron) does not even qualify as a pro tanto reason (Schroeder 2007: 96).

any rate, these reasons do not seem to outweigh the much stronger reason he had to not commit suicide, grounded in the incomparable intrinsic value of life. Granted, he may have had some (lightweight) reasons for committing suicide, but the real question is: were they *good* (*weighty*) reasons?

We now have the resources to arrive at the normative notion of a reason that will interest us in this thesis: that of a *sufficient* normative reason for action. A *sufficient normative reason for action* is one that outweighs all other relevant considerations that are at play, thereby determining what the justified course of action is: the one that is most reason supported. In terms of weight, this is the weightiest reason. It is called *sufficient* in that it *suffices* to determine what the most reason supported action is, in virtue of being the weightiest one. Finally, there are two structural possibilities for a sufficient normative reason: it can either be the weightiest pro tanto reason that is at play, or the combination of subsidiary pro tanto reasons the total weight of which is greater than that of any other combination favouring a conflicting action. In this thesis, I may sometimes use the unqualified noun 'reason' to refer exclusively to a sufficient normative reason for action, unless stated otherwise.

1.3 Thesis Question

We do things. Some of the things we do are prescribed by different normative frameworks, and the latter sometimes come into conflict with each other. Etiquette may require that I do not talk with my mouth full, whereas morality may require that I do so in order to warn you that the food contains allergens. Moreover, some of the things we do are supported by reasons, and some are not. When the two conflict, I may choose to obey etiquette and disregard morality, but in doing so I may be neglecting my normative reasons. In general, the question of what I am required to do by *this* or *that* normative framework may seem independent from the question of what reason I have to do *this* or *that* action.

Indeed, with respect to some normative frameworks, the last thought seems to be valid. Richard Joyce (2001) discusses the dramatic example of Celadus, an unwilling gladiator (34-37). Celadus is thrown into the arena against his will, prompted to endure the pain and all that comes with his unfortunate role. But he does not want to do any of that. All Celadus wants is to escape the confines of the brutal arena, and try to get away from his captors. Suppose that he finds the right moment when the guards are relaxed, and that he has a bulletproof escape plan that will definitely work. Though it makes good sense to say that Celadus is *required* by the normative framework of gladiatorial-combat to stay in the arena and (say) refrain from throwing sand into his opponents' eyes, it makes equally good sense to say that he has no reason to do the former and every reason to do the latter.

Yet some think that with morality, the two questions are not independent. In particular, some think that agents always have reasons to do as moral norms[7] say. Let us return to the previous example, and suppose this time that on his way out of the

[7] In this thesis, I shall take it that moral norms (or a considerable subset thereof) *express* requirements, and leave it open whether there really are any such requirements (see 2.5). As I will be concentrating on those moral norms that express such alleged requirements, my use of 'moral norm' will be elliptical for 'moral norms that express alleged moral requirements'.

arena, Celadus notes a newborn child on the stands and thinks about grabbing it in order to sell it on the black market. We may grant that Celadus is morally required not to do this, but that he does not care: he wants to do it anyway. Some people think that the fact that Celadus is morally required to not grab the child is enough to tell us that he has a reason to not do so. We may call the defender of this thought: *moralist*. A moralist defends a kind of entailment between the following two matters: what one is morally required to do, and what one has reason to do.

In particular, I shall understand the moralist's claim as follows: moral norms (at the least the ones expressing requirements or obligations)[8] are reason-*supplying*: invariably *supply* agents (to whom they apply) with sufficient normative reasons. This may be mistaken for the claim that moral norms *entail* that agents have reasons to do as they say. The latter claim is weaker in that, at least in theory, it may be possible that agents always have reasons to be moral which are of a non-moral kind or supplied by a different normative framework. A rosy view, which would be too much to assume at the outset, would have it that as a matter of freak coincidence agents always have reasons of self-interest to be moral (Shafer-Landau 2003: 166). On this scenario, it would be true that moral requirements *entail* reasons: are in perfect alignment, so to speak, with what agents have reason to do. If moral requirements are mostly about altruistic actions, this is a particularly unrealistic scenario. This will become clearer in chapter 2, when we look closely at the content of moral requirements. For now, note that were this unrealistic scenario to hold, it would still not be true that moral requirements *supply* agents with reasons: agents would have reasons to do as moral requirements say *not because* they are morally required but due to some other fact that obtains.

It is important to note that the moralist's claim is rather intuitive. With the same confidence that we say that contrary to what gladiatorial-combat requires from him, Celadus has no reason whatever to abide by gladiatorial rules, we are also inclined to think that Celadus does have a reason to refrain from grabbing the child: in virtue of being required to not do so by morality. This result seems to generalise to all other cases. In virtue of being morally required, animal trappers seem to have a reason to be kind to animals, pyromaniacs seem to have a reason to preserve forests, misogynists seem to have a reason to respect women, egocentrics seem to have a reason to help others whenever they can do so at minimal personal cost. That the moralist's claim is intuitive tells us something about the nature of the upcoming dialectic: the burden of proof seems to be on the opponent, who denies the intuitive thought, to ground her claim.

The moralist's opponent <u>*denies*</u> that moral norms invariably supply agents with reasons. She thinks it is possible that an agent may be *morally* required to Φ but lack a reason — i.e. a weighty, sufficient normative reason — to do so. On her view, Adolf Hitler and Attila may have been morally atrocious, but in being so they were not necessarily ignoring their practical reasons. In general, agents may be acting selfishly,

[8] There may be a distinction to make between morally obligatory or required acts, on the one hand, and morally supererogatory or charitable acts, on the other. I may be morally obligated or required to take care of my own children, but it would only be morally good or 'great' to take care of other children too. It may be disputable where we draw the line when it comes to particular cases, but at any rate, I shall here assume that the moral norms that I am concerned with stand for moral obligations or requirements (Singer 1972).

cruelly, and in inconsiderate ways, but that does not mean that there are reasons against so acting (cf. Williams 1981: 110). In thinking this, the opponent rejects the existence of reasons invariably supplied by morality. Consistent with a previous observation, this is not to deny that moral norms *may* be always supported by reasons supplied by other normative frameworks — though for reasons already suggested, this is unlikely to be true if the opponent is right.

In this thesis, the moralist's opponent defends moral error theory: the view that ordinary moral thought is embedded with error. As we see in the next chapter, there are different argumentative directions that a moral error theory may take, with each focusing on a particular alleged error. As it may transpire at this point, the alleged error that the error theorist of this context focuses upon is: the thought that moral norms invariably supply agents with reasons. In attributing this alleged error to ordinary thought, the error theorist *accepts* that moral requirements are *supposed* to — or are *thought* to — supply agents with reasons. In my view, this makes the error theorist's claim rather credible. For, at least in my view, anyone who *denies* that moral requirements are *supposed* to be reason-supplying lacks competence with matters moral. This claim becomes clearer in the following two chapters, which provide an account of ordinary moral thought. But for now, let us ask: why does the error theorist think the aforementioned thought is mistaken?

The error theorist thinks — correctly — that we can answer the question whether moral norms invariably supply agents with reason by considering two things: first, a sound account of moral requirements (the alleged requirements expressed by moral norms), and second, the correct theory of reasons (Shafer-Landau 2003: 166). An account of moral requirements would enable us to understand the *kinds* of things that morality requires from us. For example, if it is true that, at least sometimes, we may be required by morality to neglect our interests for the sake of (say) helping others, we may infer that moral requirements are, at least partly, altruistic. The correct theory of reasons will, then, tell us what *kinds* of reasons there are. For example, are reasons invariably instrumental, that is, favouring actions that are instrumental to the agent's ends, or are there at least some reasons that are non-instrumental, that is, not necessarily favouring actions that are instrumental to the agent's ends?

The error theorist thinks that from a sound account of moral requirements, we must infer that moral requirements are not principally concerned with promoting self-interest: they may occasionally require that agents act against their own ends. From the correct theory of reasons, the error theorist thinks, we must infer that all reasons (*pro tanto* or sufficient) are instrumental: favouring actions that are instrumental to the agent's (for whom they are reasons) ends. It is the combination of these two views that prompts the error theorist to deny that moral norms invariably supply agents with reasons. For if all reasons are — as she thinks — instrumental, it may be possible that we are sometimes morally required to do an action for which we lack reason. And even if our ends and interests invariably align with moral requirements, and hence we invariably have reasons to do as moral requirements say, it is not true that we have such reason *because* we are morally required. For had our ends been different, we would have lacked reason to do as moral requirements say. Hence it cannot be true that moral requirements invariably *supply* us with reasons.

Thus the error theorist's case involves two major steps: arriving at a sound account of moral requirements, and defending a certain theory of reasons on which all reasons — pro tanto or sufficient — are instrumental. To understand what moral requirements are like, the error theorist engages with conceptual analysis: give an account of what we *think* when we talk about moral requirements, what our *concept* of a moral requirement is. In the last section of this chapter, I provide a method for conceptual analysis which will be used for the purpose of arriving at a sound understanding of moral requirements. It must be noted that I accept the account that the error theorist gives — and moreover, that I shall defend it further in chapter 3.

What I am reluctant to accept is the error theorist's claim that all reasons (pro tanto or sufficient) are instrumental. In the majority of this thesis, I consider the error theorist's grounds for thinking this, while finding them unconvincing. In chapter 4, I present the theory of reasons that the error theorist proposes, *practical instrumentalism,* which holds that all reasons (pro tanto or sufficient) are instrumental — thereby ruling out the possibility of moral reasons: denying the thought that moral norms invariably supply agents with sufficient normative reasons[9]. In examining practical instrumentalism, I consider two criteria that a theory of reasons must fulfil. I conclude that it is far from clear that instrumentalism fares better than its rival, non-instrumentalism, with respect to either.

In chapter 5, we consider *reasons internalism*: a family of arguments all designed to lend support to instrumentalism. Internalism specifies a criterion or restriction on what could count as a normative reason for action (pro tanto or sufficient): something is a reason only if it can potentially motivate the agent (for whom it is a reason). If internalism is true, then there are only reasons for agents that could potentially motivate them. This gives us good reason to think two things: first, that instrumentalism is true, and that there are no moral reasons: reasons invariably supplied by moral norms. To see the first, consider this question: what could motivate one to do something if not one of her ends? To see the second, consider the possibility of the dedicated immoralist who remains entirely unmoved in the face of a moral norm that applies to him: he remains unmotivated by the supposed reason that he has, thus on internalism's view, he does not really have a reason to act morally.

In chapter 6, we consider an interesting argument that may be advanced in favour of the view that all normative reasons are instrumental. The argument attempts to show that instrumental reasons are conceptually basic: the concept of an instrumental reason *just is* the concept of a normative reason for action. In effect, this is a conceptual argument purporting to show that as a matter of definitional (or conceptual) necessity, all reasons are instrumental: a conceptual route towards a substantive conclusion.

In chapter 7, the investigation into the plausibility of instrumentalism, and by extension the question of moral reasons, ends with a consideration of a constitutivist argument. We shall understand a constitutivist argument as an attempt to ground all normative reasons in what is what constitutive of agency. The interesting claim that we shall consider will be this: aiming at the attainment of one's ends is the single

[9] The error theorist (of chapters 4-7), of course, does more than this. She also denies the thought that moral norms invariably supply agents with *pro tanto* normative reasons (see chapter 8 for an alternative view).

constitutive aim of agency, and hence agents as such have only reason to aim at the attainment of their ends. In other words, there are only instrumental reasons. This is effectively an attempt to pick out instrumental reasons and to infer, based on considerations about agency, their normative supremacy.

1.4 Conceptual Analysis

We now proceed by introducing a method for conceptual analysis which will be used in the next two chapters, where the error theorist's account of moral requirements is developed and defended. Conceptual analysis is at the heart of the analytic philosophical tradition, and an essential tool in the philosopher's toolkit. It is thus worth understanding its nature along with some of its limitations. I begin by (1.4.1) distinguishing carefully between terms and concepts. I then (1.4.2) introduce the classical theory of concepts very briefly, which expresses the traditional view about the structure of concepts. I then (1.4.3) sketch a particular method for conceptual analysis which I shall adopt. I close the section by (1.4.4) addressing briefly the paradox of analysis, and (1.4.5) considering the possibilities of defective analyses and defective concepts.

1.4.1 Terms and Concepts

I start by distinguishing between terms and concepts.[10] A term is an element of language; e.g. a sound, a word or phrase. By contrast, a concept is an element of thought; e.g. a particular meaning associated with a given term. As it transpires the two kinds of things are closely connected. When it is meaningful, a term expressed a concept, and the concept expressed is the meaning of the term. We may also say that a term can be defined, whereas a concept can be analysed. In analysing a concept, we attempt to gain further insight into the concept; to grasp a better understanding of it (more on this below). Note that a term may be meaningless, that is, when it fails to express an actual concept. An example is Lewis Carroll's term 'Jabberwocky', which serves as the title to his famous gibberish poem.[11]

Two potentially useful stipulations are in order. First, different terms may express the same concept. For example, the english word 'cat' and the french word 'chat', though different terms, are commonly taken to express the same concept: that of a feline mammal. Second, a particular term can have different meanings in different contexts. For example, the term 'cat' generally expresses the concept of a feline mammal, while in certain underground contexts it expresses the concept of a male person who is thought to be cool.

In what follows, I will be primarily interested in concepts, and the discussion will take the following format: I will start by introducing a concept, which I will then attempt to analyse. In order to express the concept in question, I will make use of a particular term. That term may be technical; used by myself (or others) to refer to a particular concept, without it being necessarily the case that ordinary people use it in the same way. This will not be a problem, as it will not be my intention to claim that the term in question belongs to the public domain. Rather, my claim will be that the

[10] I will sometimes refer to concepts by using expressions in small caps.
[11] from *Through the Looking-Glass, and What Alice Found There* (1871)

concept, to which I will be calling attention, does so. That is, following my analysis of the concept in question, I shall claim that the concept belongs — so to speak — to ordinary thought, though it may be ordinarily expressed by one term or another. To use the cat/chat example, it is clearly the case that both english and french speaking populations possess the concept CAT, though they use different terms to express it.

1.4.2 Drawing on the Classical Theory of Concepts

Following the classical theory of concepts, I understand an important subset of concepts as structured complexes (Margolis and Laurence 1999: 5). According to the classical theory of concepts, some concepts C have definitional structure in that it is composed of simpler concepts that express necessary and sufficient conditions for falling under C.[12] To give an example, consider the concept KNOWLEDGE. On one definition, KNOWLEDGE has the constituents BELIEF, TRUTH and JUSTIFICATION. The latter are understood as simpler concepts that make up the concept of knowledge, while something falls under KNOWLEDGE, or is said to be an instance of knowledge, if it is a justified true belief and only if it is a justified true belief. Though 'structure' here may be understood in different ways,[13] this rough description of what a concept is should suffice for our purposes. Another thing to note here is that an analysis of a concept C will naturally lead to a conceptual account of C, the latter understood as the summary or encapsulation of the essential parts of the concept.

The above understanding of concepts as structured complexes may be taken to suggest a certain method of conceptual analysis. In general, analysis may be understood as the breaking down of something into its components. Thus we may understand conceptual analysis as the process of decomposing a concept down to its fundamental components. Following Beaney,[14]

> [A]n analysis embodies a definition; it specifies a set of conditions that are individually necessary and jointly sufficient for the application of the concept.

An important thing to note about conceptual analysis is that it is supposed to be a distinctively *a priori* activity — that is, to be conducted from the proverbial armchair. Following Zangwill,

> To say that a truth is a conceptual truth is to say something about the way we can know it. What gives a truth conceptual status is the fact that it is knowable 'by conceptual means' (Zangwill 1994: 211).

Indeed, an analysis of a concept involves drawing certain inferences which can be best seen as 'movements in the mind'. This metaphorical way of putting it is meant

---

[12] Margolis, Eric and Laurence, Stephen, 'Concepts', *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition)
[13] (Margolis and Laurence 1999: 5)
[14] Beaney, Michael, 'Analysis', *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition)

to capture the idea that, whereas empirical inferences require perceptual observation, the type of inference which conceptual analysis should be concerned with requires 'the mere operation of thought' (Hume 1748/2007: 18). For example, in the context of conceptual analysis, *mere* reflection on BACHELOR is supposed to be enough for a competent user of the term to infer UNMARRIED and MAN. Another thing to note here is that to say that a truth is knowable by conceptual means *alone* is to specify a restriction on the method that the relevant investigation can adopt. It is to say that someone who seeks to establish a conceptual truth by means of, say, conducting a popularity poll is probably doing something wrong.

By breaking a concept down to its fundamental components, our understanding of it may be thought to be enriched. There is, however, a problem in this idea; what is often referred to as *the paradox of analysis*. In a nutshell, if UNMARRIED and MAN are somehow *contained* in BACHELOR, then the relevant analysis (i.e. BACHELOR means unmarried man) must be obvious and thereby *un*informative. But given that analysis is supposed to give us some sort of *insight* into the concept in question, how can it be that it is at the same time *un*informative?[15] The method of conceptual analysis to be adopted here is not susceptible to the paradox of analysis (to be explained in 1.1.4), and this must be considered an independent virtue of the view.

1.4.3 Smith's Method For Conceptual Analysis

I shall now sketch and adopt a particular method of conceptual analysis, which has been championed by Michael Smith in *The Moral Problem*. As I understand it, Smith's method is true to the spirit of the classical theory of concepts in that it understands concepts as consisting of simpler conceptual items. What is distinctive about Smith's method is its proposal about the exact nature of those simpler conceptual items. In a nutshell, Smith understands the constituent parts of a concept C to be certain platitudes associated with the concept. A particular a priori method is then suggested for analysing a concept: 'an analysis should give us knowledge of all of the relevant platitudes about [the concept in question]' (1994: 31). Before elaborating on the present suggestion, effort must be taken to explain what is here meant by 'platitude'.

In ordinary language, a platitude is a statement treated as *obviously* true; accepted immediately without need for justification: a truism. In the context of the present analysis, and in a similar vein, a platitude surrounding a concept C is a statement the truth of which everyone that understands C will be immediately willing to accept. Importantly, and consistent with the *nature* of conceptual analysis discussed earlier, what makes a statement a platitude is the basis upon which people accept it; it must be accepted on conceptual grounds alone. In other words, if true, the statement must be an *a priori* truth (Smith 1994: 22). This last stipulation separates the ordinary use of 'platitude' from the specific one used in this context. For a statement may be deemed a platitude without it being regarded a conceptual truth. An example of what may be called an *empirical* platitude is the statement that the chemical composition of water is H_2O. From now on, I shall use 'platitude' and its derivatives to refer exclusively to *conceptual* platitudes.

[15] The idea traces its origins to Meno's paradox of inquiry. See (Sorensen 2003: 58-70) for a nice introduction, and (Langford & Schilpp 1943) for a more detailed discussion.

Consider now the following two platitudes surrounding the concept WITCH:

(1) A witch is a person who has magical powers
(2) An old woman wearing a hat is not necessarily a witch

Following Smith, a person comes to master the concept WITCH, and by extent the vocabulary referring to the concept, by coming to treat remarks such as the above as platitudinous (1994: 30). Indeed, anyone who can be considered a competent user of the term 'witch' must be willing to accept (1) and (2), and to suppose otherwise is to manifest a failure to understand the concept (1994: 22). It 'would be exhibiting a conceptual confusion: a failure to grasp the nature of the relevant vocabulary' (Blackburn 1984: 65). Or to use yet another useful expression, denying that (1) and (2) are platitudes about witchcraft 'is to change the subject away from' witchcraft (Zangwill 1994: 211).

Conversely, think of someone who accepts the following remarks as true *a priori*:

(1*) A witch is a female person who looks mean and wears hats
(2*) Magical powers are not required for person to be considered a witch

Someone who accepts (1*) and (2*) as platitudes cannot be considered a competent user of the term 'witch', or in more ordinary language, we may say that such a person does not know what the term means. To sum up the points thus far:

> As the project of conceptual analysis is being thought of here, then, an analysis of a concept is successful just in case it gives us knowledge of all and only the platitudes which are such that, by coming to treat those platitudes as platitudinous, we come to have mastery of that concept (Smith 1994: 31).

A clarification at this point is in order. Suppose we come to analyse the concept WITCH by reference to a long list of relevant platitudes. The list can get *too* long, if we include (as we should) platitudes such as:

'A witch may not necessarily be of Dutch origin'
'A witch may not necessarily have children'
'An unmarried woman may still be a witch'

Note that each and every example of a relevant platitude specified is indeed constitutive of competence with the concept WITCH, while at the same time the list could go on and on if we, say, substitute 'Dutch' with any other ethnic denotation, or 'children' with any other relevant term. It becomes apparent that John, a competent user of the term 'witch', cannot memorise, or be able to deduce without careful thought, such a long list of relevant platitudes. But if doing so is thought to be necessary for John to be considered a competent user of the term, a paradoxical result may then be thought to follow; that John cannot be considered a competent user of the term that, *ex hypothesi*, he has already mastered. Moreover, if we instead allow that

John is, after all, not a competent user of the term given his inability to memorise all the relevant platitudes, we seem to be left with no good reason to suppose that anyone else could fare better. It then seems that our analysis of the concept WITCH implicates a concept that is seemingly impossible to master! If we do think all this, something must have gone wrong in our understanding of the present method of conceptual analysis.

Smith considers the possibility that competent users of a given term may not be able to produce the relevant platitudes 'off the top of their heads' (1994: 30). But that does not prompt him to think that such persons are not competent of the term in question. Rather, he explains:

> The point is not that if we have mastery of the word 'red' then we are able to produce a long list of remarks like these off the tops of our heads. That may or may not be true. The point is rather that these remarks capture the inferential and judgemental dispositions vis-a-vis the word 'red' of those who have mastery of that term, whether or not they are able to produce them off the tops of their heads. (ibid)

Indeed, the above clarification can be thought to be implicit in the initial formulation of a platitude provide earlier: a statement the truth of which everyone that understands C will be immediately *willing to accept.* To be willing, or disposed, to accept a statement as a platitude — that is, *after* becoming aware of it — is different than to be able to infer it without careful thought. John may be *disposed* to infer all the relevant platitudes from the concept WITCH, without it being necessarily the case that he actually does so. In other words, John may possess the inferential disposition WITCH → MAGICAL POWERS without him necessarily being aware that he is so disposed. And when John *actually does make* the inference, inferring that 'A witch is someone who has magical powers', the inferential disposition is *manifested* in John's thought. Hence a platitude in the context of the present analysis may be best seen as a type of *inference* which users that are competent with C are only *disposed* to be make.

1.4.4 Addressing the Paradox of Analysis

The above clarificatory remarks can be used to explain why Smith's method of conceptual analysis is immune to the paradox of analysis. Recall the paradigm example: if UNMARRIED and MAN are somehow contained in BACHELOR, then the relevant analysis (i.e. BACHELOR means unmarried man) must be obvious and thereby uninformative, but this cannot be the case — given that analysis is supposed to give us some sort of insight into (i.e. be informative about) the concept in question.

By taking into account that the proposed method for conceptual analysis understands the fundamental components of a concept C to be certain inferential dispositions possessed by competent users of C, we can explain why such an analysis will be both (in a sense) unobvious and informative. Following Smith:

> Because even though someone who has mastery of some concept C must have certain inferential and judgemental dispositions, it may not be

> transparent to her what these inferential and judgemental dispositions are, and so, *a fortiori*, it need not be transparent to her what the best summary …of the platitudes that describe these dispositions is (1994: 38).

In other words, the possibility that certain inferential dispositions associated with a concept C are scarcely, if ever, manifested in our thought, it may take time and a considerable amount of reflection to figure out what exactly follows from our very own concepts; in other words, what the actual meaning of the concepts that we can be said to have already mastered is. Thus an analysis which makes reference to such inferential dispositions may seem at first unobvious to us, yet after careful examination it may turn out to be true; in other words, it may turn out to be *informative* about our own concepts.

1.4.5 Defective Analyses and Defective Concepts

We must also allow for both the following possibilities. First, (i) a conceptual analysis can be defective, and second, (ii) a concept itself can be defective. By way of examples, I shall elaborate on both possibilities.

(i) Suppose we seek to analyse WITCH. An analysis can be incomplete in case there is an inferential disposition possessed by competent users which is missing from the relevant list of platitudes; e.g. 'A witch is a female person'. Indeed, anyone who has mastery of the concept must be willing to accept the inference WITCH→FEMALE. Consequently, an analysis that omits specifying the aforementioned inference must be considered defective. We may say that such an analysis under-specifies the concept in question: makes reference to less platitudes than it should. And the reverse problem is possible, in case the analysis includes an inferential disposition that does not represent a platitude; e.g. 'A witch is a person of Dutch origin'. A person who has mastery of the concept is not required to accept the inference WITCH→DUTCH ORIGIN. This time, the analysis over-specifies the concept in question: makes reference to more platitudes than it should.

(ii) A concept itself may also be defective, and there may be more than one ways that it can be so. I have already discussed the concept WITCH, and Joyce considers the concept PHLOGISTON. Both concepts can be said to be defective for the same reason. Take the latter, which featured centrally in superseded scientific theories that sought to explain the phenomenon of combustion, and consider a relevant analysis of phlogiston which is conducted partly by reference to these platitudes:

> 'Phlogiston is some kind of stuff'
> 'Phlogiston is stored in bodies'
> 'Phlogiston escapes during combustion'

An analysis of PHLOGISTON based on the above platitudes then reveals that the following is true of the concept: Phlogiston is some kind of stuff that is stored in bodies and is released during combustion. Now, having established what our concept PHLOGISTON is, we may ask the question whether there is anything in the world that corresponds to it. If the answer is no, which the revolutionary chemist Lavoisier knew

it was, we may then conclude that the concept is defective in that it is never instantiated.

But there seems to be another way that a concept can be said to be defective. Smith considers such a way:

> We would … change our inferential habits if we were shown that the judgements and inferences that we make as masters of [a given concept] — the platitudes themselves — were *incompatible or inconsistent with each other* (1994: 30, emphasis added).

Smith here is calling attention to the possibility of a concept having inconsistent parts: defined by reference to inferential dispositions that are inconsistent with each other. An example could be the concept of a round square. There seem to be at least two things to infer *a priori* from ROUND SQUARE: the concept of something being a square, and the concept of something being round. Insofar as we think the two concepts are inconsistent with each other — e.g. that something cannot be both at the same time — we may infer that ROUND SQUARE is defective. Notice that whereas PHLOGISTON and WITCH are deemed defective solely on empirical grounds, even though the platitudes surrounding the concept are seemingly consistent with each other, the defect presently under consideration involves no such empirical test but a test which is conducted solely through conceptual means.

Notice also that there need not be anything wrong with the conceptual analyses of PHLOGISTON and WITCH, on the one hand, just as there need not be anything wrong with the conceptual analysis of ROUND SQUARE, on the other. In fact, all above mentioned analyses seem to do a pretty good job of capturing our inferential dispositions regarding the concepts in question; in short, they seem to capture what the relevant terms *mean*. In general, an analysis of a concept C can be successful, and correspondingly a conceptual account resulting from it can be true, despite it being the case that the concept in question is itself in either way defective.

1.5 Summary

This chapter has introduced us to the question that will primarily concern us in this thesis: do moral norms invariably supply agents with sufficient normative reasons? In the next chapter, we use the method of conceptual analysis introduced here to arrive at a sound account of moral requirements — on which moral requirements have little or nothing to do with promoting our own interests and ends. Establishing this will license a critical examination of the theory of reasons that the error theorist proposes: practical instrumentalism. For having fixed on the nature of moral requirements, an answer to the thesis question hinges on an answer pertaining to the correct theory of reasons.

2. Moral Error Theory

2.1 A Distinct Dimension of our Experience

We look at things from various perspectives and try to assess them based on various considerations. We think that the president's speech was well-written and the way he delivered it rather captivating. It contained just the right amount of words to make his points without being wordy, and those fuzzy words whenever he wanted to get away with something. A smooth operator, we think, but a rather *dishonest* one. For when we concentrate on the meaning of what he said, he seemed to contradict his previous statements, give a false account of known facts, and let go of commitments he made in the past. From the perspective of writing speeches and delivering them, the president scored high enough to merit a positive assessment. At the same time, from the perspective of ethics or *morality* (see 1.1), he did not do that well.

Morality is a distinct dimension of our experience, and it exercises its own sovereignty (Dworkin 1996: 128). At times, it may be that our moral thoughts or reactions are seemingly incompatible with some other aspect of our experience. For instance, we may think that there is something *morally* wrong with dwarf-tossing, while thinking at the same time that it is aesthetically pleasant. We may think that even though abortion favours the woman's right to bodily autonomy, it is still morally wrong — and in thinking so, we are inclined to feel an aversion towards it. Most of us have a reliable tendency to refrain from performing acts that we deem morally wrong, and we expect others to do too. Indeed, as a matter of fact, most people strive to be moral, and most of those who do not try to conceal their indifference to the moral way of life — as moral failure is the kind of thing that speaks deeply to ordinary perception. We feel disgust in the face of those actions that we deem the worst from a moral perspective, such as genocide. We condemn whoever is responsible for them, and we want to avoid them at any cost. We expect others to feel as appalled by such acts as we do. With respect to actions the moral status of which is less obvious to us, we may wonder whether they *really* are wrong. We strive to find answers to questions pertaining to the moral status of actions, because morality matters to us.

Our characteristic ways of thinking, acting and reacting when engaging with the moral dimension are based on a set of beliefs about morality. For example, when we judge that genocide is morally wrong not because we have agreed that it is, but rather because *it is a fact* that it is wrong, we seem to be assuming that the moral status of actions is (in a certain sense) an objective matter. Or when we construct thought experiments involving difficult ethical choices, such as the choice between not doing anything and allowing a train to kill five people or pulling the leveller and bringing it about that it kills only one, we seem to be assuming that such answers are to be arrived at through correct reasoning, rather than agreed upon by considering our collective interests.

2.2 Moral Error Theory: General and Specific

Moral error theory comes as a threat to the beliefs and assumptions that are ingrained in ordinary moral thought. It purports to attribute an error to ordinary moral thought:

to say that contrary to what people think, it is not the case that (say) genocide is wrong in an objective sense, or that there are correct moral answers that are to be arrived at through correct reasoning. We may regard moral error theory as an objection which is directed at ordinary moral thought.

Like any careful objection, a well-thought-out moral error theory first needs to provide an accurate depiction of its target. Though 'target' here may be taken to mean ordinary moral thought as a whole, I use it to refer specifically to that particular thought which is deemed by moral error theory to be problematic and ultimately false. The identification of that idea, call it X, and the subsequent charge of it with error, licenses the error theoretic punchline 'Ordinary moral thought is flawed because of X'. We may call the first step of moral error theory the *fixation-thesis*, and the second step of moral error theory the *denial-thesis*. In short, the fixation-thesis typically says something to the effect that ordinary moral thought is committed to X, or that X is an indispensable part of it, while the denial-thesis is typically a statement to the effect that X is false.

2.2.1 The Many Moral Error Theories

Historically, moral error theorists have identified a number of ideas that are thought by them (and others) to be part of ordinary moral thought and which (only them) take to be flawed. John L. Mackie, a prominent moral error theorist, identified a number of such ideas in *Ethics: Inventing Right and Wrong*. By focusing on one such idea, it is possible to develop specific versions of moral error theory. In what follows, we look at an overview of Mackie's many moral error theories.

Mackie began by noting that moral judgements are thought to be 'at once prescriptive and objective', and that this is an indispensable part of ordinary moral thought (Mackie 1977: 23). What Mackie meant by the double characterisation ('objective' and 'prescriptive') may be captured in the following example:

> Suppose we encounter an isolated civilisation living on some distant island in the Pacific, and that we find out that they have no special reservations toward the act of rape — though they tend to be distressed in the event of a rape victim trying to escape sexual assault by punching the attacker in the head. We tell them that rape is wrong, and that *as a result of this* they *should not* behave in such ways.

We think this because we take moral facts to be *prescriptive*; the classification of an act as morally right or wrong is taken to be immediately relevant to what an agent should or should not do (Garner 1990: 139).

> The strangers then respond that *to them* rape is not morally wrong. As an act of politeness, we agree that our moralities (theirs and ours) are just different — at least in regard to their stance toward rape. That is, perhaps in a similar way that the partisans of basketball and football differ in the stance toward handballs: the former accept is as a legitimate act, whereas the latter condemn it as wrong.

But what we do not say to them, we nevertheless think and feel deeply about: that morality is more than just subjective — or arbitrary — opinions on how we choose to classify acts in terms of right and wrong. We think that our morality is the *right* morality, and that they are in some sense *mistaken* about their ethical beliefs. We think this because we take morality, unlike football, to be an *objective* matter. We may say that moral values are both prescriptive and *objective*.

Based on his understanding of moral values as objectively prescriptive, Mackie called attention to four ideas which he took to be part of ordinary moral thought, and which he crucially thought to be problematic. The first concerns the way that we come to know objective moral values: objective moral values are knowable. That is, there must be a way of knowing what the morally right thing to do is, and as a matter of fact some people possess a considerable amount of moral knowledge, while some others do not fare as well. Mackie looked at this first idea with suspicion, and felt that a denial-thesis may well apply to it: it would take 'a special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else', to be aware of objective values (Mackie 1977: 38). Things would be rather straightforward if moral values were thought to be subjective or intersubjective. That is, if they were supposed to be founded on things like thoughts or desires that we happen to have or ones that everyone in our community happens to have. For we are in a good position to know what we desire (e.g. introspection), or what others happen to desire (e.g. by asking them, or conducting a poll). Insofar as the idea that moral values are both objective and knowable in principle is part of ordinary moral thought, an explanation of how we come to know them may be necessary for making sense of the idea.

The second idea that Mackie called attention to concerns the way objective moral values relate to the natural world: objective moral values must somehow relate to the natural world. Indeed, it seems to be a conceptual commitment of ordinary moral thought that the moral is somehow tied to the natural. For how else would we be able to explain the intuitive thesis that two states of affairs which are identical in their natural properties are also identical in their moral properties. For example, consider the action of hitting someone in the head as a means of self-defence, and the action of hitting someone in the head in order to cause deliberate harm for fun. Intuitively, only the second instance of violence is obviously wrong. Note that this must be true not just because we say so, or because we are disposed to approve of the one and not the other. If that were the case, moral values would be subjective and not objective. Rather, it must be true because of some *natural* feature of the situation itself. Again, after considering this ordinary commitment, Mackie wonders sceptically: 'What is the connection between the natural fact that an action is a piece of deliberative cruelty … and the moral fact that it is wrong?' (Mackie 1977: 41). Mackie thinks that 'the wrongness must somehow be 'consequential' or 'supervenient'; [that] it is wrong because it is a piece of deliberate cruelty', but he then wonders 'what in the world is signified by this 'because'?' (Ibid). Again, insofar as the idea that moral values are somehow related to the natural world is part of ordinary moral thought, an explanation of how exactly they relate to it may be necessary for making sense of the idea.

The third idea relates to moral motivation: mere recognition of moral values (e.g. judging or realising that kicking dogs is morally wrong) would *ensure* that one

abstains from (say) kicking dogs, 'overruling any contrary inclinations' (Mackie 1977: 23). In thinking that the aforementioned idea is an indispensable commitment of ordinary moral thought, Mackie suggested that the latter is committed to something like motivational internalism: the thesis that the recognition of the truth of a moral judgement necessarily or intrinsically motivates (Brink 1984: 113). Unlike the first two ideas, whether this one is a conceptual commitment of ordinary thought is dubious (see below). But as with the previous two, on the assumption that the idea that mere recognition of moral values guarantees motivation is part of ordinary moral thought, an explanation of how exactly moral values motivate in such a rigorous manner may be necessary for making sense of the idea.

Arguments for moral error theory fixing on any of the above ideas tend to be persuasively criticised. Beginning backwards, one must question whether something like a commitment to motivational internalism really is part of ordinary moral thought. Motivational internalism is a thesis that is 'both ambiguous and easy to criticise' (Garner, 1990: 142). For one thing, the idea that moral values are intrinsically motivating may be held by some people while not by others, and this is likely to lead to an impasse (Joyce 2001: 20). Crucially, both groups of people may well be regarded as competent when it comes to moralising, and this strongly suggests that commitment to motivational internalism is far from constitutive of competence with matters moral. Furthermore, even if a commitment to motivational internalism really was a requirement for competence in moral matters, the fact that humans often remain unmotivated to pursue what they judge to be good would hardly show that there was anything problematic about ordinary moral commitments. At best, the latter would point to a certain inadequacy in the human psychology, as opposed to any incoherence or falsity exhibited by ordinary moral concepts. Compare: we think that good strikers absolutely and unconditionally should score from a penalty kick; but as a matter of fact, some of them occasionally do not. Would the latter be an indication that our ideas about good strikers ought to be revised, or would it be an indication that strikers in general are after all not as good as we might have initially thought?

In dealing with the remaining two ideas, opponents can point to other domains of thought in which similar ideas can be identified as conceptual commitments. If this can be done successfully, moralists can admit that these ideas may be hard to make sense of, but that this is not a problem that is peculiar to moral thought. This claim is part of a wider strategy that effectively amounts to 'seeking companions in guilt' (Lillehammer 2013). The claim has the interesting result that either both domains are embedded with error or that neither is; in any case, an error theory that fixes on them is not an objection to morality *per se*.

Addressing the first objection, opponents can refer to our knowledge of things the acquiring of which is not thought to presuppose the existence of a 'special' faculty. Mackie himself considers a few such examples in *Ethics* (39), and I shall mention here our knowledge of logic. Logical truths, it seems, are as causally inert as objective values — and what is more, our grasping of them seems to depend crucially on self-evident intuitions. That is, we do not seem to infer such truths from more basic intuitions. If these considerations do not prompt us to postulate a 'special' *logical* capacity, why should they prompt us to postulate a 'special' *moral* one? More importantly, if these considerations are not taken to suggest that there exists no

knowledge about logic, why should they be taken to suggest that there exists no moral knowledge? As Lillehammer put it, 'as for logic, so for value' (Lillehammer 2013).

A similar strategy may be employed to defend the second idea: the dependence of the moral on the natural. In fact, there seem to be more than one option to do this — with some options being more intuitive than others. One way to explain the alleged connection between moral values and the natural world is by claiming that moral properties are identical or reduce to natural properties. In other words, to claim that moral wrongness just is (say) infliction of deliberate pain. This explanation, however, threatens to make moral properties explanatorily redundant, and as such is often considered a non-starter. That is, if all we mean by 'moral wrongness' is (say) 'deliberate infliction of pain', why do we not say so? Perhaps a more intuitive way to account for the relation between the moral and the natural is by claiming that moral values supervene on natural properties in a similar way that sociological facts supervene on psychological facts, and psychological facts supervene on biological facts. If social scientists and psychologists are not pressed to say anything more interesting to explain such connections, why should we give moralists a hard time? The error theorist may still press that, unlike thought characteristic of these other domains, moral thought postulates brute connections: that are insufficiently explained and thus we have no good reason to accept (McPherson 2012: 228). Depression supervenes on certain physical states, and neuropsychology provides us with good enough explanations as to why this is the case. On the other hand, this line goes, there are no good explanations as to why moral wrongness supervenes on (say) causing deliberate pain. Opponents, in turn, can respond to this by turning to first order ethical explanations, effectively suggesting an analogy between explanations from normative ethics and other sorts of explanations which are not considered brute (Olson 2014: 97). An example of a first order ethical explanation may be this: Causing deliberate pain is morally wrong, i.e. the property of moral wrongness supervenes on the property of causing deliberate pain, because it amounts to harming someone without bringing about any worthwhile consequences.

In what follows, we take on the task, on the error theorist's behalf, of working towards a moral error theory which centres on normativity. This involves arriving at a fixation-thesis. My intention is not merely to *assume* that the fixation-thesis is true, but rather to convince the reader that it is. To do this, I shall call attention to some distinctive aspects of ordinary moral behaviour and then work towards an interpretation of them.

2.3 Ordinary Moral Behaviour

This section is devoted to presenting what may be called *linguistic evidence* for the claims pertaining to ordinary moral thought which follow in 2.4: *how* we use moral concepts, distinctive verbal utterances of participants in moral discourse, as well as some of their characteristic ways of treating transgressors. Based on these, we will attempt to infer in 2.4 the concepts accounting for such behaviour. I begin here by acknowledging two simple observations, each relating to a particular feature of ordinary moral behaviour. Note that these are by no means supposed to constitute an exhaustive account of ordinary moral behaviour. Rather, they are merely the ones that seem to underlie Joyce's account of ordinary moral thought, which in turn underlies his argument for moral error theory. I shall label the observations as follows: (i) the

apparent unconditionality of moral judgements, (ii) the practical 'oomph' of moral judgements. Whereas the former will be found to be not strictly peculiar to moral behaviour, the latter will be deemed so.

2.3.1 The Unconditionality of Moral Judgement

We may start by noting that when making a moral judgement, we are mainly concerned with what morality wants — so to speak — as opposed to what the agent wants. Consider the unlikelihood with which we hear statements like these:

'Stop torturing the poor man on the street, *if* you want to.'

'It is morally wrong to break promises, unless one wants to'

'You must not kill your boss — *unless* you have an interest in doing so!'

The above statements strike us as odd — and may even provoke amusement — partly because they are unheard of, but most importantly due to the semantic tension between the first clause of each statement and the second: the unconditionality implicit in the first and the conditionality explicit in the second. That is, regarding the first sentence, whether one ought or ought not torture the poor man on the street — it would appear — is not conditional on whether one wants to or not. This is, in linguistic jargon, the *implicature* of the statement. The tension occurs when the second clause makes it rather explicit that whether one ought or ought not torture the poor man on the street actually depends on whether one wants to or not. This is part of the explicature of the statement, and it comes in direct conflict with the implicature.

Indeed, the following claim seems true: When ordinary people express moral opinions or judgements, they do so while remaining unconcerned about the addressee's personal ends. Information regarding the addressee's ends is typically treated as irrelevant when we make a moral judgements.[16] Imagine being told by a friend of yours that you shouldn't join the military 'because it is wrong to kill people', and you replying: 'But I have a liking for the occasional killing'. Would you expect this last bit of information to have any impact on your friend's moral views? That is, would you expect your friend to change her view on whether killing is morally wrong or on whether it would be morally wrong for you to join the military? Ordinary people are not only unwilling to change their views upon acknowledging such information, I submit, they also find it inconceivable that they could do so. As far as moral opinion is concerned, they think that such information is trivial.

Note that the observation is not undermined in the least by the fact that we are sometimes unsure whether what we think is a true moral judgement actually is so. In cases when our confidence in our moral judgement falls drastically, we may sound reluctant or less than passionate about our moral convictions. We may say, for instance: '*Perhaps* you should not join the military'. In such cases, we may not be sure whether it would be appropriate to advice our friend to (say) join the military — precisely because we are not sure (in this example) whether joining the military really is morally wrong. But our reluctance to express a strong moral opinion is not due to

[16] (Joyce 2001: 99)

any intention on our part to take into account our friend's ends before arriving at our final moral judgement. Lack of strong moral conviction here must not be mistaken for conditionality.

It is noteworthy that things are often different when we give non-moral advice to someone. There are cases, that is, in which we may be willing to treat information about the addressee's ends as relevant, if not *highly* relevant, and correspondingly form or revise our initial judgement in the face of it.

- It is raining; you shouldn't go out tonight.
- But I find it extremely pleasant whenever I get soaked while going somewhere.

The above exchange may be odd, but how much odder would it be if we insisted our friend still not go out tonight? In general, there are times that we treat information pertaining to the addressee's ends as highly relevant when making a judgement about what she should do. In such cases, information about the addressee's ends tends to shape decisively what we think she should do. Such advice is personal in the sense that it is solely concerned with the addressee's ends. By contrast, we might add, moral advice is impersonal in that it is typically unconcerned with the addressee's ends.

Now, we must note that the apparent unconditionality of morality — thus construed — is by no means an exclusive, let alone striking, feature of moral judgements. Philippa Foot points to the imperatives of etiquette which do not 'fail to apply to someone who … does not care about what, from the point of view of etiquette, he should do' (Foot 1972: 308). In other words, we also do not treat information pertaining to the addressee's ends as relevant when making certain non-moral judgements. This is confirmed by our familiarity with sentences such as the following:

'You ought not to speak with your mouth foul during dinner.' (as opposed to '…, unless you want to insult our guests')

'Go on, son, take the penalty and give us the title!' (as opposed to '…, if you feel like it!')

Indeed, whether the player wants to take the penalty or not, we think, is not immediately relevant to whether he ought to. The upshot here is this: by being an observation that is not exclusively about moral utterances, the present observation is unlikely to reveal anything peculiar about the commitments of ordinary moral thought. In her search for a potential peculiarity, the error theorist may move to the second feature of ordinary moral behaviour.

2.3.2 The Practical 'Oomph' of Moral Judgement

The starting point of this second observation is this: there seems to be a striking difference in the way we treat the demands of morality, on the one hand, and in the way we treat the demands of etiquette and those of any other normative framework, on the other.

Before I elaborate, it would be helpful to start by noting some further related similarities regarding our way of dealing with both. Consider the following case: Sid Vicious is at the Royal Palace having dinner with Queen Elizabeth II. Suppose Sid is not particularly fond of good manners, and that he has a reliable tendency to disobey etiquette. Even though we are well aware of Sid's indifference to the rules of etiquette, wouldn't we still insist that he'd better behave at the queen's table (and for that matter, at anyone's table)? Following Joyce, 'irrespective of a person's feelings for etiquette, [we think or say that] he or she ought not speak with a full mouth' (Joyce 2006: 61). Similarly, in the case of morality, even though we may be aware of a sadist's indifference to the rules of morality, we would still insist that he or she ought to behave morally.

Note that the above point is slightly different from the one previously mentioned regarding the apparent unconditionality of our judgement that 'It would be rude for Sid to speak with his mouth full'. Granted, we do not treat information about Sid's ends as relevant when we make an appraisal about etiquette; just as I do not treat information about your ends as relevant when I say that, according to the game I invented five minutes ago, you must read each line twice in order to win five points. That is, whether you want to play the game or not seems to be irrelevant to whether, according to the game, you should reach each line twice. At same time, however, it is unlikely that I would *insist* that you read each line twice — for it would be insane for me to do so. I would be willing to allow, in other words, for what I acknowledge to be a breaking of the rules of the game. The point stressed above concerns our *insistence* that transgressors change their ways — and this insistence is something that *often* characterises both our attitude towards morality and etiquette.

But there is a particular difference that can be considered striking and therefore of potential interest to the error theorist. On the one hand is our attitude toward etiquette: In short, we not only allow for the occasional breaking of the rules of etiquette, but we also tend to reward it on some occasions. That is, whenever someone breaks a rule of etiquette in order to (or for) *X*, where *X* signals a special cause that the action of breaking the rule seeks to secure, we tend to say something like 'It's fine!' — if not 'Well done!'. To see this, imagine if Sid spoke loudly with his mouth full in order to warn Queen Elizabeth that she was about to eat a wasp.[17] Granted, in order to spare the Queen from the unpleasantness of eating a wasp, Sid had to break a rule of etiquette. But, at the same time, wouldn't we expect the Queen to say something like 'Good Job, Sid' — rather than emphasising the fact that he broke a rule? The observation here is not that we fail to acknowledge the breaking of the rules of etiquette whenever it means securing a special cause; rather, it is that on such occasions we do not insist that transgressors behave in line with etiquette. On the contrary, on such occasions we tend to encourage the breaking of the rules of etiquette in order to secure that other end that we deem special.

And on the other hand is our attitude towards moral rules: We do not encourage or retrospectively reward behaviour that goes against one's moral obligations — even when it means securing a 'special' cause. Rather, we insist that others obey (what we judge to be) moral requirements *no matter what*. To begin with, there seem to be no

[17] This is the example that Joyce gives to make the current point (Joyce 2006: 61).

special causes when it comes to outweighing the moral cause — and this is the reason I initially put the word in brackets. We treat moral rules — and for that matter, morality itself — as if it is *the* special cause. Furthermore, our experience tells us that it is hardly, if ever, the case that we allow for the breaking of a moral rule for the sake of securing a non-moral one — no matter how important the latter is deemed by us or others. Consider again the previous example, but suppose this time that Sid Vicious sets the palace on fire — thereby endangering the lives of many — for the sake of satisfying his friend's wish of seeing him on tomorrow's front page news. Suppose also that Sid regards his friend as a very special person, and that he regards her wishes as very important. Even if these were the facts, would we expect anyone acquainted with the incident to condone Sid's immoral action — let alone retrospectively reward it? To be sure, the following statement is very unlikely to be said by any serious participant in ordinary moral discourse:

> 'In doing so, you broke a rule of morality — but that meant securing a very important end of yours — so well done!'[18]

In general, a breaking of a moral ought — i.e. obligation — for the sake of securing a non-moral end is hardly ever condoned by ordinary people who seriously participate in moral discourse. This is consistent with the possibility of ordinary people condoning the breaking of what appears to be a moral obligation for the sake of following an *actual* — or perhaps more basic — one. For example, if I think that I am morally obliged to give all of my savings to charity, and I am told by Noam Chomsky that doing so is not really morally necessary, then I may well not do so. In doing this, I am still in line with my moral obligations — it just happens that what I thought was a moral obligation turned out not to be one. In a similar observation, Bernard Williams notes that 'morality encourages the idea [that] only an obligation can beat an obligation' (Williams 1985: 200).

To sum up: there seems to be a striking difference with regard to our attitude towards the breaking of moral rules, on the one hand, and the breaking of the rules of etiquette, on the other.[19] Whereas we occasionally condone rude behaviour, we hardly ever condone immoral behaviour. Morality, we may say at this preliminary stage, appears to possess a serious and special backing when it comes to practical decisions and the assessments of past actions. Unlike etiquette and any other set of rules, morality appears to possess a practical 'oomph'. Following Joyce, 'that morality has practical oomph is a simple observation' (Joyce 2006: 62).

2.4 Ordinary Moral Thought

[18] Joyce notes that a likely response to Celadus' breaking of the gladiatorial rules would be one of disregard, though he notes how unlikely such a response would be to the breaking of a moral rule. By contrasting the case of Celadus to that of Gyges — an aspiring immoralist, Joyce notes: 'But what we are content to say about Celadus we are far from happy about admitting of Gyges' (Joyce 2001: 41).

[19] I take it that, in this respect, our attitude towards the breaking of the rules of any other non-moral framework should be similar. That is, I have trouble thinking of a non-moral framework which would elicit a response of disregard when it comes to violating its rules for the sake of a special cause.

Having listed two basic observations regarding ordinary moral *behaviour*, the error theorist may now ask: What can these tell us, if anything, about ordinary moral *thought*? That is, here we attempt, on the error theorist's behalf, to provide an interpretation of the moral behaviour listed in the previous section in the form of concepts which can account for it. In general, I do think that based on observations about some behaviour, we can infer with some degree of certainty something about the thoughts influencing that behaviour. And in particular, I think that the *above* observations *can* license a hypothesis regarding the contents of ordinary moral thought.

2.4.1 Moral Inescapability

Building on the observation in 2.3.1, we can reasonably infer that our concept of a moral requirements involves the concept of a *categorical* requirement. This point calls for some further terminology, namely a distinction that is exhaustive with respect to imperatives[20]:

> Hypothetical Imperative: 'If you want to Φ, then you ought to Φ'
>
> Categorical Imperative: 'You ought to Φ, irrespective of whether Φ serves any of your ends or desires'

The if-clause is the defining feature of a hypothetical imperative, whereas what both hypothetical and categorical imperatives share is the ought-statement. Through its distinctive logical form, a hypothetical imperative communicates a very specific application-condition for its ought-statement. That is, the application of an ought-statement contained in a hypothetical imperative is contingent on the agent's satisfying the conditions expressed by if-clause. Should the agent to whom a hypothetical imperative is directed fail to satisfy the conditions contained in the if-clause, the ought-statement would also fail to apply to her. Things are quite different with respect to categorical imperatives. That is, the application of an ought-statement contained in a categorical imperative is not contingent on the agent's satisfying any conditions expressed by the imperative — for on the face of it, there are no explicit conditions to be met. An ought-statement contained in a categorical imperative which is directed to an agent, we may say, applies anyway.

The particular subset of practical advice that I earlier called attention to can suitably take the form of a hypothetical imperative. As we have already seen, when we are disposed to retract 'Do not go out tonight; it is raining' upon acknowledging the addressee's desire to get wet, we are effectively giving advice which is willing to take such information as relevant. The form of a hypothetical imperative, specifically its containing of an if-clause, makes such retracting possible. Quite relatedly, the form of a hypothetical imperative can account for the fact that the ought-statement (e.g. you ought not go out tonight) can be escaped insofar as the agent can alter her ends (e.g. by choosing to value one thing over another).

---

[20] In an ordinary sense, an imperative is a mood that expresses an intention to influence the listener's behaviour. But an imperative is here understood as a sentence implicating an ought-statement. This slight modification is useful as it brings a traditional philosophical distinction into the discussion.

By contrast, moral advice can only take the form of a categorical imperative. As we have seen, information pertaining to the addressee's ends is not considered relevant when we make a moral judgement. As a result of this, the ought-statement (e.g. you ought not go out tonight) is never retracted upon acknowledging the addressee's ends. A person cannot, in Joyce's words, 'evade a moral imperative by citing unusual desires or aberrant goals' (Joyce 2006: 60). Following Joyce, 'we express our moral condemnation of wrong-doers using categorical imperatives', and I might as well add, for it seems inconceivable that any non-moral consideration would make us retract it (Joyce 2001: 35). A consequence of this is that moral oughts are inescapable, in the qualified sense that by altering her ends, the agent is unable to escape its application.

To sum up: our concept of a moral requirements is the concept of a categorical, inescapable requirement.

2.4.2 Moral Bondage

As understood so far, inescapability does not seem to be a unique feature of moral requirements. A simple observation regarding our attitude towards etiquette reveals that we also do not take information pertaining to the addressee's ends as relevant when making an appraisal about etiquette. Consistent with the above suggestions, this means that the oughts of etiquette are best understood as categorical imperatives — and as a consequence of this, they are in the above qualified sense also inescapable.

But apart from thinking that moral requirements are inescapable, ordinary people must also think something *further* of them. For they do not only think that moral requirements are inescapable, but considering the observation in (2.3.2), they also *insist* that they be followed *regardless*. The former statement denotes an element of thought — a particular commitment of ordinary moralisers. The latter, however, the insistence, cannot be properly understood as a thought but rather as a pattern of behaviour. That is, we *insist* that others follow moral oughts regardless of the situation, but what is the thought underlying this particular insistence?

Building on the observation in (2.3.2), we may reasonably infer that moral requirements are thought to be *binding*: to bind each and everyone of us. And drawing on the observation in (2.3.1), we may add that it is part of ordinary thought that we are *inescapably* bound by morality: bound irrespective of our desires and interests. This is more or less the proposal advanced by prominent error theorists who concerned themselves with the observation.[21] As a first approximation, the proposal that ordinary thought takes moral oughts to be binding seems promising to explain why ordinary people insist that we follow moral oughts regardless of our desires, and more importantly, without allowing for any exceptions. If bindingness is thought to be a unique feature of moral oughts, then we seem to have a hint toward accounting for the striking difference between morality and etiquette captured in (2.3.2). In a nutshell, we think that moral oughts are binding in a way that those of etiquette are not.

[21] See (Joyce 2001: 31), (Garner 1990).

But unfortunately for this premature proposal, the thought that we are bound by moral oughts is by no means clear — to quote Garner, it is 'important, metaphorical, and misleading' (Garner 1994: 54). Following Garner, 'even though the meaning of 'moral bondage' is unclear, the thought that we are inescapably bound is an indispensable part of the institution of morality' (Garner 1990: 139). In a similar discussion, Joyce imagines a case of a parent addressing his child's question 'But why mustn't I do what's wrong?' by insisting that they 'simply mustn't'. Joyce notes that the idea underlying the parental response — if the insistence is to be considered a response — must be 'an illusive notion' (Joyce 2001: 31).

In general, whenever an idea seems unclear, it is natural to seek an explanation of it — which would ideally uncover a *further* idea that lies behind it and which is more easily to make sense of. Just to be perfectly clear on what the nature of the present quest is, consider the following example. Imagine someone sets a building on fire, and after the police catch him, they ask him why he did it. He responds: 'Because *buhalula*'. The police search for the word 'buhalula' on the web and nothing comes up, so they infer it is not a foreign word, and they then ask him for clarification. The culprit does not say anything, besides repeating that he did it 'because buhalula'. The police hypothesise that 'buhalula' is not just a string of sounds but that it must express a concept in the culprit's head. They then turn to people that know him, seeking clarification on what 'buhalula' might mean. No one can offer any help, and they feel obliged to conclude that the word expresses no concept at all — or that if it does, the concept is probably *nonsense*. The police then feel licensed in wondering exasperatedly: 'What was he thinking (when setting the building on fire)?'

Something similar goes on with our present quest. When people insist in saying things like 'The criminal ought not to have stolen from the bank, even though he got away with it and he now leaves peacefully in an exotic island enjoying his newly acquired wealth', they may be interpret as saying that moral requirements are *binding*. We may grant that in saying so they are employing a certain concept, say MORAL BONDAGE, and we may then endeavour into clarifying what that concept amounts to. We may ask: 'What are we thinking when we say that criminals ought not to have done the things they did, even though they only seem to have profited from doing them?'

A few absurd proposals

In their quest to make sense of the concept MORAL BONDAGE, moral error theorists have entertained various proposals; some of them more absurd than others. A first proposal might be this: when we say (or imply) that we are all bound by moral rules irrespective of whether we like it or not, and in a way different that we are bound by the rules of etiquette, what we mean is that there is an invisible moral cord tying us to the moral path and rendering us unable to escape it. Aside from the obvious reasons this proposal is absurd, it cannot be true simply in virtue of the fact that sometimes we do escape doing what is morally right — thus breaking a moral rule must be deemed at least intelligible by the analysis of the concept.

A second proposal is that MORAL BONDAGE implies that in acting immorally people somehow harm themselves (Joyce 2001: 33). Thus, the idea goes, insofar as *some form* of self-harm reliably follows immoral behaviour, and given that no

reasonable person desires self-harm, then we are 'bound by self-harm' not to act immorally. In turn, there may be different forms of self-harm for us to consider. One form of self-harm relates to supernatural rewards – e.g. An all-powerful being punishes the immoral at this life or the other. A second form of self-harm relates to the sacrifice of inner piece of mind — e.g. by acting immorally, we incur in ourselves the highly unpleasant feeling of guilt (cf. Plato 380 BC/2008: books II and X). And we may also add yet a third form that the proposal can take, namely that in acting immorally people are likely to suffer 'earthly' punishment.[22] In addressing the first form, we must note that it is merely an appeal to magic — thereby by no means a reasonable proposal. In addressing the second, we must note that even if it is the case that the feeling of guilt sometimes follows our acting immorally, it cannot explain our moral condemnation of those who show no reluctance or do not feel in the least bad after having acted immorally. And in addressing the third form, we must note that even when someone can get away with murder without getting punished, we still think that he ought not to have done it.

And we may add — and subsequently reject — yet another proposal, namely that part of what we mean by MORAL BONDAGE is that we are bound to be moral by *morons*. This is based on the absurd thesis that 'among its numerous particles of energy and matter, [the universe houses] some special particles — morons — whose energy and momentum establish fields that at once constitute the morality or immorality [of particular acts]'(Dworkin 1996: 104). Of course, apart from the obvious absurdity of the idea of such particles existing, there is a distinct difficulty with the current proposal: even if it were the case that such things existed, what would it mean to be *bound* by them? That is, could we not reasonably ignore them and act however we wanted?

A *prima facie* reasonable proposal

An obvious proposal may be that when we say that people are inescapably bound by moral requirements, and not by (say) the requirements of etiquette, what we mean is that only the former carry with them a reason — or, that from the two we only have a reason to follow the former. This is the proposal that is taken seriously by prominent error theorists.[23] It is *prima facie* reasonable (no pun intended) given that morality has a lot to do with normativity — and normativity, in turn, has a lot to do with reasons.

In short, the proposal is this: our concept MORAL BONDAGE means or implies that there are weighty, sufficient normative reasons for us to do as moral requirements say. Given that the same is not true of the requirements of etiquette, we can account for the observation in 2.3.2, namely the contrast between our insistence that others *always* be moral and our occasional indifference to the breaking of the rules of etiquette. We insist that others always act in accordance with moral requirements, without allowing for any exceptions, because we think that moral requirements *invariably* supply agents with *reasons*. Conversely, we do *not* insist that others always follow the rules of etiquette, and so we do allow for occasional exceptions, because we do *not* think the same about those requirements.

---

[22] Joyce rightly dismisses all three forms of the current proposal (Joyce 2001: 33).
[23] Joyce (2000, 2001), Olson (2012).

At this point, we need to note that there is a problem with the current proposal as it stands. To see this, we may start by noting that reasons, or reason-talk, is notoriously ambiguous (as explained in 1.2.1). In clarifying reason-talk, we must start by granting that 'only in the normative sense does assent to a reason statement commit one to holding that an action has some consideration that stands in its favour' (Sinclair 2016: 1.1.2). Consider this statement: 'The reason the economic collapse happened was the growth of subprime mortgages'. The growth of subprime mortgages caused the economic collapse, but it did not justify or make it right in any interesting sense (quite the contrary).

And even when it is restricted to its normative sense, reason-talk is still quite ambiguous. In a sense, you seem to have a reason — albeit a very weak one — to read each line twice: to win five points. In other words, the fact that my game requires you to read each line twice seems to be a consideration — albeit *defeasible* — that favours your reading each line twice. At any rate, the fact that my game requires you to read each line twice it is enough *for me to think* that you have a reason to do so. 'Similarly', following Joyce, 'the gladiatorial enthusiast who asserts 'Celadus ought not throw sand' will also assent to 'There is a reason for Celadus not to throw sand' (Joyce 2001: 40). At the same time, it is highly unlikely that either you or Celadus would accept that you have such reasons. It seems, therefore, safe to assume that you must have a different conception of reasons in mind when you beg to differ. The upshot here is that by leaving it unspecified which particular conception of reasons we have in mind, the current proposal seems unclear. Furthermore, if all we mean by 'reasons' in the current proposal is 'reasons [that] are contained implicitly in the rules just in virtue of the presence of 'ought' statements', why do we not also always insist that others follow the oughts of etiquette?[24] For in a rather weak sense, there also seem to be reason for others to follow the oughts of etiquette: the reason is the fact that etiquette requires them to do so.

Following the above realisation, we are seem to be left with two options: we must either concede that the concept moral bondage does not really account for the observation in 2.3.2, as it fails to explain why we treat morality and etiquette in such different ways, or that our concept MORAL BONDAGE really is incomplete, as it may be taken to imply that the reasons of morality and those of etiquette are substantially different while it has no convincing story to tell on how exactly they differ.

Perhaps we can improve the prospects of the current proposal by modifying our understanding of the reasons that moral oughts are supposed to bring with them. A hint comes from Joyce when he notes that:

> It may be legitimate for an onlooker to say 'Celadus has a reason not to throw sand,' but this doesn't provide Celadus with a *real* reason (Joyce 2001: 41, my emphasis).

An enthusiastic follower of gladiatorial combat may think that Celadus has a reason to throw sand, as this is an act required by the rules, but Celadus — who is unwilling to play the game — may think differently. In particular, Celadus may

[24] Joyce considers this difficulty in (Joyce 2001: 40).

reasonably think that he does not have a 'real' reason to do so. By adopting the characterisation 'real' and applying it exclusively to reasons of morality, we may be able to differentiate between the reasons of morality, on the one hand, and those of gladiatorial combat and etiquette, on the other. Thus if MORAL BONDAGE is taken to imply that there are always *real* reasons to be moral, whereas it is not the case that there are always *real* reasons to follow etiquette, we can both account for the observation in 2.3.2 and also spare our proposed concept from being deemed incoherent.

But yet another problem seems to follow the above proposal. In general, but also in this context, 'real' may be taken as genuinely honorific. That is, different reasons may be deemed 'real' by different persons depending on their particular normative alliances. Thus a devoted moralist may well regard the reason to follow the ought of morality as 'real', and the reason to follow the dictates of capitalism as 'not real'. The present sense of 'real' merely denotes the speaker's respect or affinity for morality or whatever the case may be. Now, if it is strictly in terms of this honorific sense of 'real reason' that ordinary people are disposed to explain why they think they are bound by morality (and not etiquette), then they seem to be incoherent. For different people may well regard different kinds of reasons as real (e.g. devoted moralists may regard moral reasons as real, cult members may regard the reasons to follow their cult as real and so on). Why does what *we* take to be real reason bind *them*? Ordinary people, it seems, cannot be that incoherent.

The final adjustment to our modest proposal comes from an intuitive distinction that separates normative reasons into two classes: *pro tanto* or defeasible, and weighty or sufficient. A *pro tanto* normative reason is a consideration that justifies an action but which may be overruled by stronger ones. There may be a normative reason for me to go to my favourite cafe this evening (in virtue of my wanting to), and a normative reason for me not to go to my favourite cafe this evening (in virtue of my having promised to stay home with you), and yet a conflicting normative reason for me to go to church this evening (in virtue of my being a Christian). A sufficient normative reason is one that outweighs all other relevant considerations that are at play, thereby determining what the justified course of action is: the one that is most reason supported.

Moral bondage, the current proposal goes, implies that moral reasons — reasons to obey moral requirements — are weighty or sufficient ones. This modification seems to explain why we sometimes find it appropriate to neglect the requirements (and reasons) of etiquette, while we never find it appropriate to neglect those of morality. For only the latter carry with them — or supply us with — sufficient reasons for action. The error theorist has now a plausible proposal to account properly for the observation in 2.3.2.

2.5 An Argument for Moral Error Theory

We know what moral requirements are supposed to be, because we have at our disposal a conceptual account of moral requirements — which is based on observations about moral behaviour and a careful interpretation of that. We now have the resources to arrive at a promising argument for moral error theory — one which concentrates on

normativity. This argument will be in support of the error theorist's stance on the central question of this thesis: do moral norms invariably supply agents with reasons? The error theorist's answer is 'no'. In its simplistic form, the argument is the following:[25]

> (P1) Moral requirements *are supposed* to apply to agents irrespective of their ends.
> (P2) Moral requirements *are supposed* to invariably supply agents with sufficient normative reasons.
> (P3) There *are* only instrumental reasons.
> (C1) There *are* no moral requirements.
> (C2) Moral norms – which express those alleged requirements — do *not* invariably supply agents with reasons (or it is *not* the case that: moral norms invariably supply agents with reasons).
> (C3) Ordinary moral thought embodies error.

The first two premises, P1 and P2, are *conceptual*: they call attention to what moral requirements *are supposed* or *are thought to* be. They tell us that *if* there are any moral requirements, they must apply to agents irrespective of their ends and supply them with reasons. Effectively, they purport to tell us what our concept of a moral requirement *is*. Compare: our concept of a witch is a woman with magical powers — that is to say a witch *is supposed* to be a woman with magical powers. Therefore, *if* there are any witches, they must be women with magical powers.

Both premises have already been defended in previous sections. I accept these premises, as I believe it is an indispensable part of ordinary thought that, first, moral requirements do not depend on the agent's having a relevant end that might be served by following them, and second, that when one is morally required to perform some action, one has sufficient normative reason to do so. In the following chapter, I defend further these conceptual premises against an attack by Stephen Finlay (2008, 2011).

Both the third premise and the conclusions are *substantive*: they purport to tell us what things there *are* (in the case of C1), or what *is* the case (in the case of C2 and C3). The third premise is a statement of a certain theory of practical reasons: practical instrumentalism. Practical instrumentalism holds that there are only reasons favouring actions that are instrumental to the agent's (for whom they are reasons) ends. To see that the first conclusion C1 follows, observe that for moral requirements to exist, they must supply agents with sufficient normative reasons to perform actions that need not have anything to do with the furtherance of their ends. But given that (P3) there can only be reasons favouring the agent's (for whom they are reasons) ends, it follows that there are no moral requirements. What there *is* is moral norms: certain prescriptions derived from the moral status of actions, and which are placed upon us. Norms which, importantly, do not necessarily supply us with sufficient normative reasons. Hence the second conclusion, C2. Finally, the third conclusion C3 states that the thought expressed by P2, which is part of ordinary moral thought, cannot be defended, as it

[25] I take it that the argument is similar to the one given by Joyce (2001, 42).

finds no place in the correct theory of reasons: practical instrumentalism. Hence ordinary moral thought embodies error.

2.6 Summary and Planning

Moral error theory comes as a threat to ingrained assumptions in ordinary moral thought. The version of moral error theory that I am here focusing upon centres on normativity. In particular, it denies a thought that seems implicit in our understanding about morality: that moral norms invariably supply agents with reasons. The error theorist of this context accepts that this is a thought that is non-negotiable, or an indispensable part of, ordinary moral thought. What she denies is that it can be defended: that it is really true that there are moral requirements, and hence that moral norms invariably supply agents with reasons. This thought is the error that the error theorist attributes to ordinary practitioners in moral discourse. In chapters 4-7, I examine the error theorist's grounds for denying that there are reasons supplied by moral norms. I argue that the error theorist's case is unconvincing, as there seem to be no good arguments for thinking that there are no such reasons. The thought that there are good arguments for thinking this is the error that I shall attribute to the error theorist.

Before we move on to the substantive part, I shall defend the conceptual premises in the error theorist's argument against attacks by so called moral relativists. This is the task of chapter 3. I find this important as a successful defence of those premises ensures that this thesis, which is primarily concerned with examining the thought that moral requirements invariably supply agents with reasons, operates on a firm foundation. For if it were to be found that this thought was never part of ordinary thought in the first place, a work examining whether the thought is true would seem rather trivial. Just think about how much *less* interesting considering the thesis-question would be when knowing that this thought is precious to no one: no one thinks it is true, and no one cares if it is not.

3. In Defence of the Conceptual Claim

3.1 A Conceptual Question

This thesis defends the thought that moral norms – which express alleged moral requirements - invariably supply agents with sufficient normative reasons. That the dedicated immoralist, that is, who has no concern for moral considerations and only ends towards immoral acts, has nevertheless adequate reason to act morally. It does so by considering the plausibility of arguments aiming to disprove it. All the arguments to be considered (chapters 4-7) aim at showing that there are only instrumental normative reasons: ones favouring actions that are instrumental to the agent's (for whom they are reasons) ends. If this is so, and given that moral requirements have little to do with the furtherance of the agent's (to whom they apply) ends, it cannot be true that dedicated immoralists invariably have sufficient normative reason to act morally (see 2.5). Before we examine the plausibility of those arguments, we should first establish that the thought really is shared by the majority of people — or at least those who engage in moral discourse. Doing so is strictly speaking not important to the substantive purposes of ascertaining whether there are moral reasons, but just think how much more interesting would be granting that the thought that is under examination is precious to all of us. This chapter defends the claim that the aforementioned thought is shared by the majority — if not all — of people who think and talk in moral terms.

In terms of error theory, the view that denies the plausibility of the above thought, this chapter defends its conceptual premises, thereby paving the way for its conclusion. So for the error theorist's purposes, defending that the thought is shared by ordinary moral practitioners is important. This thesis maintains that attributing this thought to ordinary people is *not* the error in the error theory. Rather, where error theorists err — on my view — is in thinking that there are good arguments for the all-important claim that all normative reasons are instrumental.

Stephen Finlay (2008, 2011) thinks otherwise. He thinks that the error in the error theory is located in the conceptual premises: those specifying the meaning of moral terms. Finlay thinks, thus, that despite it being true that all normative reasons are 'relative' to the agent's ends, such that an agent has a reason only if she has an end served by acting on that reason, ordinary thought never erred as it never thought otherwise. Here, we look at Finlay's grounds for disputing the conceptual premises of error theory.

3.2 The Conceptual and the Substantive

Stephen Finlay (2008, 2011) thinks that there is 'an error in the error theory', and he makes it clear whereabout in the error theorist's territory we should look for that error. He has no objection to the error theorist's substantive premises and conclusions, though he 'no longer believe[s] that everyday moral discourse embodies a fundamental error' (Finlay 2008: 347). It is in those parts of the argument fixing on ordinary moral thought, Finlay thinks, that we should look for the error in the error theory. In objecting

to the conceptual part of the argument, Finlay is effectively complaining that the error theorist has misinterpreted ordinary thoughts.

As we saw in the previous chapter, an error theory[26] is a view that is made up of two parts: the *conceptual* and the *substantive*. Consider the following argument for moral error theory:

(P1) Moral requirements are supposed to apply to agents irrespective of their ends.
(P2) Moral requirements are supposed to invariably supply agents with sufficient normative reasons.
(P3) There are only instrumental reasons.
(C1) There are no moral requirements.
(C2) Moral norms – which express alleged moral requirements - do not invariably supply agents with sufficient normative reasons.
(C3) Ordinary moral thought embodies error.

As explained in 2.5, the first two premises are *conceptual*, whereas the rest and the conclusions that follow are *substantive*. A conceptual premise calls attention to what something *is supposed* or *is thought* to be. It tells us that *if* there is *X*, it must be *such-and-such*. Effectively, it purports to tell us what our concept of *X* *is*. By contrast, a substantive premise purports to tell us what things there *are* (in the case of C1), or what *is* the case (in the case of C2 and C3). A standard example incorporating a conceptual and a substantive premise is this: A witch *is supposed* to be, or our *concept* of a witch is, a woman with magical powers, but there *are* no witches.

The above argument tells us that moral requirements are *supposed* to both apply to agents irrespective of their ends and invariably supply agents with sufficient normative reasons. This is the conceptual part of the argument. The argument also tells us that there are only instrumental reasons, and from this it infers a number of interesting conclusions (C1-C3). This is the substantive part of the argument.

As our interest here will be in sufficient normative reasons. As I explain in *1.1.2*, in being *normative*, a sufficient normative reason contrasts with motivating and explanatory reasons: a reason *for which* someone does something, and a reason *why* someone does something, respectively. A normative reason for action is a reason there *is* for doing something: a consideration that justifies or supports a prospective course of action. A sufficient normative reason is one that outweighs all other relevant considerations that are at play, thereby determining what the justified course of action is: the one that is most reason supported. In being sufficient, it contrasts with *pro tanto* normative reasons: a consideration that justifies an action but which may be overruled by better ones. In what follows, I may use the unqualified noun 'reason' to refer to sufficient normative reasons, unless stated otherwise.

[26] Unless indicated, and for the sake of brevity, I will use 'error theory' to mean the moral error theory of the kind presented in the previous chapter.

3.3 Finlay's Terminology

Finlay complains that the error theorist has erroneously attributed 'an assumption of the *absolute [normative] authority of moral value*' (2008: 348, my emphasis}. By 'absolute normative authority of moral value', he means:

> Moral discourse assumes that moral… requirements provide or entail reasons for acting that apply categorically (absolutely), independently of one's desires and ends (ibid).

Effectively, this the combination of P1 and P2. The thought that moral requirements provide or entail reasons for acting is the thought (P2) that moral requirements are supposed to invariably supply agents with sufficient normative reasons[27], and the thought that moral requirements entail reasons *that apply categorically (absolutely), independently of one's desires and ends* involves the thought (P1) *that moral requirements are supposed to apply to agents irrespective of their ends.* On Finlay's terminology, anyone who subscribes to the combination of P1 and P2 may be thought of as possessing *an absolutist conception of morality* (or ordinary moral discourse), and anyone who does not subscribe to it may be thought of as possessing *a relativistic conception of morality*. As I take it, the terms *absolutist* and *relativistic* are meant to denote one's commitments on what types of sufficient normative reasons there are. If you are an *absolutist* on Finlay's view, you must think that it is possible that an agent has a sufficient normative reason to perform some action despite it being true that no end of hers is served by doing so. If you are a *relativist* on Finlay's view, you must think that it is not possible that an agent has a sufficient normative reason to perform some action if it is also true that no end of hers is served by doing so.

Note that in our discussion with Finlay, you can be a *relativist* in two ways. First, by thinking that moral requirements are not supposed to apply to agent's irrespective of their ends, and second, by thinking that even though they are, no one thinks that they invariably supply agents with reasons. In both cases, one's commitment to a relativistic nature of sufficient normative reasons is preserved. That is, by thinking either, one is not committed to the existence of sufficient normative reasons that have nothing to do with the furtherance of the agent's (for whom they are reasons) ends. This observation takes us straight to the next point.

3.4 Degrees of Eccentricity

It is open to anyone to attack the error theorist's and my interpretation of ordinary moral thought, which may be called *absolutist*, and which assumes that ordinary folks are committed to the existence of sufficient normative reasons that need not be furthering to the agent's ends, by rejecting either P1 or P2. To reject P1 is to dispute that the underlying logical form of moral imperatives is that of a categorical imperative, and maintain that it is that of a hypothetical imperative (see 2.4.2). There are two forms that a sentence implicating an ought-statement may take:

[27] Though on my understanding, the claim that moral requirements *entail* reasons is weaker than the claim that moral requirements *supply* or *provide* reasons (see 1.3).

> Hypothetical Imperative: ‘If you want to Φ, then you ought to Φ’
>
> Categorical Imperative: ‘You ought to Φ, irrespective of whether Φ serves any of your ends or desires’

On this option, whether or not moral requirements invariably supply agents with reasons, they do so by applying to agents hypothetically or contingently on their ends. On Finlay's terminology, they may (or may not) be reason-supplying, but they are certainly *not categorical* or *absolute*. Anyone who disputes P1 must accept that the general moral judgement 'People ought to be moral' can sensibly be reconstructed in the following hypothetical form: 'People ought to be moral, *if* they want to'. As I noted in 2.3.1, such hypothetical moral judgements are far from a realistic possibility. When people make moral judgements, they do not condition them on the addressee's ends, nor it makes sense to reconstruct moral judgements in a hypothetical manner. Moral judgements seem to carry with them a certain unconditionality. At any rate, Finlay does not dispute the categorical form of moral judgements:

> Certainly moral judgements are characterized by categoricity in some sense; we do not explicitly condition our moral assertions on agents’ desires or concerns, for example (Finlay 2008: 349).

In accepting that moral judgements are categorical, what I take to be a truism about morality, Finlay avoids a serious charge of eccentricity. For anyone who denies that morality is ‘characterised by categoricity in some sense’ must be taken to reveal a highly unconventional understanding of morality. This brings us to the second option for anyone wishing to attack our interpretation of ordinary moral thought: to reject P2. This is to deny that categorical moral requirements are supposed to *invariably supply* agents with reasons. This is the route that Finlay takes.

Now, in denying P2, Finlay risks sounding eccentric nonetheless. For as we saw in the previous chapter, there seems to be a striking difference with regard to our attitude towards the breaking of moral norms, on the one hand, and the breaking of the norms of etiquette, on the other.[28] Whereas we occasionally condone rude behaviour, we hardly ever condone immoral behaviour. Morality, we may say at this preliminary stage, appears to possess a serious and special backing when it comes to practical decisions and the assessments of past actions. Unlike etiquette and any other set of rules, morality appears to possess a practical 'oomph' (cf. Joyce 2006). An implication of Finlay’s claim is that the normative weight of moral norms is on a par with that of, say, norms of etiquette, provided that on Finlay’s view neither framework invariably supplies agents with reasons. This leaves Finlay with the burden of accounting for intuitive observations such as the ones made in 2.3.2.

But Finlay is not an eccentric — for in his interpretation of ordinary moral thought — the one in which moral requirements are not supposed to invariably supply

[28] I take it that, in this respect, our attitude towards the breaking of the rules of any other non-moral framework should be similar. That is, I have trouble thinking of a non-moral framework which would elicit a response of disregard when it comes to violating its rules for the sake of a special cause.

agents with reasons — he does acknowledge the fact that there is a striking difference between the way ordinary people treat moral norms and the way they treat the norms of nearly any other normative framework (2008: 354). In other words, even though he accepts much of the 'evidence', by which we mean observations pertaining to common behaviour of those engaging in ordinary moral discourse (see 2.3), he nevertheless attempts to provide an alternative interpretation of it that is compatible with the rejection of 'absolute normative authority'. In doing so, Finlay can be taken to imply that the rival hypothesis (moral requirements are supposed to invariably supply agents with reasons) provides the basis for a good interpretation of ordinary moral discourse, though he chooses to argue 'defensively' that his own interpretation 'can explain the evidence at least as well' (2008: 352). And herein lies what is admirable about Finlay's interpretation of ordinary moral discourse: in his subtle way of explaining the 'evidence' which resists 'absolutist' commitments while simultaneously avoids sounding eccentric. But no matter how subtle a magician's trick is, we can always spot its inadequacies when the matter is examined in a careful and thorough manner. This is the task of the next section.

3.5 Methodology

So I (and the error theorist) think that moral requirements are supposed to invariably supply reasons, that morality possesses absolute normative authority, and Finlay thinks not. How do we go about resolving the disagreement? The first thing to note is that the disagreement is partly conceptual and partly empirical. It is partly conceptual insofar as Finlay and I understand our concept of a moral requirement in substantially different ways. On my understanding, it is the concept of a requirement that invariably supplies agents with reasons, and on Finlay's, it is the concept of a requirement that does not. The disagreement is also empirical insofar as we both ascribe our concept of a moral requirement to ordinary moral thought — assuming, that is, that the vast majority, if not all, of ordinary people engaging in moral discourse share our concept of a moral requirement. Importantly, what is of primary concern here is the empirical disagreement. For the purpose of this chapter is to examine the grounds for thinking that the thought that moral norms invariably supply agents with reasons is part of ordinary thought.

An obvious way to resolve the empirical disagreement is to *ask* ordinary practitioners in moral discourse *think* of normative authority of morality to be. The question here is: can we plausibly ascribe to ordinary moralisers the thought that moral norms, which apply irrespective of the agents' ends, invariably supply agents with reasons? Would the majority, if not all, of ordinary moralisers think that Adolf Hitler, regardless of his interests and concerns, had reason to abstain from causing the misfortune of so many people? A positive answer, namely that Hitler did have reason to abstain from causing the misfortune of so many people, would lend support to my interpretation of ordinary moral thought, whereas a negative answer, namely that, given his peculiar interests and concerns, Hitler did not have reason to abstain from causing the misfortune of so many people, would lend support to Finlay's.

3.5.1 Problems with Introspection

There are, unfortunately, difficulties associated with the process of reflection or *introspection*. Introspection is the kind of reflection that provides us with insight into

our thoughts, hence it can be of use in analysing our concepts. But it can also be defective. To see one such way that introspection can be defective, we may consider the phenomenon of implicit bias. According to a standard definition, an individual harbours an implicit bias against some stigmatised group G, when she has automatic cognitive or affective associations between her concept G and some negative property P, which are influencing her behaviour without the conscious awareness of the agent (Holroyd 2012: 275). 'Cognitive associations' here can be understood as inferential dispositions, and the alleged automatic character of them can be seen as contributing to the agent's lack of conscious awareness of them. To give an example of a case, suppose John is walking down the road when she notices Mary, a person belonging to a stigmatised ethnic group. Upon noticing Mary, John crosses the road pursuing a different path, thus indicating a nervousness of sorts relating to the group in question. When asked about his ethnic associations, John declares innocence; that he harbours no bias against that ethnic group. On the assumption that he says the truth, John may very well still harbour *implicit* (as opposed to explicit) bias against the relevant stigmatised group. In fact, we can account for his nervousness without having to suppose that John is also consciously aware of the relevant inferential dispositions which may be producing his anxiety. In other words, the example illustrates how a person may possess certain inferential dispositions (e.g. STIGMATISED GROUP G → DANGER) while at the same time not being able to identify those dispositions by means of introspection. The moral is that it may be possible that we are not aware of some of our inferential dispositions, though the latter are exhibited in our behaviour.

Second, even if ordinary practitioners in moral discourse are able to discern with a certain degree of clarity the relevant inferential dispositions of theirs, it may still be difficult for us to interpret their testimony in appropriate ways — given that they do not normally think or speak of their spontaneous intuitions in exactly the same way that philosophers do. For example, it may be possible that people who subscribe to the thought that moral requirements are supposed to invariably supply agents with reasons have their own ways of pledging their allegiance. People may use the language of 'oughts', as in 'Hitler really ought not to have done such-as-such', or the language of both 'oughts' and reasons in a misleading way, as in 'Hitler really ought not to have done such-and-such, even though he had good reason to do so'. Regarding the former, it is an open question whether we should translate statements about 'oughts' as implying that agents have sufficient normative reason to do such-and-such, and regarding the latter, it is unclear what sense of 'reason' is being called attention to, and moreover, what is the relation between what the agent really ought to do and those reasons.

Finlay himself recognises something like the above difficulties when he observes that 'introspection is known to be untrustworthy' (Finlay 2008: 353). In downplaying the importance of (what he calls) introspective 'evidence', Finlay is not alone either. Joyce too thinks that the best way to adjudicate between competing interpretations of ordinary moral discourse is not by relying on what people will actually *say* if asked, but rather on 'which conception of morality … best explains and makes sense of the dominant majority of moral practises' (Joyce 2001: 98). That is, considering common behaviour that may be attributed to the vast majority of moral practitioners, and attempt to infer the thoughts or concepts that best account for it. Compare: why did John drink petrol? What beliefs and intentions makes best sense to

attribute to him? My best guess would be: he thought that the glass contained water, and he intended to satisfy his thirst by drinking it.

3.5.2 The Evidence, and Inference to the Best Explanation

Drawing on a distinction made between ordinary moral behaviour and thought (2.3, 2.4), we may arrive at a plausible methodology for resolving the disagreement. We start by considering ordinary moral behaviour, our *evidence*: how we use moral concepts, distinctive verbal utterances of participants in moral discourse, as well as some of their characteristic ways of treating transgressors. Having listed such observations regarding ordinary moral behaviour, we will then ask: What can these tell us, if anything, about ordinary moral *thought*? That is, based on the evidence, we will attempt to infer the concepts accounting for it. The rationale is the following: to the extent that an interpretation (relativistic or absolutist) is better suited to explain the evidence, we seem to have reason to think that the concept which features in that interpretation really is part of ordinary moral thought.

In what follows, we consider each of the distinct features of ordinary moral behaviour 'that can be or have been advanced as evidence' (Finlay 2008: 352)[29].The evidence that is deemed *not* telling, will be dismissed. This is understood as particular aspects of moral behaviour that are equally compatible with both interpretations, and which hence cannot determine which one is right. By contrast, the evidence that is deemed telling (i.e. possibly revealing of the crucial commitments) will be subject to further argumentation. The latter will be comprised of arguments designed to support an inference to the best explanation.

3.6 Weak Evidence

3.6.1 Linguistic Evidence

Finlay begins by describing what he calls *linguistic* evidence which 'consists in the nonrelativized form of the sentences that people utter when they make moral claims'. He then narrows it down to cases in which 'the relevant end or standard is shared and [is] readily identifiable' (2008: 353). An example of such evidence is the following statement: 'You ought to keep your promise'. Note how such declarations are hardly ever relativized — or to put it otherwise, cannot be interpreted in the form of a hypothetical imperative. We hardly ever hear anything like: 'You ought to keep your promise , *if* you want to' — nor do we ever think along the lines of: 'Your responsibility to keep the promise is conditioned on your wanting to do so'.

But as Finlay rightly notes, the non-relativisation of linguistic assertions is by no means a unique feature of moral utterances. He gives the example of the captain of a rugby team where it would be strange for him to prefix his instructions to his teammates with expressions like '… *if* you want us to win the game'. As he puts it, 'the obvious need not be stated' (2008: 353). In general, we often phrase our linguistic assertions about other (non-moral) frameworks in nonrelativized form. This mainly

[29] Note that Finlay talks of seven such features, with the first being reflective evidence. This has already been dismissed - thus I will consider the remaining six.

occurs, it seems, when there is an implicit agreement among us and our interlocutors that we all share the same concerns. For example, we say 'Put your feet off the table because it's rude', implying something like '… we both know that you don't want to be rude'.

The above observation is of some importance as it shows that, just as we cannot infer that the normative authority of etiquette (or rugby) is absolute based on the nonrelativized form of some linguistic assertions, so we cannot infer that the normative authority of morality is absolute based on assertions such as the above. But there is an equally important thing to be said in response to the current observation.

As I have already noted, the current evidence is rather specific in that it relates strictly to those cases in which discourse occurs among 'subscribers': people who share the same ends or concerns. Restricted in this way, the current linguistic evidence is unsurprisingly insufficient to tell us anything interesting regarding what ordinary people think of the requirements of morality, on the one hand, and those of etiquette (and other non-moral frameworks), on the other. Specifically, it is unable to give us any insight on whether ordinary people take the requirements of morality, and (say) not those of etiquette, to possess absolute normative authority. To see this, note that even on the assumption that moral requirements possess absolute normative authority, we would still not expect ordinary people to invoke this normative consideration when it is not called for — for that would be weird or even out of order. Imagine how strange it would be if we kept reminding our morally conscientious friends that they must act morally because moral requirements invariably supply one with reasons. This could — unsurprisingly — provoke them, as it would seem as if we were questioning their commitment to the moral way of life. Rather, it is when dealing with non-subscribers (or moral indifferentists), people who do not share our concern for morality, that appealing to the supposed absolute normative authority of morality seems in order. Compare: a policeman need not remind us, law-abiding citizens, that he has a gun which she could use to restrain our behaviour if needed — but when occasion calls for it she may very well do so.

Thus in order to know what they take the normative authority of morality or etiquette to be, we need to press ordinary people — so to speak — and inquire about their response or reactions in regard to those who do not share their concern for morality or etiquette: 'What would you say to someone who behaves immorally and does not share the concern of decent people for humanity?', or in the case of etiquette, 'What would you say to someone who behaves impolitely and does not share our concern for good manners?' To the extent that their response remains the same (or alters), we may be able to draw interesting conclusions about what ordinary people think of the normative authority of each framework. For linguistic evidence that is telling in the aforementioned way, we must wait for subsequent sections.

3.6.2 Appraisal Evidence

Finlay calls *appraisal evidence* that which relates to 'observation … of what information we treat as relevant and what we treat as irrelevant when making a moral appraisal' (cf. Joyce, 2001: 99). Like before, Finlay narrows down *appraisal evidence* to the following kind of appraisal: that which relates to the judging of a person's action or character as morally bad or wrong (2008: 353-354). Finlay rightly notes that:

> We might be convinced that the Nazis, the emperor Caligula, Charles Manson, or Fred West lacked any concern for or commitment to moral ends or standards, but this would not deter us from judging their actions wrong and their characters bad (2008: 354).

But he quite sensibly remains unimpressed by the observation:

> Is this anything that a [relativistic] interpretation cannot explain? (ibid)

I agree with Finlay that this is something that his interpretation can explain.. Our moral appraisal (construed narrowly, such as to include only appraisals pertaining to the judging of a person's action or character as morally bad or wrong) should be expected to take exactly the same form irrespective of whether the assumption of absolute normative authority was part of ordinary moral thought. For example, we would still think that, despite his indifference to moral standards, Charles Manson's actions were morally wrong even if we simultaneously thought that the normative authority of morality was not absolute. After all, this is certainly true of other kinds of appraisals. To see this, suppose that you are in the bus on a busy day, and in order to kill time you invent a silly game according to which whenever someone touches their nose they win five points. There are at least two kinds of appraisals you could be making about your game: whether actions qualify as right or wrong (or successful or unsuccessful) according to your game, and whether strangers have reason to touch their nose. You may say that the person sitting next to you is doing reasonably well — as she is winning quite a few points, but at the same time, it would be insane for you to think that she necessarily has reason to touch her noise more frequently. After all, what is your game to her? This shows that your appraisals pertaining to whether behaviour merely qualifies as right or wrong according to your game, says nothing about your commitments regarding the normative authority of your game's demands. In fact, it seems to be perfectly consistent to hold the following judgements at the same time: (i) one ought (according to the rules of my game) to touch her noise, and (ii) one does not have any reason to do so. If something like this is true in the case of silly games, why should it not be true with respect to morality? The upshot here is that appraisal evidence thus construed seems to be consistent with both hypotheses that are on the running: for the simple reason that considerations pertaining to (ii) are not supposed to be revealed in the practises that reflect such evidence.

Before we discard this evidence, it would be interesting to consider a further feature of *moral* appraisals thus construed. The suggestion is hinted at by Finlay when he says:

> We are much more serious and intransigent about our moral appraisals than we are about our appraisals of manners or chess strategy (2008: 354).

Indeed, there is a striking difference between the way we talk about our moral appraisals and the way we talk about appraisals of other kinds. We say 'That was morally wrong!' in a much more serious and passionate way than we let others know that (say) their chess move was illegitimate. Could this observation be indicative of

what we think about the normative authority of morality, on the one hand, and of chess, on the other? I submit that it could not, and therefore I tend to agree with Finlay:

> Moral standards or ends are of pressing concern to ordinary, decent people, and their importance to us typically overrides the importance of other standards and ends (ibid).

We tend to feel more strongly about moral appraisals than we do about (say) chess appraisals because we happen to care about morality more than chess. I am not suggesting that this explanation is all there is to the striking difference — just that it is sufficient to explain it. The upshot is that even this further feature of moral appraisals — our unsurpassed emotional investment in them — cannot give us any insight on whether we take the normative authority of morality to be absolute or not. The point can be put in terms of a distinction between *psychological* and *normative* importance (cf. Williams 1985: 203-204). It makes equal sense to say that one cares for morality because she acknowledges that morality is important *for her* — as opposed to because she acknowledges that morality is normatively important — *irrespective of what she or anyone else happens to think or feel* about morality. The former sense of importance may be regarded as psychological, whereas the latter as normative .

3.6.3 Reactive Evidence

Finlay also considers what he calls *reactive evidence*, which consists in:

> … our reactive attitudes of blame towards those we believe to have acted wrongly, and our judgements of blameworthiness, responsibility, and the deservedness of blame and punishment. The fact that our reactive attitudes towards and judgements about moral transgressors are not contingent upon what they desire or the standards to which they subscribe… (2008: 358).

The question hinted at by Finlay is:

> Is the practice of holding people blameworthy irrespective of their interests or concerns revealing of an assumption (on the accuser's part) that they violated reasons that had authority for them?

I do not believe that it is. I begin by noting that the evaluative dichotomy between blaming and praising others — irrespective of their interests and concerns — is not exclusive to *moral* thought. After all, our actions can be approached from different perspectives, and morality comprises only one such perspective. Upon finding out that my ice cream is missing, I may ask who took it: who is responsible for its missing? If it turns out that John, who promised me not to touch it, ate it, I may infer that he is *morally* blameworthy for eating it — even though in doing so he was not violating any of his interests or concerns. If, moreover, we lived under a regime that had it *illegal* for anyone to eat ice cream in the evening, I may as well infer that John is *legally* blameworthy for eating it — even though in doing so he was not violating any of his interests or concerns. And if, to stress it even further, we were playing a game according to which whoever eats the other person's ice cream wins ten

points, then John's action would also qualify as *praiseworthy* in relation to that game — irrespective of whether in following the rules of the game John was satisfying or sabotaging any of his interests or concerns.

Now, if it is part of ordinary thought that moral culprits necessarily violate reasons that they have, it must also be part of ordinary thought that any ascription of moral blameworthiness would *ex hypothesi* entail the judgement that the culprit violated reasons that she had. But the bare fact that we morally blame culprits cannot by itself vindicate either hypothesis pertaining to what ordinary people take the normative authority of morality to be. After all, we also blame or praise people based on other non-moral considerations (e.g. legal) and this cannot be taken to be indicative of any commitment on our part regarding the normative authority of such other frameworks.

To sum up the weak evidence: Whereas *linguistic evidence* is construed very narrowly such as to include discourse occurring only among subscribers, *appraisal evidence* is also construed narrowly enough as to exclude appraisals pertaining to what we think others have reason to do. Moreover, we can explain certain striking differences between our moral and non-moral appraisals without having to postulate an assumption of absolute normative authority. Reactive evidence was also found to be non-telling, as it reflects practises that are not exclusive to the ascription of *moral* responsibility.

3.7 Strong Evidence

Up to this point, Finlay's defensive strategy has worked just fine, as he has managed to show that the above kinds of evidence do not lend support to my interpretation of ordinary thought. Finlay now moves on to consider three closely related kinds of evidence, and precisely because they are in this way related, I am inclined to treat them as 'facets of the same argument'.[30] Unlike the first three kinds of evidence, these three are clearly about practises in which ordinary people are concerned with matters of normative authority — e.g. questions about what others *really should* do, as opposed to questions on whether certain acts merely qualify as right or wrong from the perspective of morality. Furthermore, and importantly, these also include discourse occurring among subscribers and non-subscribers.

In short, strong evidence consists in the following related observations: ordinary people do not retract nonrelativized moral judgements upon acknowledging that the addressee has no concern for following the injunctions of morality (address evidence), ordinary people expect that such people nonetheless conform with moral requirements (expectation evidence), and ordinary people still think that transgressors are in a deep sense mistaken in not following the injunctions of morality — even upon acknowledging that they are totally indifferent to it. As we see, these three related kinds of evidence lend *prima facie* support to the interpretation of ordinary thought according to which moral requirements are supposed to invariably supply agents with sufficient normative reasons.

[30] Though Finlay complains that they are not (2008: 352).

The structure of this section will be as follows: I first present the evidence as Finlay understands it. Following Finlay's initial concession that these practises may be incompatible with his interpretation, I will proceed to considering four arguments that I discern and which are offered by Finlay in support of his interpretation. I then explain why each is unconvincing.

3.7.1 Address Evidence

Finlay understands *address evidence* as relating to the practise of addressing moral claims to an audience who do not share our moral ends or standards (2008: 355). We may regard the following example of ordinary conversation as a fine representative of such evidence:

> -Don't lie to her, tell her the truth.
>
> -Why?
>
> -Because it is morally wrong to do so!
>
> -But I really like lying — and I share no concern for morality.
>
> -Both these facts are irrelevant to what you really should do on this occasion.

We must note that, unlike previous kinds of evidence, the current relates (first) to judgements about how others should or should not act, and (second) to individuals whom we know to be indifferent to moral ends or concerns. Both these features are crucial for adjudicating between Finlay's and my interpretation of ordinary thought. For, as suggested earlier, it is in relation to these two considerations that the absolute normative authority would naturally come into play. More specifically, it is precisely in regard to judgements about how *indifferentists should* or *should not* act that the notion of the absolute normative authority of morality would find natural application. To give an example, upon acknowledging that Charles Manson had no concern whatever for acting morally, the idea that moral requirements supply reasons is precisely what would dispose us to think that he nevertheless should not have killed so many people. At risk of repetition, it is the thought of absolute normative authority that would naturally tip the balance from our thinking that morally alien people should behave immorally to the judgement that they should not.

Indeed, I submit, the reason why we always address (and hardly ever retract) our normative judgements to moral indifferentists is because we think that the normative authority of morality is absolute. By contrast, the reason we sometimes fail to address (or choose to retract) our normative judgements about etiquette to others is because we do not think that the normative authority of etiquette is absolute. Consider our familiarity with something like the following conversations:

> -Don't put your feet on the table, behave!
>
> -Why?
>
> -Because it is rude!
>
> -But the man sitting at the corner is paying me a large sum of money to do so.

-Well, in that case, you really should put your feet on the table — even though by conventional standards it is still rude.[31]

-Don't lie to her, tell her the truth.
-Why?
-Because it is morally wrong to do so!
-But the man sitting at the corner is paying me a large sum of money to do so.
- You should still not lie on this occasion.

By contrast, consider our lack of familiarity with both the following ones:

-Don't lie to her, tell her the truth.
-Why?
-Because it is morally wrong to do so!
-But the man sitting at the corner is paying me a large sum of money to do so.
-Well, in that case, you really should lie to her — even though by the standards of morality you should not.

-Don't put your feet on the table, behave!
-Why?
-Because it is rude!
-But the man sitting at the corner is paying me a large sum of money to do so.
-You should still not put your feet on the table on this occasion.

The acknowledgement of peculiar ends or concerns of others may occasionally prompt us to retract our judgements on how they should act according to non-moral frameworks such as etiquette. This is not true, however, with respect to morality — and the natural explanation for this is that we think that moral requirements invariably supply agents with reasons.

Finlay concedes — at least initially — that:

> [Address evidence] may seem problematic for a relational interpretation: when the audience does not share the speaker's concerns, those concerns might not be in the context. The nonrelativized character of moral utterances in such cases may therefore be taken as evidence that the moral judgements being expressed are not intended as relational [i.e. relativistic] (2008: 355).

In 3.7 and 3.8, we consider Finlay's attempt to account for address evidence.

3.7.2 Expectation Evidence

[31] Or as Joyce would very eloquently put it: despite *a remaining sense* in which you should not do so (2000: 465).

Moreover, we may consider the *point* of engaging in the practise of addressing nonrelativised moral judgements to nonsubscribers (2008: 355). As Finlay puts it:

> [Ordinary people] address their moral claims to nonsubscribers, apparently in the expectation that in doing so they might possibly succeed in influencing the other by the provision of absolutely authoritative reasons, the authority of which the other can recognize and obey (ibid).

In other words, it seems true that, regardless of their interests and concerns, we nevertheless *expect* moral indifferentists to acknowledge that moral considerations deserve to be taken seriously in their practical deliberation. As Joyce puts it:

> Calling an action 'morally correct' or 'virtuous' or 'wrong' or 'just' is (putatively) to draw attention to a deliberative consideration that cannot be legitimately ignored or evaded (Joyce 2006: 57).

It would be the cause of both bafflement and irritation were others to treat our judgements about what they morally should to do with characteristic disregard.

But on the assumption that it is part of ordinary thought that the normative authority of morality is absolute, we can explain in a very natural way why we expect indifferentists to be influenced by moral norms in a way and to a degree that we do not expect them to be influenced by (say) the norms of etiquette: moral norms, unlike those of etiquette, are thought to supply agents with reasons. If Finlay was right in thinking that the assumption of absolute normative authority is not part of ordinary moral thought, I submit, we would not find it permissible to insist that others treat moral norms as directly relevant to what they really should or should not do. Moreover, if Finlay was right, we would also expect others to dismiss our characteristic insistence as eccentric. That is, the addressee who does not share our moral ends or concerns would be likely to reply: 'What are you even talking about? Why would you think that I *really* should not lie given that you know that I have no interest or concern whatever in doing so?'

It must be emphasised, of course, that my interpretation of ordinary moral thought does not require that people *actually* conform with moral norms. Rather, this would be highly consistent with my interpretation provided at least that moral indifferentists do not treat our insistence that they act morally as eccentric or out of order. In short, that they are not likely to respond to it with 'What are you even talking about?' If ordinary moralisers think of moral requirements as reason-supplying, which I submit they do, they must be at least inclined to feel 'bound by' their normative authority.[32]

Again as before, Finlay concedes that expectation evidence may be lending prima facie support to my interpretation (2008: 355). In 3.7 and 3.8, we consider Finlay's attempt to account for expectation evidence.

[32] See 2.4.3 where the notion of moral bondage is discussed.

3.7.3 Disputation Evidence

Finally, we may consider *disputation evidence*. This is 'the practise of fundamental moral dispute — i.e. that which occurs between people who subscribe to fundamentally different moralities' (ibid). More specifically, this relates to disputes over what we or others *really should* do. This provision is important — and Finlay seems to be neglecting it. This is the relevant passage:

> If people with incommensurable moral ends or standards engage in a dispute over *what is 'morally right'*, this is evidence that they assume that moral rightness is an absolute, not a [relativistic] matter (ibid, emphasis added).

Finlay summarises disputation evidence as disagreement over 'what is morally right' — but I find the italicised remark somewhat misleading. Remember, the issue here is whether it is part of ordinary moral thought that the *normative authority* of morality is absolute or relative. Given that this is our current focus and concern, it may be misleading to treat disagreement over which actions *merely* qualify as morally right or wrong as telling evidence. After all, it would be compatible with Finlay's interpretation of ordinary thought that people disagreed over which actions qualify as morally right and wrong without thinking at the same time that their conclusions would necessarily be indicative of what they have reason to do. The two issues seem logically independent (cf. Shafer-Landau 2003: 165). Compare: people may disagree over whether typing a message on your phone while sitting at the dinner table is rude — without thinking at the same time that the normative authority of etiquette is absolute. After all, even if using your phone at the dinner table really is rude, it would still be the case that one may occasionally have a reason to break this rule (e.g. to let your boss know that you are going to be late for work).

It is, then, disagreement relating strictly to judgements on how ourselves and others *should* or *should not* act that must concern us. On the assumption that Finlay's interpretation is right:

> Were people to recognize that they subscribe to fundamentally different moral standards, then they should recognize that there is no genuine dispute between them (ibid)

That is, were people to recognise that they subscribe to fundamentally different moral standards, then they should be also expected to recognise that there is no genuine dispute over what they should or should not do — or over what reason they have. For if the normative authority of morality is ordinarily not taken to be absolute, we should expect people to treat the two discussions; which actions qualify as morally right or wrong, on the one hand, and what they have reason to do, on the other, as somewhat unrelated. That is, people may be disagreeing over who is morally right, without necessarily disagreeing over what they have reason to do. The latter is the sense in which there would be no 'genuine dispute'. Of course, as Joyce accurately notes, '[this] hardly describes moral disputation' (2001: 99). Ordinary people take the question of what actions are morally right and wrong to be highly informative to the question of what they and others have reason to do.

To give a famous example, consider the ongoing controversy surrounding the *moral* status of abortion. The 'pro-life' movement holds that abortion is morally wrong, while the 'pro-choice' disagrees. In disagreeing over this, the two sides seem to be presupposing that the dispute runs deep: the disagreement is not merely about which actions qualify as right or wrong according to the 'moral game', rather it is disagreement over how we should conduct our lives. That is, it is (by extension) disagreement over what we have adequate reason to do. If Finlay was right in his interpretation of ordinary thought, we would expect the pro-lifer to be disposed to concede something like the following to her opponent: 'Even though by the standards of morality is wrong to kill a foetus, this has nevertheless no immediate implication over what you *really should* do'.

Again as before, Finlay concedes that the existence of disputation evidence may be lending prima facie support to my interpretation:

> If people with incommensurable moral ends or standards engage in a dispute over what is 'morally right', this is evidence that they assume that moral rightness is an absolute, not a relational matter (ibid).

To sum up the strong evidence: The existence of address, expectation and disputation evidence lend *prima facie* support to my interpretation of ordinary moral thought. The availability of such evidence must be considered undeniable given our strong familiarity with the practises that it reflects. These three related practises seem to be part, if not *the* point of engaging in moral discourse. We often struggle to reach important moral conclusions (e.g. is abortion morally justified — and if so, under which conditions?). Crucially, we do not regard our conclusions as merely intellectual property of ours. On the contrary, we think that they have direct application to the way we and others *really should* act — and we do not cease to think this upon acknowledging people's indifference to morality. Importantly, what people *really should do* and what people *have sufficient normative reason to do* seem intertwined. Or to put it otherwise: what ordinary people could mean by *really should do* if not *have adequate reason to do?* The pressure is on the relativist to account for such instances of behaviour in a way compatible with her interpretation.

To his credit, Finlay does not dispute the existence of the strong evidence, but he rather tries to explain it. This takes up to the next task, which is to examine whether Finlay's interpretation can explain the evidence at least as well.

3.8 Finlay's First Move: Disputing the Pervasiveness of the Evidence

Though he initially concedes that the existence of the above practises constitutes evidence against his interpretation, Finlay does not go down without a fight. After considering the strong evidence, Finlay asks:

> Are these three features essential to moral practice as we know it, and are they genuinely incompatible with a relational interpretation of morality? (2008: 355)

Finlay's attempt to account for the evidence in a relativistic manner comes in two parts. First, he disputes the existence of what he calls 'genuine' address, expectation and disputation evidence. In effect, this is the claim that the practises reflected in the evidence are neither pervasive nor essential to moral discourse. The practises are neither, the suggestion seems to be, reflected in the behaviour by the vast majority of moral practitioners, nor are they important or anything like the *point* of moral practise. Second, and this time on the assumption that such practises *are* both common and essential, Finlay counteracts the claim that they are really incompatible with his own interpretation. In effect, this strategy takes the form of an inference to the best explanation (or interpretation). More specifically, this involves the use of two arguments: one based on the distinction between the semantic and the pragmatic point, and one based on the notion of *psychological importance*.

3.8.1 Are Such Practises Characteristic of Moral Discourse?

Finlay's first move is to question the pervasiveness of the relevant practises. Finlay wonders:

> Are these three features essential to moral practice as we know it…?

Before he adds:

> … they are nowhere near as common or characteristic of moral practice as one might think (355).

In effect, Finlay makes two claims. First, that as a matter of fact, ordinary people rarely, if ever, have moral discussions with individuals whom they know to be totally indifferent to their moral concerns. As he vividly puts it, 'try to recall the last time you engaged in moral discourse with someone like Charles Manson or a neo-Nazi' (2008: 356). Second, that even if such practises scarcely occur, they are nevertheless *not essential* to ordinary moral discourse. I may accept that the relevant practises (reflected in address, disputation, expectation evidence) are quite rare, but I shall press that, first, they are *not* nonexistence, and second, that they are of central importance to moral discourse.

Relating to the first claim, the relevant practises are indeed rare in virtue of the fact that the overwhelming majority of people are decent and happen to share the same ends. As a matter of fact, moral indifferentists form only a small minority. But this does not mean, of course, that we never engage in discourse with people who are totally indifferent to our moral ends. More specifically, the scarcity of the practises reflected in address, expectation and disputation evidence does not rule out the possibility that we sometimes *do* address our nonrelativised moral judgements to nonsubscribers, that we expect them to change their ways, and that we enter fundamental dispute when they disagree with our strong moral convictions pertaining to what they should do.

To give a notable example, some of the finest moral representatives of their generation addressed their moral convictions to actual nazis, in the famous trial of

Nuremberg. Let us, first, grant that the trial of Nuremberg can be seen (for the most part) as a moral dispute between people who did not share the same moral concerns, *and who were also aware of this fact*. The latter addresses Finlay's concern, namely that 'to qualify as [address, expectation, disputation] evidence, it has to be the case that the speakers *accept* that they are involved in discourse with such an interlocutor' (2008: 356). The trial of Nuremberg, therefore, constitutes a prime example of address evidence — and Finlay needs a story to account for the genuine moral disagreement between the judges and the war criminals.

It would do no harm to the spirit of this discussion to dwell a little on the specifics of the above example. For the purposes of a thought experiment, suppose that history took a twist, and upon hearing the nazi officer's clarification that he has no concern whatever for the welfare of non-caucasians, the judge explicitly *retracted* the judgement that the nazi officer *really ought to have acted differently*. Moreover, suppose that an ordinary person who can consider herself a competent moraliser comes to find out about the alleged historical incident (e.g. by reading about it on wikipedia). The interesting question here is: Is it likely that the person would embrace the moral stance of the Nuremberg judge, specifically the judgement that the nazi officer *really ought not to have acted differently*, or is it likely that an ordinary moraliser would choose to separate her moral stance from that of the Nuremberg judge, specifically in relation to the aforementioned judgement?

The answer, I think, is obvious: any competent participant in moral discourse would find the judge's retraction of the initial judgement *scandalous*. That is, if upon acknowledging the nazi officer's complete indifference to morality, the judge was inclined to concede that he did not have reason to abstain from the enormous crimes, the judge would 'be saying too little' (Joyce 2001: 527). Anything like the following verdict would be deeply uncharacteristic of ordinary moral thought: 'Mr. Göring, even though we acknowledge that in acting in a genuinely monstrous fashion you were simply doing what you had adequate reason to do, we nevertheless choose to sentence you to death, as your actions, even though in complete accordance with what you had adequate reason to do, came in conflict with *our* own moral concerns'.

Relating to the second claim, it seems that such practises are also in an importance sense *essential* to ordinary moral discourse. For one thing, the mere fact that something happens rarely in one's lifetime does not entail that it is not important or essential to those involved. To give an example, one does not give birth everyday, but it is very unlikely that one would consider the act of giving birth as in any sense insignificant to one's life. Something analogous goes in our case. That is, though instances such as the trial of Nuremberg — in which decent people stood firm in their conviction that those who do not share their ends nevertheless had reason to do otherwise — are rare, they are nevertheless of great importance to those participating in ordinary moral discourse. More specifically, the practises reflected in address, expectation and disputation evidence are important in the sense that they are part of what *defines* our moral stance. Not only do we separate our stance from the alternative verdict sketched above, but we are eager to point out that we disagree with it — as if endorsing it would make us less moral than we take ourselves to be. Moreover, that such practises are of central importance to us is also evinced by the importance that we attach to the trial of Nuremberg in which people who subscribed to fundamentally different ends engaged in moral dispute. In generally, we tend to regard a historical

opposition to morally depraved individuals or establishments as something significant and deeply telling of our general moral stance.

3.8.2 Such Practises Are Odd

Finlay also utilises the observation that the related practises seem fruitless, or even odd. This line is reminiscent of an argument given by Harman in a similar discussion. Here, I conjoin the two lines in an attempt to bolster Finlay's case. Finlay observes:

> Were a person to come to accept fully that another does not share her fundamental moral concerns, would she bother to address her moral claims to him? (2008: 358)

Finlay then — correctly — notes that discourse with morally alien people is 'a fruitless waste of time and effort', and feels confident in declaring that:

> We only enter into moral disputes with people, generally speaking, if we think that we can find some moral premises that they may accept (ibid).

The idea here seems to be that discourse with those who do not share our moral concerns seems fruitless *precisely because* we consider the normative authority of moral norms to be contingent on their concerns: that moral norms entail reasons only when others have a relevant end served by following them. The claim here may be accompanied by a similar observation made by Gilbert Harman. Harman observes that 'it sounds odd to say that Hitler should not have ordered the extermination of the Jews' (1975: 7). That is, *incongruous* with some aspect of the situation. Could it be because we do not think that Hitler-like individuals really should have acted otherwise — in virtue, that is, of thinking that moral considerations fail to supply them with reasons? We may sum up both observations: It seems both fruitless and sounds odd to tell to morally atrocious individuals that they really should not have acted immorally.

Granted, in the face of the atrocity of such individuals' character, it both sounds odd and seems fruitless to engage in moral discourse with them. But this is by no means an indication that our conception of the normative authority of morality is relativistic. Moral discourse with gangsters seems fruitless simply because we realise that the chance of influencing their behaviour is minimal, if any at all. As Burgess puts it, it is 'somehow out of place in view of the inconceivability of either of these gangsters being influenced by the kind of moral considerations that influence, say, readers of the Philosophical Review' (2007: 430). For the same reason we cannot teach a cat the Pythagorean theorem, we also cannot reason morally with unrepentant criminals: it would be a complete waste of time.

Additionally, such discourse may also sound odd, for a similar reason. The oddity is not due to any 'misuse of our own language', or a violation of our own concepts — as Harman would have us believe (1975: 5). That is, it is not because we think that morality lacks authority for such people, rather it is due to our acknowledging of the fact that we are offering advice which is very likely to be dismissed, or even laughed at. It is the chilling thought of Satan laughing at our moral advice, with an unrepentant smile and characteristic coldness, which may make such

talk seem unfitting under the circumstances. Telling Hitler that he ought to have done otherwise would have indeed been *semantically* appropriate, though it would have probably been *pragmatically* fruitless. That is, given that part of the pragmatic point of our utterance is to make Hitler change his ways, our utterance seems doomed to fail. But at the same time, the semantic point of our utterance, part of which is to communicate the fact that Hitler has adequate reason to change his ways, is *right in place*.

3.9 Finlay's Second Move: Inference to the Best Explanation

Finlay's next move is to suppose that even though the practises reflected in address, expectation and disputation evidence really do occur, they can be explained in a way that is compatible with his interpretation. That is, even though people do sometimes engage in moral discourse with moral indifferentists, this does not necessarily mean that they take the normative authority of morality to be absolute. The particular interpretation that Finlay offers makes use of a distinction between the *semantic* and the *pragmatic* point of a practise.

3.9.1 The distinction Between the Pragmatic and the Semantic Point of Moral Practise

Why do we sometimes behave as if we take moral requirements to be reason-supplying, despite thinking — as Finlay would have it — that they are not? Following Finlay:

> Granting, that is, that sometimes the point for which we engage in moral discourse is 'a desire to say something more—to imbue the moral imperative with a greater authoritative force' [Joyce 2001: 41], we can challenge the error theorist's assumption that this is part of its semantic point. It may instead simply be its *pragmatic* point, or what we aim to accomplish by our speech acts (ibid).

In understanding Finlay's point, I find useful recalling a point made earlier by him, namely that we sometimes leave our normative claims nonrelativised relying on context to contribute to their meaning. The idea is that instead of saying 'Take the penalty and score for your team, *if* you haven't lost all interest in football yet', we merely say the former clause while implying something like the latter. The understanding of course is 'there' as both us and our interlocutors are in implicit agreement regarding our common goals in the situation. Finlay then rightly stipulates that 'in fundamental moral discourse … this condition is not met' (ibid). That is, to go back to the example of the Nuremberg trial, the judges as well as the culprits are aware that they do *not* share the same concern for humanity. Based on this common understanding, the judge simply cannot say 'You ought not to have done it' when this implies '… given our common concerns'. On Finlay's view, however, the judge may nevertheless speak *as if* such agreement exists in order to:

> … pragmatically communicate certain things; in this case, it is a rhetorical way of expressing the *expectation* (demand) that the audience subscribes to the speaker's ends or standards (ibid).

That is, Finlay story goes, we often talk as if the normative authority of morality is absolute — treating moral norms as reason-supplying even though we are well aware that they are not — nor because we think that they are, neither because we want to communicate the thought that they are. In a nutshell, that is not the *semantic* point of our speech acts. Rather, we speak as if moral norms are reason-supplying in order to express an expectation or demand that others share or subscribe to our moral ends (ibid). To give a similar example, suppose I tell my friend who is sitting near the door: 'The cold wind is killing me'. In saying this, I do not mean that the cold wind is killing me — for that would be false. What I am really saying, or rather what I really want to accomplish by my speech act, is to express an expectation or demand that my friend close the door.

On Finlay's interpretation of moral discourse, when we address our moral claims to moral indifferentists, we do not mean that morality has authority for those who do not care for the ends we happen to care about, but we do so in order to merely express the expectation or demand that they begin to care about them more than they already do. Following Finlay:

> This story explains address, expectation, and disagreement evidence in a way that is not only compatible with, but also presupposes a relational interpretation of moral concepts (Ibid).

We address our nonrelativised judgements to moral indifferentists, not because we think or mean to say that that morality has authority for those who do not care for the ends we happen to care about. Rather, the point of our addressing such nonrelativised judgements to others is in order to express the relevant expectation or demand. Moreover, the observation that such rhetorical demands are of 'potential psychological efficacy' 'accounts for any expectation that we may have an influence on the other' (ibid). And when others resist subscribing to our moral ends, they may counteract our rhetorical trick by insisting that *we* subscribe to *their* (supposedly morally atrocious) ends. That is, on Finlay's story, when the neo-nazi insists that we do as he says, what he is effectively doing is expressing a demand that we subscribe to the nazi-morality. This suggests a way of accounting for disputation evidence, whereby the disagreement consists in contradictory expectations or demands: the demand of the decent person that the nazi subscribes to the end of human welfare, versus the demand of the nazi that the decent person subscribes to the end of promoting the interests of the aryan race.

> When the address is reciprocal, leading to fundamental moral dispute, we have the 'disagreement in attitude' that is the centrepiece of expressivist explanations of moral discourse (ibid).

Finlay's story can be rejected on at least three counts. First, it seems to presuppose a notion of psychological importance (as I shall call it) which is crucially

not sufficient to underwrite the expectation or demand that others subscribe to our moral ends. Second, it seems to neglect the fact that the acknowledgement of the relevant persuasive ploys is highly uncharacteristic of ordinary moral behaviour. Third, it fails to account for the shock that the alternative verdict of the previous section would provoke. Below, I elaborate on each point.

Psychological Importance

As we saw earlier, Finlay can invoke the notion of psychological importance in his attempt to explain the observations he sums up under *appraisal evidence*. Psychological importance, I suggested, may well suffice to explain that kind of evidence. At this point we may ask: Can psychological importance also explain the strong evidence? I believe the answer is clearly *no*.

Let me first briefly elucidate the notion as I understand it: One would consider a normative framework to be *psychologically* important for an agent just in case she considers its importance to consist solely in the fact that it happens to be important *to her*.[33] Finlay may be thus tempted to claim that we are much more serious and passionate about our moral judgements than we are about judgements of other kinds, and hence insist that indifferentists *really should* behave as we would have them behave, *solely in virtue* of our overwhelming appreciation for the former and our comparative disregard for the latter. That is, precisely because morality happens to be of increasing importance *to us;* not necessarily because of any other consideration which we associate with morality, such as the idea that moral norms (unlike other kinds of norms) are *normatively* the most important.

But to insist that some isolated group of people behave morally *simply because* moral ends happen to be important *to us* would be highly incongruous. Granted, moral ends are 'of pressing concern' *to us* and their importance *to us* 'typically overrides the importance of other standards and ends'.[34] From this, it does not seem to follow that *others* too *ought* to care about the moral ends that we subscribe to. The inference is extremely reasonable.[35] It is one thing for me to (a) feel strongly about some norms, another thing for me to (b) also want others to care about those norms, and quite another for me to (c) *insist* that others do so. Whereas the inference from (a) to (b) seems reasonable, the inference from either (a) or (b) to (c) seems unreasonable and far-fetched. Similarly, a person who is disposed to want others to care about her standards on the basis of her affection for them seems reasonable, whereas a person

[33] In saying this, I am not contradicting Smith's and Pettit's reasonable proposal that people often act on desires without recognising that they have such desires (1990: 565). That is, I am willing to accept that things such as desires, personal values etc. do not always feature in the 'foreground of [our] deliberation' (ibid: 580). Be it in the background or foreground, however, my claim is that with respect to the moral evaluations in question it cannot be the case that people are moved solely by desires (personal valuing and the like). For in our most reflective moments, we - reasonable people - cannot insist that others act as we do when our justification for our preferred acts consists *solely* in the fact that those acts are important *to us*. There has to be something more in the way of a consideration.

[34] (Finlay 2008: 354)

[35] Note that I here do not mean to contrast 'reasonable' with 'formally invalid'. Nor do I think, however, that an inference needs to be formally valid in order to qualify as sensible or reasonable.

who insists that others care about her standards on the basis of *her* affection for them seems aggressive or dogmatic. Compare: Though I occasionally enjoy watching volleyball, my liking for football is unsurpassed. Given my unsurpassed liking for football, I want others to follow football more than they do volleyball so that (say) football becomes the dominant sport and thereby each aspect of it improves as time progresses. At the same time, I do not *insist* that others too like football as much as I do. I would certainly encourage them to give it a chance, provided they have not done already, so that they might come to like it as much as I do. But if they end up not being as passionate about it as I am, then I would be highly unreasonable were I to insist that they do so.

To elaborate: if you do not think that there is something special about your ends, i.e. *beyond* the mere fact that *you* and your like-minded friends subscribe to them, why would you *demand* that *I* too subscribe to those ends? In the absence of a *normatively* special consideration for holding your ends to such high esteem, your *demanding* that I subscribe to them may very well be interpreted as overly aggressive and highly unreasonable. On Finlay's view, we are disposed to demand that others subscribe to the ends that we happen to have, without necessarily supposing that there is something *normatively* special about those ends. That is, without necessarily thinking that our ends are also the *right* ends. If such disposition was ubiquitous, ordinary people would be highly unreasonable.[36] What is more, it would be supposing that ordinary people are also likely to exhibit behaviour that is overly aggressive. Ordinary people, I submit, are neither that inconsistent nor that aggressive. I am, then, forced to conclude that appealing merely to psychological importance cannot account for the strong evidence.

The practise of admitting persuasive ploys

Moreover, and perhaps more importantly, even if ordinary moraliser exhibited such unreasonable behaviour, we would expect them to at least acknowledge such conversational implicatures some of the time. Recall the previous example, where I tell my friend that 'The cold wind is killing me'. Suppose someone who is not familiar with the relevant convention asks me what I mean by that. In response, I would be willing to concede that by saying that the cold wind is killing me, I did not mean *that the cold wind is killing me*. Rather, my saying that was just a way for me to communicate the expectation that my friend closes the door, and that the fact that we both share that conventional understanding of the relevant statement is what enables him to grasp the actual meaning of my utterance.

Similarly, if Finlay is right in his interpretation, when we are confronted with similar questions on what we really mean by (say) 'Hitler ought not to have ordered the extermination of the Jews', we would expect ourselves and others to admit of the relevant rhetorical or persuasive ploys. For example, suppose a person who is not familiar with our moral discourse asks: why do you address nonrelativised moral judgements to those whom you know they do not share your ends? If Finlay's story was accurate, we would expect ourselves and others to clarify that the real meaning of

[36] As Falk would put it, by omitting specifying what others have genuine reason to do, urging others to do something is on par with 'goading' them; as opposed to 'guiding' them. From the two, only the latter involves reference to what they have genuine reason to do (1953: 151).

our utterances is to express an expectation or demand that others subscribe to our moral ends, as opposed to declaring our conviction that indifferentists really should do as we think, or that moral considerations supply agents with reasons. Recall the last time you heard *anything like*: 'All I mean by 'you ought not to kill' is that I demand that you subscribe to the end of human welfare' or 'We almost had him subscribe to our moral ends and concerns, but we failed — thus he does not have very good reason to abstain from murder'. Clearly, the practise of admitting such conventional ploys is highly uncharacteristic of ordinary participants in moral discourse.[37]

The shock provoked by the alternative verdict

Finally, and relatedly, Finlay's explanation fails to account for the shock (or the amusement, depending on your sensibilities) that is provoked by the alternative verdict considered in the previous section. Following Joyce, '[Finlay's story] doesn't explain the outrage that the audience will feel when the suffix is explicitly added': i.e. when the judge admits that 'I am merely expressing the expectation that you share our concerns, not claiming that you had actual reason to do otherwise!' (2011: 528).

3.10 Conclusion

To sum up: we have considered each kind of evidence that Finlay calls attention to in support of his relativistic interpretation of moral discourse. *Reflective evidence* was found to be untrustworthy, whereas the next three kinds (*linguistic, appraisal, reactive*) was deemed non-telling. Things were quite different with respect to the next three kinds (*address, expectation and disputation*), as they all seem to relate to considerations of normative authority. As Finlay concedes, the evidence seems to lend *prima facie* support to the absolutist interpretation, and this meant that the burden of proof was on Finlay's side to come up with an equally good explanation of it. Finlay's strategy consisted of two moves. Finlay's first move was to dispute the existence of the telling evidence — a move which was deemed of low significance to the dialectic. Finlay's second move was to account for the evidence directly and in a way that is consistent with his interpretation. Finlay's alternative explanation was persuasively rebutted for three reasons. Following the work of this chapter, we shall hold dear to the claim that the thought that moral norms invariably supply agents with sufficient normative reasons really is part of ordinary thought. More carefully, ordinary thought has it that moral norms express moral requirements – and that the latter are supposed to invariably supply agents with sufficient normative reasons. It now remains to be seen whether this thought is plausible, and relatedly, whether in holding it ordinary thought errs.

[37] I am tempted to go as far as to call Finlay's heroes in his story as manipulative and insincere (though of course he does not use such characterisations). For they attempt to make others subscribe to their ends, while simultaneously phrasing their utterances in deceitful ways; e.g. 'You have reason to be moral, so be moral' while meaning 'Even though you don't have reason to be moral, nevertheless subscribe to our morality!'. And if this is what they always do, then they do it on a regular basis. For a related thought, see (MacIntyre 1981: 73-74).

4. Practical Instrumentalism

We do things. Some of the things we do are prescribed by different normative frameworks, and the latter sometimes come into conflict with each other. Etiquette may require that I do not talk with my mouth full, whereas morality may require that I do so in order to warn you that the food contains allergens. Moreover, some of the things we do are supported by good practical reasons, and some are not. When the two conflict, I may choose to obey etiquette and disregard morality, but in doing so I may be neglecting my good practical reasons. In general, the question of what I am required to do by *this* or *that* normative framework may seem independent from the question of what reason I have to do *this* or *that* action. Yet some think that with morality, the two questions are not independent. In particular, some think that moral norms (at the least the ones representing obligations) are reason-giving: invariably supply agents (to whom they apply) with genuine reasons. The error theorist of this context denies just that; she accepts that moral norms are *supposed to be*, but she denies that they *actually are*, reason-giving (more on this distinction shortly). In this chapter, I examine the error theorist's grounds for thinking this: the truth of practical instrumentalism about reasons. I argue defensibly that the error theorist's case is premised on a theory of reasons the truth of which we have no good reason to accept.

4.1 From the *Conceptual* to the *Substantive*

In this chapter, I am interested in two *substantive*[38], and connected, questions. First, are there any moral reasons: do moral norms invariably supply agents with reasons? Second, what kinds of reasons are there? The *substantive* here contrasts with the *conceptual*. Whereas the latter kind of discussion concentrates on what a term *means* — or what a concept *is* — the former concentrates on whether there is anything (in the world at large) instantiating the concept in question, and if so, what. In asking a conceptual question, we are not begging any substantive questions; even if our concept of X is the concept of a Y, the world might not have Y and hence disappoint us (Smith 1994: 166).

In our case, we have established that if there are moral requirements, they are both non-desire-based and reason-supplying (chapter 2). The first thought is that moral requirements are supposed to have little or nothing to do with the furtherance of the agent's (to whom they apply) ends: they may occasionally require that agents act against their own ends. The second thought is that moral requirements are supposed to invariably supply agents with sufficient normative reasons for action. My opponent here accepts this conceptual account of moral requirements. More specifically, my opponent accepts that:

> If there are moral requirements, then moral norms – which express alleged moral requirements - invariably supply agents with reasons.

She then objects to the possibility of there actually being such requirements. More specifically, my opponent denies that:

[38] I borrow this term from Joyce (2001: 19) and Smith (1994: 63-66).

> There actually are moral requirements: that moral norms – which express alleged requirements – invariably supply agents with reasons: in short, that there are moral reasons.

She says: the thought that there are *non-desired based* moral requirements that *invariably supply agents with sufficient normative reasons* is one that we cannot defend. Before I explain why my opponent thinks this, I shall explain what I take 'sufficient normative reasons' and 'non-desire-based requirements' to be.

Sufficient Normative Reasons for Action

A *normative reason for action* is a reason there *is* for doing something: a consideration that justifies a prospective course of action.[39] Normative reasons point to the right or justified course of action, and they often do so regardless of how agents tend to behave or how they are like.[40] Following others, we may distinguish between *pro tanto* (or defeasible) normative reasons and *sufficient* (or *adequate)* normative reasons. A *pro tanto* normative reason is a consideration that justifies an action but which may be overruled by stronger ones. There may be a normative reason for me to go to my favourite cafe this evening (in virtue of my wanting to), and a normative reason for me not to go to my favourite cafe this evening (in virtue of my having promised to stay home with you), and yet a conflicting normative reason for me to go to church this evening (in virtue of my being a Christian). A *sufficient normative reason for action* is one that outweighs all other relevant considerations that are at play, thereby determining what the justified course of action is: the one that is most reason supported. In what follows, I may use the unqualified noun 'reason' to refer exclusively to a sufficient normative reason for action, unless stated otherwise.

Norms

I use 'norm' to merely denote a requirement that a normative framework may place on us, and I leave it an open matter whether any purported norm is also reason-supported: whether there invariably are reasons for agents to obey it. For example, in the moral case, I take it that a moral norm expresses an alleged moral requirement. This understanding of 'norm' enables us to discern the unquestionable similarities between morality, on the one hand, and any other set of requirements on the other. Examples may include anything from games such as football to well-thought-out legal

---

[39] 'Reason' is a polysemous term. First, there is the notion of an *explanatory* reason: a reason *why* someone does something. Second, there is the notion of a *motivating* reason: a reason *for which* someone does something. These two senses of 'reason' will not concern me here. See 1.2.2.

[40] This is at least an intuitive thought that we may preserve at the outset, even though some people (instrumentalists and internalists about reasons) may beg to differ. If you left your wallet at my house, and you are not sure exactly how much money you left inside, I can take a few notes without you noticing it. Given how much I want the money and that I can get away with stealing a few, it is a safe bet to suppose that I am going to do so. I may be motivated to do it, and my doing it may make perfect sense given what we know about me, but I may still have normative reason for refraining from doing so.

frameworks such as the laws of the United Kingdom. This much should be uncontroversial.

Requirements: desire and non-desire-based

If there invariably are reasons to follow moral norms (substantive matter), those reasons must be non-desire-based. We can hold on to the latter claim because we have already assumed a conceptual account of the alleged requirements expressed by moral norms — on which their content makes little if any reference to the agent's (to whom they apply) desires or concerns. In general, a practical requirement — again, in our technical sense — may be either desire-based on non-desire-based. A desire-based requirement is one that makes some reference to the agent's (to whom it applies) desires: takes some of the agent's desires as input when determining what the agent must do. A non-desire-based requirement is one that makes no such reference: its output does not depend on any such information.[41] On this understanding, etiquette and morality are sets of non-desire-based requirements, whereas the requirements of instrumental rationality are desire-based (cf. Foot 1972: 307). To give an intuitive example, we may be prepared to withdraw our judgement that our friend must go to the cafeteria after finding out that he does not want a coffee after all, but we may be inclined to hold on to the judgement that he must fulfil the promise that he made to us despite finding out that he does not want to do so.

The moral error theorist's claim

The error theorist, my opponent, claims that there are no moral reasons — by which she means: moral norms do not invariably supply agents with sufficient normative reasons for action. Call this *the error theorist's contention*. To see what she means, consider Andy:

> … who as he exits a burning building, does not feel the slightest inclination to pull the fire alarm. He sees no reason to save others when it does not benefit him, and he is unmotivated to pull the fire alarm when doing so would save dozens from dying in the fire (Bedke 2008: 102).

Let us grant the modest proposal that Andy is morally required to pull the fire alarm. On my opponent's view, it is not necessarily the case that Andy has sufficient normative reason to pull the fire alarm, though it is consistent with her thesis that Andy may (or may not) have such reason. On her story, the latter always depends on other things about Andy — namely, the particular things that Andy happens to value or care

[41] I take the distinction to be compatible with other well-known ones. A desire-based requirement may be understood as a *hypothetical* imperative, and its counterpart as a *categorical imperatives* (Foot 1972). The former is hypothetical in that its application is contingent on the agent's having a relevant end, whereas the latter is not. Another way of understanding a desire-based requirement is as an escapable prescription, and its counterpart as an inescapable (Joyce 2000). The former is escapable in that the agent can escape its application by either abandoning some of her desires or adopting others, whereas the latter is inescapable in that the agent cannot do so by doing either.

about: his *ends*.[42] So for example, the following claim would be consistent with my opponent's view: Andy has sufficient normative reason to pull the fire alarm in virtue of his desire (or concern) to save the lives of others. But — crucially — absent such desires or concerns on Andy's part, my opponent would have to claim that Andy has no reason to pull the fire alarm, if she had to be consistent with her commitments on *reasons*. This is so because my opponent is an *instrumentalist* about reasons: she thinks that the correct theory of reasons for action is *practical instrumentalism*.

4.2 Practical Instrumentalism

Why think that the error theorist's contention is true? The short answer is: practical instrumentalism. Practical Instrumentalism (henceforth, just *instrumentalism*) is the view that:

> The only reasons there can be are *instrumental reasons*: reasons favouring actions that are instrumental to the fulfilment of the agent's (for whom they are reasons) ends.

Instrumentalism, as it transpires, implies a certain criterion (or restriction) on a theory of practical reasons — which we may call *instrumentality*: the quality of a practical reason of favouring an action which is *instrumental* to the agent's (for whom it is a reason) ends.

If practical instrumentalism is true, then moral norms do *not* invariably supply agents with reasons: in short, there are no moral reasons. And if there are moral reasons, practical instrumentalism is false (Shafer-Landau 2009: 189). To see this, recall that moral requirements apply to agents irrespective of the latters' ends (e.g. Andy is still required to pull the fire alarm regardless of whether he wants to). Thus if it is always true that agents have reasons to do as moral requirements say, it must be also true that at least some reasons are non-instrumental: hence instrumentalism is false. If instrumentalism is true, on the other hand, then all reasons are hostage to the agent's (for whom they are reasons) ends, and hence moral requirements are at least sometimes unsupported by reasons: agents do not have reasons to do as they say. If this is so, the following argument for moral error theory may well be sound:

(P1) Moral requirements are non-desire-based.
(P2) Moral requirements are supposed to invariably supply agents with sufficient normative reasons for action.
(P3) Practical instrumentalism is true.
(C) There are no moral reasons: sufficient normative reasons invariably supplied by moral norms.

My opponent in this thesis is the error theorist who insofar as she bases her contention on the truth of instrumentalism, she also happens to be an instrumentalist. Henceforth, I shall therefore refer to her as the *instrumentalist*.

[42] I shall use the terms 'ends' and 'desires' here interchangeably to cover things like desire proper, valuing, having an end etc. (Dreier 1997: 87).

We must note at this point that it may be at least possible that — as a matter of perfect coincidence — instrumentalism and the view that agents have reasons to do as moral norms say are true. That is, that despite it being the case that all reasons are instrumental, it is still true that agents *happen* to have reasons to follow moral norms. For this to be true, it would have to be the case that agents happen to have some desire or concern for doing as moral norms say. Even if that were true, it would certainly not entail that such reasons are supplied by moral norms. That is, even if agents happened to have reason to do as moral norms say, that would not mean that there are moral requirements. Now, such a view may be possible, but it is certainly far from probable — and it would certainly be too much to assume at the outset (Shafer-Landau 2003: 166). At any rate, I shall here assume that if practical instrumentalism is true, there are no moral reasons: sufficient normative reasons invariably supplied by moral norms.

The unintuitiveness of instrumentalism

Insofar as it denies some common intuitions about reasons, instrumentalism is certainly unintuitive. To illustrate this, examples can be multiplied:

> Dan drives his car through a puddle so as to splash dirty water all over three pedestrians and ruin their pretty clothes. Dan sees the puddle coming, could avoid it easily and safely, knows that driving through it would harm the pedestrians, and has nothing against these pedestrians; but he does not care at all about the pedestrians, and so he does not bother to avoid the puddle (Sinnott-Armstrong 2006: 48).

We know that Dan is inconsiderate. What is particularly important for our purposes, is that Dan has no desire or concern whatever for the happiness of the pedestrians — let alone the freshness of their pretty clothes. Thus if Dan has a reason to avoid the puddle, that reason cannot be instrumental: for there is no end that Dan has such that the action of avoiding the puddle would favour it.[43] Given her commitments regarding the kinds of reasons that there can be, the instrumentalist must deny that Dan — and others fitting his description — has a reason to avoid the puddle. At the same time, and assuming that Dan desires to be home a bit earlier this afternoon in order to catch his favourite TV show, the instrumentalist may be inclined to grant that Dan actually has a (further) reason to speed up a little — thereby causing a little more misery to the pedestrians by splashing the puddle a little bit harder. In general, the instrumentalist must accept only those reasons that favour the agent's ends, and reject those that do not.

The instrumentalist does not deny the intuitiveness of the thesis that there are moral reasons, she only denies that this thought is philosophically defensible. But why think that her own view, instrumentalism, is true?

4.3 Why Think that Instrumentalism Is True?

[43] Perhaps, Mark Schroeder (2007) would object to this. In 4.6.1, I examine his liberal account of 'promotion'.

We have already noted a disadvantage of instrumentalism: it fails to account for some common intuitions about reasons. In cases when an agent has no interest or concern for doing as morality says (e.g. not stealing when doing so will only be only beneficial to her, can get away with it etc), we think that the agent has nevertheless good practical reason to be moral. The agent may not have an end which will be served by (say) not stealing, we think, but she nevertheless has a reason not to do so: namely, that doing so will be unfair to others.

In general, a theory of reasons may be judged depending on how well it fares with respect to this first criterion: *extensional adequacy*. A theory about X is *extensionally adequate* insofar as it measures up to our intuitions about the extension of X: gets the (intuitively) *right* results. Suppose there is a definite number of 100 such intuitions, and that a theory gets all of them right: the theory is extensionally adequate *as such*. Yet a theory may well get some right and some wrong results. For example, a theory may account for the intuitive thought that I have a reason to go to the cafeteria in virtue of my desire to have coffee, while fail to account for the idea that I have a reason to pull the fire alarm to save the lives of others. Insofar as it gets right results, it is *to that extent* exensionally adequate. Thus a theory A may be *more* (to a greater extent) extensionally adequate than a theory B, in case it gets more intuitive right results than its counterpart, while it may be true that neither is extensionally adequate *as such*, in cases where are some intuitive results that both fail to account for.

As we see in upcoming sections, instrumentalism is extensionally adequate to *a certain extent*, though not extensionally adequate *as such*. The problem for instrumentalism is that it has a powerful rival: non-instrumentalism, which can account for a greater number of — if not all — intuitions about what agents have reason to do.

> Non-instrumentalism: there are reasons favouring actions that are instrumental to the agent's (for whom they are reasons) ends, just as there are reasons favouring actions that are not instrumental to the agent's (for whom they are reasons) ends.

Non-instrumentalism seems to be in a strong position with respect to *extensional adequacy*. For it seems to be able to recognise the existence of reasons for agents to do things that are instrumental to their ends (e.g. go to the cafeteria when they want some coffee), just as it can recognise the existence of reasons for agents to do things when there is no end of the agent that can justify the favoured action (e.g. pull the fire alarm when they have no end served by doing so). By contrast, instrumentalism seems to be unable to account for the latter types of reasons. Though the two theories are contraries — they cannot both be true — non-instrumentalism's superiority with respect to this first criterion need not be decisive as to which is the true substantive theory of reasons for action. For instrumentalists can accept that their theory fares worse with respect to the first criterion (extensional adequacy), and still maintain their claim to the true theory of reasons. They can do so insofar as they acknowledge the existence of a second — equally important — criterion on a theory of reasons: (what we may call) *metaphysical grounding*. If their theory fares better with respect to this second criterion, instrumentalists may well cast doubt on our initial verdict resulting the consideration of the first criterion.

Instrumentalists who propose the second criterion (*metaphysical grounding*) accept the status of a large subset of claims about reasons — specifically those that have little to do with the agent's ends — as intuitive, but demand that we offer a good story about how these judgements and the theory that hosts them relate to the world. They ask: what in the world makes such claims true? This is a quest for a metaphysical foundation of reason-claims: a challenge to locate their basis in something that is part of 'the furniture of the world' (Scanlon 1998: 56). As we will see, instrumentalists typically object to the metaphysical commitments of non-instrumentalism, while they remain confident about their own. If they can show that their theory is thus superior, they can justifiably claim that as intuitive as some thoughts are (e.g. that agents have reasons for action irrespective of their ends), they cannot be defended. I shall address their claim in 4.7.

Sections 4.4-4.6 are devoted to examining instrumentalism's plausibility with respect to the first criterion of *extensional adequacy*. In 4.4, I list some problems that arise for a crude version of instrumentalism. In 4.5, I develop a theory of instrumentalism that is true to its spirit — that reasons are tied to ends — yet modified as to account for a greater number of intuitive cases. This will effectively bolster the theory's plausibility with respect to *extensional adequacy*. Even though instrumentalism will be found to be in a worse position with respect to this criterion than its rival, instrumentalists will claim that their theory fares better with respect to metaphysical grounding. This claim will be scrutinised and rejected in 4.7. The conclusion of this chapter will be that instrumentalism is less plausible than non-instrumentalism, and hence that reasons need not always be instrumental to the agent's (for whom they are reasons) ends. The upshot will be that the error theorist's contention that moral norms do not invariably supply agents with reasons, which is based on the supposition that there can only be reasons that are instrumental to the agent's ends, remains unsupported.

4.4 Extensional Adequacy: Problems for Crude Instrumentalists

The formulation of instrumentalism provided in 4.2 is rather simplistic, as it *only* specifies a *necessary* criterion on what could count as a practical reason: the quality of favouring an action which is instrumental to the agent's ends. Left as it currently stands, this simplistic account of instrumentalism may lead directly to an account on which *any* end or desire can be part of what grounds a reason. Drunky Tom has indeed a present desire (or craving) for a pint of beer this morning, yet it is highly unintuitive to think that he also has a reason for doing so. Intuitively, it is Tom's desire for giving up drinking that seems to provide him with a reason this morning: namely to refrain from drinking. The curious case of Lucifer Sam seems to reveal a similar problem of the explanatory capabilities of the theory. Sam has no desire (or any other kind of psychological inclination) towards refraining from hurting others. Insofar as instrumentalism is committed to the view that the only reasons there can be are those favouring actions that are instrumental to the agent's own ends, the intuitive thought that Sam has indeed a reason to refrain from kicking his dog finds no place in the theory. The crude version of instrumentalism that emerges here is the one that its archetypal figure, the historical Hume, seemed to have held:

> Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger (Hume 1738).

A theory of instrumentalism inspired by the above motto — which implies that *any* desire, preference or the like is part of what grounds a reason — cannot be the whole story of an account of practical reasons. It may have become clear at this point that instrumentalism — thus far understood — is extensionally inadequate. In fact, the theory as it stands may seem weak enough to render it false on independent grounds: without even the need for examining how well its rivals do with respect to *extensional adequacy*. However, before we deem the theory inadequate, we shall take the required steps to strengthen the theory to the best of our ability.

I understand a crude instrumentalist to be someone who is disposed to accept the spirit of instrumentalism (i.e. reasons are tied to ends), but who does not take an effort to work out her theory as to make it more intuitive. In what follows, I isolate a major problem for crude instrumentalists in an attempt to motivate a refined version of it. I then draw on attempts by Smith and Joyce to develop a neo-Humean version of instrumentalism — which will enable the defender of the spirit of instrumentalism to accommodate basic intuitions on reasons. This work is supposed to strengthen the case for instrumentalism as a theory of reasons. Having strengthened the theory to the best of my ability, I will have paved the way for my crucial claim: instrumentalism fails as a theory of practical reasons, not because its crude version fails to accommodate some basic intuitions; rather its failure must be attributed to its *spirit*. Elaborating on this claim will have to wait for 4.8.

4.4.1 Failure to Exclude 'Bad' Desires

A major problem for crude instrumentalists is their inability to exclude desires or ends that (intuitively) seem to be blatantly irrational from the correct explanation of what agents have reason to do. A relevant group of such things are impulses: desires that are sudden and unreflective. By failing to exclude such desires, instrumentalists may allow for reasons that there are not (cf. Schroeder 2007: 84). After losing a couple of hands in a row, a poker player may become increasingly frustrated and develop a sudden urge to increase her stake, despite having bad cards. In doing so, she ends up losing a large portion of her bankroll. It does not take more than a few minutes for her to calm down and acknowledge the silliness of her move. 'I wasn't thinking straight', she says, 'I had no reason to do it, I was being irrational'. Whereas what she *really* cared about was the safety of her bankroll, that sudden impulse urged her to go against that basic concern of hers. By failing to differentiate between 'good' and 'bad' desires, the crude instrumentalist is unable to explain why only one desire (her caring for her bankroll) can be deemed reason-giving, while the other desire (to increase the stakes) must not be deemed so.

Note that impulses are not the only instances of 'bad' desires. In their more reflective moments, people may rightly choose to separate themselves from addictive desires, compulsive desires, among others.[44] What all these groups have in common is their perception by the agent as 'external forces' alien to her real identity: things they would rather get rid off than to follow (Shemmer 2007: 328). Importantly, these seem

[44] For a longer list of such psychological items, see (Smith 1994: 154).

to be desires that seem intuitively not reason-giving: ones that should not feature in the correct explanation of what agents have reason to do. A notable instance of those is what Goldman refers to as '*just* feeling like it': 'Just feeling like it is not a reason to do something, but reflects the lack of a reason' (Goldman 2009: 108). The modifications supplied in the next section will enable the instrumentalist to overcome such difficulties.

4.5 The Neo-Humean (Version of) Instrumentalism

By overcoming the deficiency discussed above, instrumentalists can move in the right direction as a theory of practical reasons. In what follows, I take on the task of improving crude instrumentalism, effectively elevating the theory to a family of views which often goes under the name of *neo-Humeanism*.

4.5.1 Distinguishing Between Objective and Subjective Reasons

To make instrumentalism more plausible and hence a serious contender for a theory of practical reasons, certain modifications need to be made. The first modification — or rather an acknowledgement — is the introduction of a distinction between subjective and objective reasons. As Joyce elegantly puts it:

> Common sense demands a distinction between reasons that an agent is aware of and those he is ignorant of (Joyce 2001: 53).

Recall the previous example of the poker player, and suppose that (unbeknown to her) it turns out that the other players hold worse cards than her, thereby making a potential increase in her stakes advantageous. Let us begin by asking: would it now be *rational* for our player to act in an impulsive way and increase the stakes? To some people, it may seem *un*intuitive to say that increasing the stakes would be here an instance of a rational action — and I am inclined to agree. Given that (for all she knows) the chances of the other players holding worse cards than herself is very low, it would be somewhat irrational for her to increase the stakes. In a certain sense of *rationality*, it 'pertains to what we do with the desires and beliefs that we have; it doesn't guarantee or require that the agent is armed with truth' (Joyce 2001: 74).

At the same time, however, it would be a serious omission of a theory of practical *reasons* not to acknowledge the fact that (in some sense) the player has a reason to increase the stakes; as doing so will dramatically increase her bankroll, thereby satisfying one of her principal desires (cf. Kolodny 2005). Instrumentalism may begin to accommodate both of the two observations made here in the following way: The player has an *objective* reason to increase the stakes, even though her negligence of that reason makes it rational for her to not do so. As Williams would put it: given that the other players hold worse cards than herself, it is just very odd to say that she has no reason to increase the stakes, and natural to say that she has a reason to do it, *although she thinks* that she does not have any (Williams 1981: 102). This last sense of reason — the one that an agent *thinks* she has — we may follow tradition in calling a *subjective reason*. In our example, that is the reason to *not* increase the stakes. An objective reason must be here understood as one that would be recognised as such

by the agent (for whom it is a reason) just in case the agent enjoys a state of full relevant information. Cullity and Gaut spell out the relation between the two senses of reason:

> I have a subjective normative reason to Φ whenever I believe that I have an objective normative reason to Φ (Cullity 1997: 2, cf. Schroeder 2007: 10-13).

As a theory of practical reasons, instrumentalism may identify the agent's reasons with her *objective* reasons. The key thing to note here, however, is that the agent's objective reasons are still tied to her ends; an agent cannot have an objective reason which would not further *any* of her ends. This makes the theory of reasons that emerges from the distinction still an instrumentalist one. For the sake of simplicity, I will henceforth use 'reasons' to refer to objective reasons unless stated otherwise.

4.5.2 The Neo-Humean Instrumentalism

As we have already seen, the crude instrumentalist's inability to distinguish between 'good' and 'bad' desires renders her theory unintuitive with respect to a number of cases (e.g. impulses may be just as good indicators of reasons as any other desire). This brings us to a second modification of instrumentalism: the introduction of a principled way of distinguishing between those desires or ends that give the agent reason to act, on the one hand, and those desires or ends that do not give the agent reason to act, on the other. A modification along these lines will effectively elevate crude instrumentalism to the neo-Humean family of views. We may attribute the following motto to the neo-Humean instrumentalist: Not every desire or end gives you a reason for action.

In the context of a theory of practical reasons, *distinguishing* between desires amounts to specifying the conditions under which desires give reasons, and those under which the do not. Following Goldman, for a theory which maintains that 'all practical reasons derive from our motivations or concerns', it 'becomes pressing to clarify when and which desires create reasons and which do not' (Goldman 2009: 108). A central aspect to such projects is the introduction of a double restriction: what Goldman calls the *information requirement* and the *coherence requirement* (Goldman 2009: 45, 57). The former amounts to a demand that the agent be in a state of full relevant information when deliberating about her reasons: i.e. in the context of instrumentalism, when considering which desires of hers provide her with a reason.[45] The latter is a demand for a certain kind of harmony between the agent's desires as a whole: for '[having] certain desires or motivational states *given other* desires or concerns' (Goldman 2009: 57, emphasis added). We are thus owed an account of those *other* desires.

[45] Goldman's own example from (Goldman 2009: 35): 'I can have a reason to apply for a grant given my desire for research support, but if I am unaware of the availability of the grant, I will not be motivated to apply for it'.

A promising neo-Humean account of the relevant criticisability/appraisability of desires is developed by Joyce, and it comes roughly down to the following description[46]:

> For any actual agent, S, we may speak of an idealized counterpart of S, S+, who deliberates from existing motivations in a flawless manner (is fully reflective and imaginative) and who has all and only relevant true beliefs. There will be certain actions that S+ would advise S to perform. Those actions are the ones that S has normative reasons to perform (Joyce 2001: 75).

On this account, what distinguishes our idealised counterparts is their ability to be fully reflective, fully imaginative and to have all and only relevant true beliefs. We may call those particular desires that survive the aforementioned process as the agent's *ultimate* desires; in the sense that the agent is disposed to identify with them in her most reflective and imaginative moments — that are not considered by the agent as 'alien external forces'.[47] As Joyce rightly notes, the resulting theory is still a version of instrumentalism, as it understands reasons only as means to some of the agent's ends (Joyce 2001: 77). To see this, consider that despite the fact that the resulting theory includes a certain method of criticisability of desires, the standard by which the criticism is conducted is always comprised by one (or more) of the agent's *own* desires. On any account deserving of the name 'instrumentalism', such ultimate desires must fall outside the jurisdiction of criticisability. As Joyce puts it, on an instrumentalist's account, '[a bad] desire has not been, on some manner, properly examined by the agent' (Joyce 2001: 158). Once such an examination has been properly conducted, the characterisation 'bad' ceases to be fitting when ascribed to any desires that have survived it (cf. Lillehammer 2000).

4.5.3 The Apparent Merits of the Resulting Theory

The above modification makes the resulting version of instrumentalism far more extensionally adequate than its crude ancestor, and as a result the theory is now a far more serious contender for the true substantive theory of practical reasons. Let me briefly illustrate why this is the case, before I embark on explaining what exactly I think is wrong with the theory.

Recall the case of the poker player, who after becoming increasingly frustrated with her recent losses, increases dramatically the stakes for the next hand — ignoring how unlikely it is that her competitors hold weaker cards than herself and, as a result, how likely it is that she will lose a large sum. This is a paradigmatic case of someone having a present desire to do something (increase the stakes), but at the same time has (intuitively) no reason to do so. (Neo-Humean) instrumentalists have now the resources to explain just that. To do so, the need to observe that the player's idealised counterpart, who deliberates in conditions of full reflection, imagination and

[46] Note that neo-Humean accounts of rational criticisability can differ regarding their exact content. As it will later transpire, this will not be of much consequence here. For alternative such accounts, see (Sinclair 2017: 211) and (Schroeder 2007: 131).
[47] I borrow the phrase from (Shemmer 2007: 328).

information, would suggest that she do not increase the stakes. She would do so because having full information would enable her to know how unlikely it is that the other players hold weaker cards than hers, being fully imaginative would enable her to see the likely consequences of taking such a risk for her bankroll, and being fully reflective would enable her to realise what her ultimate desire is: to do well in the game. The fact that the player's idealised counterpart would suggest that she does not increase the stakes explains that: (i) the player has no reason to increase the stakes, even though she has a desire to do so, (ii) the player has a reason not to increase the stakes, even though she has no (present) desire to do so.

Now, suppose that it is true that the other players hold weaker cards than our player and that, as a result, following her impulse on this particular occasion could very likely be beneficial to increasing her bankroll. Suppose also that we subtract from the player's idealised situation the condition of full information, so that now the player's idealised counterpart is fully reflective and imaginative but unaware of some of the fact that the other players hold weaker cards than hers. Through the distinction between subjective and objective reasons, the (neo-Humean) instrumentalist can now account for the following two intuitions: (i) the player has *some kind of* reason to not increase the stakes, (ii) the player has nevertheless no reason to increase the stakes, and a reason to not increase the stakes. That is: (i) a subjective reason to do so, which is a result of what she knows (or does not know) of her circumstances, and (ii) an objective reason to not do so, which is a result of her being 'armed with truth' (cf. Joyce 2001: 74).

4.6 Extensional Adequacy Revisited: Problems for Neo-Humean Instrumentalists

Crude instrumentalists had the problem of being committed to saying that 'bad' desires — ones that we intuitively have no reason to satisfy — nevertheless entail reasons[48]. The problem was mitigated through the introduction of a process of criticisability in which an ideal counterpart would suggest which desires are worthy of pursuing and which are not. The upshot was that a lot of 'bad' desires could well be eliminated through proper reflection, imagination and full information: conditions characteristic of our ideal counterparts. The resulting theory was found able to rule out a great number of unintuitive results; e.g. to say that the gambler's impulse to increase the stakes is not reason-supported, in virtue of the idealised counterpart's suggestion that the player do not do so.

At the same time, however, it seems at least possible that some such desires — ones that we intuitively have no reason to satisfy — will survive the process of criticisability proposed, and for that matter any such process (cf. Schroeder 2007). Consider Manson's desire for massacre, which survives the process of criticisability. Given that Manson's idealised counterpart, Manson+, would advice Manson to massacre, it seems to follow that even neo-Humean instrumentalists would be

[48] It is tempting here to speak of desires that are reason-*giving*. But to be perfectly neutral on whether - on the instrumentalist's view - desires give or merely ground reasons, I shall prefer to speak of desires that *entail or ground* reasons. For on a plausible instrumentalist view, a desire can ground a reason without itself being the reason-giving feature (cf. Pettit & Smith 1990, Schroeder 2007).

committed to saying that Manson has a reason that he intuitively does not seem to have: to commit massacre. In general, there seem to be cases in which an agent seems (intuitively) to have no reason to Φ, despite it being the case that according to the theory she does have a reason to Φ. For being thus permissive, instrumentalism is sometimes accused of allowing for *too many reasons* (Schroeder 2007: 84).

Moreover, consider the fact that the (fictional character) Schmanson has no desire of concern for the welfare of his children, and that not even his idealised counterpart would suggest that he care for them. It seems to follow that neo-Humean instrumentalists would be committed to saying that Manson has no reason to help his own children, yet another unintuitive result. In general, there seem to be cases in which an agent seems (intuitively) to have a reason to Φ, despite it being the case that according to neo-Humean instrumentalism she does not have a reason to Φ. For being thus restrictive, instrumentalism is sometimes accused of allowing for *too few reasons* (Schroeder 2007: 103).

Thus for all its strengths, the resulting theory still faces challenges with respect to the first criterion, *extensional adequacy*: it is still inadequate by being either too permissive or restrictive. I do not think these challenges can be overcome by neo-Humeans. But some people do. In what follows, I consider Smith's (1994) and Schroeder's (2007) proposals for an improved (with respect to extensional adequacy) neo-Humean instrumentalism.

4.6.1 Smith's Account of the Idealised Counterpart

It would be an omission at this point not to mention Michael Smith's (1994) optimism that the kind of deliberation in which the criteria of *full* reflection, imagination and information obtain 'can both produce new and destroy old underived [ultimate] desires' (Smith 1994: 158). For Smith, in the kind of deliberation specified in the neo-Humean's account, desires are not beyond the scope of rational criticisability (Joyce 2001: 70). Our idealised counterparts, who deliberate correctly in the presence of the aforementioned conditions, the idea goes, can and occasionally will suggest that we either diminish or generate new ultimate desires. As he puts it:

> [W]e can not only explain how we might come to have new [ultimate, underived] desires as a result of such reflection, but that we can also explain how we might come to lose old [ultimate, underived] desires as well (1994: 160).

Smith seems to suggest that our idealised counterparts will not only see things differently — as it were — but that they would see them differently to such a degree that they would (on some occasions) desire entirely different things from their actual counterparts. They would do so, he thinks, because they would realise that certain ultimate desires of their actual counterparts do not fit well into an overall set of desires that is more justifiable than the one that they actually have. As I understand it, the idea boils down to this: one set is judged by the ideal counterpart as being (in and of itself) more desirable than another. Before I say more on this, it is noteworthy how this gives Smith's version of neo-Humean instrumentalism all that a defender of the dogma would want. Namely, that there could be a convergence of reasons such that any agent

— no matter what her present or actual desires might be — would have the same reasons that we (intuitively) think she has. To see how Smith's account of criticisability would bring this result, we only have to remind ourselves of the following, distinctively instrumentalist, proposal: a convergence of desires among possible agents would entail a convergence of reasons among them (cf. Joyce 2001: 76, 78). For simply put, if every agent who deliberates correctly — on Smith's view — would end up having the same desires, either by generating new ones or extinguishing others, it follows that every agent would also have the same reasons for action, insofar as on the instrumentalist's view reasons are tied to the desires of the agent (for whom they are reasons).

To use an example, suppose that Manson has an ultimate desire for hurting others for fun, and that he comes to deliberate correctly (with full reflection, imagination and the like). On Smith's view, Manson will eventually come to desire the elimination of his desire, possibly because he would realise that the desire does not fit well into a more justifiably desiderative set. Correct deliberation — on Smith's view — may even supply Manson with a wholly new desire; e.g. to care for others. And given that this is entirely generalisable, the resulting view is (perhaps surprisingly) still a version of instrumentalism, in that it understands reasons only by reference to some of the agent's desires: i.e. the ones the agent would come to have were she to deliberate correctly.

If Smith's account of the idealised counetrpart can be shown to be plausible, the resulting version of neo-Humean instrumentalism will likely be able to accommodate our intuitions about the extensions of reasons. For even if some reasons that we think agents have (e.g. Manson's reason to care for others) are not in the business of furthering the desires of actual agents, they may well be furthering the desires of idealised agents. For example, Manson may *actually* not care for the lives of others, and hence may seem that he does not have a reason for caring, but Manson+ may well suggest that Manson care for the lives of others, and hence make it the case that Manson has a reason to care after all.

Now, I shall not take sides on whether there are some desires (or a set thereof) that are (in and of themselves) more desirable than another. What I shall object to is the possibility that the formation or elimination of new ultimate desires can be suggested by our idealised counterparts. This is the reason I do not find Smith's account of the idealised counterpart convincing: his suggestion that the latter would (or could) suggest that we (actual counterparts) have *entirely different* desires than they ones we presently have. For I do not think that the conditions in terms of which we understand an idealised counterpart (full reflection, imagination and information) are substantive enough as to alter the content of our actual desires *entirely*. As Joyce puts it:

> I see no ground for doubting that this may be entirely generalizable: that the things that the ideal version of myself wants for me always depend on what I actually desire (Joyce 2001: 76).

My own view is that were agents to think correctly, in conditions of full rationality (what in 4.8 I refer to as rationality in the broad sense), they would eventually come to recognise what their reasons for action are. And crucially, I submit,

those reasons need not be connected to any of their actual ends, or those of their idealised counterparts. What I do not find plausible, is that agents would also come to the formation of desires that would be in alignment with (what they would correctly recognise as) their reasons. For the following two seem to be independent matters: what I think I have reason to do, and what I want to do.

4.6.2 Schroeder's Way

Schroeder's account is indeed a version of instrumentalism, insofar as it maintains that:

> Every time someone has a reason, she must have some desire that explains why (Schroeder 2007: 108).

In the context of improving the theory's extensional adequacy, Schroeder is interested in two things. First, to explain why agents who have strong desires for things that they intuitively have no weighty reason to pursue, actually have no such reasons. In explaining, for instance, why Manson does not have weighty reason to commit massacre despite his strong desire for doing so, Schroeder will have effectively dealt with the *too many reasons* objection. Second, to account for the existence of so called *agent-neutral reasons:* reasons for action for everyone who finds herself in some fixed circumstances *C* (Schroeder 2007: 17). Consider again Schmanson's complete disregard for the welfare of his children. Intuitively, it seems that Schmanson still has a reason to care for his children regardless of whether he cares about it or not, just as everyone has a reason to care for their own children regardless of whether they care or not. In accounting for such reasons, Schroeder will have effectively dealt with the *too few reasons* objection. In what follows, I explain how Schroeder deals with a specific version of the too many reasons objection, and with the too few reasons objections. I am with Schroeder on the former, but I object to his way of dealing with the latter objection.

The *too many reasons* objection

We must note that the too many reasons objection comes in two versions: one about *pro tanto* reasons and one about *sufficient* reasons. The former concerns cases in which an agent has foolish desires for things that he intuitively has no reason *at all* to pursue: i.e. not even a pro tanto reason. Schroeder thinks that it is a problem for any theory of reasons to hold that foolish desires, such as Manson's desire to commit massacre, entail that agents have even *some* (i.e. pro tanto) reason to satisfy such desires[49] (84). As I am here primarily interested in *weighty*[50] (i.e. sufficient) reasons for action, I shall remain neutral on whether this is a problem for a theory of reasons. That is, I shall not

[49] In dealing with this, Schroeder acknowledges that agents may occasionally have desires that are not reason-providing in the strong sense: they do not provide weighty or sufficient reasons for action (e.g. Manson's desire for massacre). But he wants to maintain, at the same time, that such desires do provide *some* reason for action in virtue of the agent's (to whom they provide reason) desires. Schroeder adds that for pragmatic purposes, we tend to not mention these very light-weight reasons, and that this gets confused with the thought that there are no such reasons at all. This is an interesting claim developed carefully in pp. 92-98, though I think ultimately flawed.

[50] By 'weightiest' or 'strongest' reasons I shall here mean *sufficient:* i.e. 'as weighty as the reasons that are obviously sufficient to decide what we should do' (94).

take sides on whether agents have pro tanto reasons for satisfying their foolish desires. What I will be interested in is the second version of the too many reasons objections, what I shall call the *too many sufficient reasons* objection: the one concerning cases in which an agent has foolish *strong* desires for things that he intuitively has no *strong* reason to pursue.

Let us grant that Manson has a very strong desire for committing massacre — even stronger than his desire to stay out of prison. Why do neo-Humeans (or at least some of them) think that (say) Manson has a weighty reason for committing massacre? Because of their adherence to a certain account of the weight of reasons, *proportionalism*: 'when a reason is explained by a desire, … its weight varies in proportion to the strength of that desire' (Schroeder 2007: 98, 102, 123). In short: proportionalism holds that the *stronger* the desire, the *weightier* the reason. Thus, on proportionalists' view, Manson's strong desire for massacre is *proportional* to how much reason he has for committing massacre: *weighty* (or *sufficient*) reason.[51]

Proportionalism, or any account similar to it, is a standard way for neo-Humeans to account for the weight of reasons (ibid). Those who accept it open themselves to the too many sufficient reasons objection. Schroeder rejects proportionalism, thereby seperating himself from this infectious thesis.[52] He is now not committed to saying that agents with *strong* desires for foolish or despicable things have *weighty* or *sufficient* reason for doing such things. But he now needs an alternative account of the weight of reasons for his neo-Humean theory of reasons to work. For if, as the neo-Humean thinks, reasons are explained by desires (of the agent for whom they are reasons), and some desires are stronger than others, the pressure is on the neo-Humean who denies that strong desires provide weighty reasons to explain why this is so[53] (99).

The account of the weight of reasons that Schroeder proposes boils down to this:[54]

---

[51] Note that the idealised counterpart model introduced earlier may well be considered a version of proportionalism. First, the idealised counterpart model may be understood as an account of the weight of reasons, insofar as those actions recommended are the ones favoured by the agent's strongest reasons (i.e. *sufficient* or *weightiest*), and those actions disregarded may be either favoured by the agent's not-so-strong reasons or no reason at all. Second, and importantly, the model seems to guarantee that the agent's weightiest reasons will go along with the agent's strongest desires. For why would a fully reflective, imaginative and well informed counterpart of Manson*,* who deliberates from *existing* motivations, not recommend massacre when the actual Manson has a very strong desire for doing so? The too many sufficient reasons objection arises for neo-Humeans who subscribe to the idealised counterpart model because it is at least *possible* that foolish or despicable desires (e.g. Manson's desire for massacre) would be recommended by an idealised counterpart (cf. Schroeder 2007: 86).

[52] Schroeder cites as the main reason for rejecting proportionalism the fact that 'the weight of reasons is a normative matter' (100). As proportionalism effectively reduces the weight of reason to the strength of a psychological state (desire), the account fails to fulfil this expectation.

[53] Some would go as far as to say that the viability of Schroeder's modified neo-Humeanism hinges crucially on that alternative account (Shackel 2014: 2).

[54] Schroeder provides a complex account with intricacies, many of which I shall not unveil for our purposes. Notably, he introduces a formal framework in presenting his account, which

> A pro tanto reason for doing A is weightier than a pro tanto reason for doing B just in case it is *correct* to place more weight on A than on B (140).

Consider a potentially problematic case for neo-Humeans: Manson who has a very strong desire for massacre. On proportionalists' view, Manson has *a very weighty* reason for massacre. This is extensionally wrong — it seems that Manson has no such reason. How does Schroeder get around this? By reminding us that on his account of the weight of reasons, strong desires do not entail weighty reasons. Thus Manson may have a very strong desire for massacre, which gives him a pro tanto reason for doing so, but it may well be incorrect to place weight on it: it is not a weighty reason. I have no reservations in accepting this result, nor do I have any in accepting Schroeder's account of the weight of reasons.[55]

The *too few reasons* objection

The too few reasons objection is the charge that neo-Humeans are unable to account for reasons that agents intuitively have. In particular, this is the charge that neo-Humeans cannot account for the existence of *agent-neutral reasons*: reasons for everyone who finds herself in some fixed circumstances *C*. Insofar as neo-Humeans think that every time someone has a reason, she must have some desire that explains why, they open themselves to this objection. To see this, consider Schmanson's complete disregard for the welfare of his children. Intuitively, it seems that Schmanson still has a reason to care for his children regardless of whether he cares about it or not, just as everyone has a reason to care for their own children regardless of whether they care or not. In the absence of a desire that would explain why such agents have the reasons they intuitively have, neo-Humeans cannot consistently maintain that reasons are always explained by desires and that such agents have such reasons.

Schroeder attempts to get around the problem. He rightly acknolwedges that neo-Humeans must explain both how any agent has the agent-neutral reasons she intuitively has, and how those reasons (that are for everyone) are also weighty reasons (for everyone) (106). If Schroeder can show that agents with no apparent desire served by the reasons they intuitively have, nevertheless have those reasons, I will grant that those reasons are also weighty ones. The main reason for this is that I subscribe to Schroeder's account of the weight of reasons mentioned above. In short: it may well be correct for Schmanson to place more weight on that reason (to help his children) than in any other that may be at play, irrespective of how strongly or lightly he feels

---

includes among other things talking about *sets* of reasons instead of just reasons: pro tanto or sufficient.

[55] It is noteworthy that Schroeder (136) and his critics (Shackel 2014) worry that the account of the weight of reasons is viciously circular. On my view, the account is of course circular, as it defines the weight of reasons in terms of correctness, and correctness in terms of the weight of reasons (140-141). But I do not think this is a problem for any account of the weight of reasons. I think that the notion of a sufficient normative reason for action is primitive, and hence that any explanation of why some consideration favouring Φing is a weighty reason for Φing is doomed to make reference to the fact that it is simply correct to place more weight on it than on other pro tanto reasons that may be at play (and correspondingly, any explanation of why it is correct to place more weight on a consideration favouring Φing is doomed to make reference to the fact that that consideration is a weighty reason).

about either: irrespective of the strength of his desires. But as I see it, the real challenge is to account for the existence of pro tanto reasons for agents to do things that they apparently have no desire for (or served by doing them) in the first place. What is the desire that Schmanson has and which explains why he must help his children?

The 'massive ovedetermination of reasons'

Schroeder accepts that: for R to be a reason for X to do A requires X to have some desire whose object her doing A *promotes* (108). What he does not accept, and nor should he, is that: for every reason, there must be some *single* desire that explains why each person who has that reason has it (ibid). To see what he means, suppose we both pass by the river and notice a drowning child. Intuitively, we both have a reason (an agent-neutral reason) to help the child. On Schroeder's view, we may have that reason in virtue of different desires; your reason may be explained by your desire for the child's well-being, and my reason may be explained by my desire for appearing in the evening news. In this sense, there need not be a single desire on which an agent-neutral reason depends. This seems to me like a genuine possibility.

Moreover, Schroeder rightly notes, it is possible that an agent-neutral reason for X to do A is explained by more than one desire that X has. For example, suppose that I want both to see the drowning child live another day, and to appear in the evening news. Insofar as the action of saving the child promotes both desires, my reason for doing the action is 'doubly determined' (108). In general, a reason-giving fact is *overdetermined*, in case it can be explained by more than one desire (of the agent for whom it is a reason-giving fact). As before, this seems to me like a genuine possibility.

Schroeder makes use of the above observations in introducing the notion of *a massively overdetermined reason*: a reason for anyone, no matter what she desires, simply because it can be explained by *any* possible desire (109). In short, Schroeder's proposal for how there can be agent-neutral reasons is that such reasons are *massively overdetermined*. *If* agent-neutral reasons, such as Schmanson's reason for helping his children or yours and my reason for saving a drowning child, are massively overdetermined, it follows that *any* agent with *any* desire will have them. This is a hypothesis that, at least in theory, may enable the neo-Humean to overcome the too few reasons objection. At the same time, it features a big 'if'. For it seems unlikely that such reasons *are* massively overdetermined. For what is the desire that *Schmanson* has and which explains why he has the agent-neutral reason to help his children?

The weak promotion relation

Schroeder realises that his proposal is in need of further motivation, and so he attempts to make reasons 'easier to come by': to weaken the necessary conditions for some facts to count as reason-giving for agents. In order to do this, he introduces a 'sufficiently weak' account of the promotion relation that features in the neo-Humean's view of how reasons are explained by desires: for R to be a reason for X to do A requires X to have some desire whose object her doing A *promotes.* As he notes, the relevant notion of promotion has been historically understood in terms of either *identity*: an action counts as promoting a desire just in case it is the object of the desire, or *necessary means:* an action promotes a desire just in case it is a necessary means to the object of that desire (111). Schroeder's favoured account is in terms of *increased*

likelihood: an action counts as promoting a desire just in case it increases the likelihood of the object of that desire relative to the agent's (who has the desire) doing nothing.

The account of the weak promotion relation is important towards accounting for the existence of agent-neutral reasons. To see this, suppose that Schmanson is passing by the river and notices the drowning child, and note that the example is a paradigm of an apparently hard case for neo-Humeans. We know that he has no desire for seeing the child live another day, nor for appearing in the newspaper news. Thus he has no reason for saving the child which is explained by *those* desires. Still, it is reasonable to think, Schmanson has other desires. Let us suppose that he has a desire for meeting his friend at the cinema, a desire to eat a burger tonight, and a desire for a good night sleep, and that these are the only desires he has. On a strong conception of the promotion relation, in which promotion is understood ash either *identity* or *necessary means*, it seems highly unlikely that Schmanson's reason for saving the child can be grounded on any of these desires. First, he has no desire for saving the child, and second, the action of saving the child is not necessary for achieving the object of any of his desires.

Now on Schroeder's weak conception of the promotion relation, perhaps things *could* be different. All the neo-Humean needs is that Schmanson has a desire the satisfaction of which could be made more likely by the action of saving the child relative to Schmanson's doing nothing. For example, take his desire for meeting his friend: would the action of saving the child increase the likelihood of satisfying that desire were he to do nothing else? If the answer is 'yes', then Schroeder will think his theory shows that Schmanson has a reason to save the child. If 'no', then we will have to ask the same question with respect to Scmanson's remaining two desires.

Assessment

Schroeder gives a formal scheme of an account of agent-neutral reasons, which amounts more or less to suggesting a mere possibility that such reasons can be given a neo-Humean defence. But at the same time, he does not give enough detail to convince us that this account can generate the agent-neutral reasons that we think agents have. That is, when it comes down to those actual cases that are considered hard for the neo-Humean, such as that of Schmanson's reason to help his children or save the drowning child, it seems perfectly possible that such agents will have no desire whatever served by doing what they intuitively have reason to do. And though this is a problem that neo-Humeans subscribing to a strong promotion relation are particularly vulnerable to, those like Schroeder who subscribe to a weak promotion relation are vulnerable as well.

It is noteworthy that some actions may well be *massively overdetermined*: promoting any possible desire. For example, breathing (if it is to be considered an action) may be promoting any desire such that any agent with any desire will have a reason to breath. But crucially, what Schroeder needs to show is that the actions that *we intuitively judge that we all have reasons to perform,* e.g. the action of saving a drowning child, are like that. In order to do that, Schroeder must examine the particularly hard cases of Schmanson passing by a drowning child and illustrate how *Schmanson* has a reason *to save the child.* In the absence of doing this, Schroeder gives

a recipe which by itself does not seem to guarantee that we will get the (extensionally) right results.

Perhaps, one of Schmanson's desires can be promoted in the above weak sense. For example, were he to save the child, he would *perhaps* make the afternoon news, the restaurant owner will *perhaps* see the news and hence *perhaps* be able to recognise him. This will *perhaps* lead to making sure that Schmanson gets served on such a busy night ahead of other customers. And if he were to do nothing else, e.g. go to McDonalds, the likelihood of satisfying his desire for eating a burger would *perhaps* be increased. But one thing that the multitude of all these 'perhaps' suggests is that even in this weak sense of promotion, Schroeder's account may succeed at generating agent-neutral reasons with respect to some cases, but it need not do so with respect to all cases that we want it to. For whether agents have desires promoted (in any sense available here) by doing what they intuitively have reason to do seems to be a genuinely contingent matter. As I have noted in an earlier section, to assume that self-interest (in some sense) and morality (or at least agent-neutral reasons) always coincide is a rosy view which would be too much to assume at the outset (Shafer-Landau 2003: 166).

It may be also noteworthy that whereas the strong sense of promotion (*identity* or *necessary means*) seems to be the standard sense, the weak sense of the promotion relation that Schroeder proposes seems eccentric and far-fetched. I have a desire for ice-cream, and there are two possible actions: go to the ice-cream shop, or go to the beach and collect 100 000 tiny rocks. Intuitively, which of the two actions promotes my desire for ice-cream? The former. But on Schroeder's weak conception, the latter may be also thought of as promoting my desire. For collecting 100 000 tiny rocks will *perhaps* lead to the ice-cream shop owner who will be *perhaps* passing by see me, and that will *perhaps* impress him and … (so on and so forth). And therefore, relative to my doing nothing else, collecting 100 000 tiny rocks may increase the likelihood of satisfying my desire. Granted: in a weak sense, the latter action may be thought of as promoting my desire, but it simply does not promote it enough to merit the standard sense of promotion.

4.6.3 The Dialectic As It Stands

We started with: Are there any moral reasons — i.e. do moral requirements invariably supply agents with sufficient normative reasons? I assumed that if the instrumentalist is right, and there can only be reasons (pro tanto or sufficient) favouring actions that are instrumental to the agent's ends, the answer is 'no'. Call this *the incompatibility* of instrumentalism (or some form thereof), and the possibility of there being moral reasons: sufficient normative reasons invariably supplied by moral norms.

I then asked: why think that instrumentalism is true? In answering this question, I picked out two criteria that any theory of reasons should be held accountable to. In relation to the first (extensional adequacy), *crude* instrumentalism seemed to do very badly. Having fixed some problems of their crude ancestors, *neo-Humean* instrumentalists were then in a position to claim that their theory surpasses certain deficiencies relating to *extensional adequacy*. Yet as we have just seen, even neo-Humean instrumentalism seems to be committed to some unintuitive results, and

to that extent failing with respect to this first criterion.[56] Crucially, the theory's failure to deliver the right results comes in stark contrast with its rival's, non-instrumentalism's, success at doing just that. For non-instrumentalists can hold on to the claim that, say, Manson has no reason to commit massacre despite him having a strong desire for doing so, or that Schmanson has a reason to support his children despite him having no desire or concern for doing so. They can do so given their lack of commitment to the thesis that all reasons must be somehow explained by some of the agent's (for whom they are reasons) ends.

This marks an interesting turn in the dialectic. Instrumentalists may now concede that their theory is committed to *some* unintuitive results, and that it is hence *to some extent* extensionally inadequate, but blame instead those very intuitions that their theory cannot account for: 'so much the worse for those intuitions', our neo-Humean instrumentalist now says (cf. Goldman 2009: 22). The instrumentalist may support this claim through the introduction of a second criterion that a theory of reasons must be held accountable to: *metaphysical foundations*. Depending on how well instrumentalism fares to that criterion, and how relatively bad its counterpart does, instrumentalists may be able to say that albeit extensionally inadequate to some extent, their theory is nevertheless the *correct* theory of reasons. And if all this can be shown to be the case, then so much the worse for a defender of moral reasons: for all we know, moral norms may not invariably supply agents with sufficient normative reasons.

4.7 Metaphysical Grounding

The work of the previous sections enables me to infer that instrumentalism, the theory upon which the error theorist's contention is based, fares worse than its rival with respect to the first criterion of extensional adequacy. As I am interested in a case that would suggest that instrumentalism and, by extension, the error theorist's contention is true, I shall lay out the following claim and proceed by considering support for it: instrumentalism is nevertheless superior with respect to the second criterion — metaphysical grounding — and hence we have good reason to think that it may be true. If instrumentalists can support this claim, they may still elevate their theory as the true — or at least the most plausible — theory of reasons.

Some effort must be taken to unpack the notion of a metaphysical grounding. A grounding claim is a claim about about what grounds (or explains) what.[57] Consider

[56] It must be pressed that our criteria for having reservations as to the truth of instrumentalism are by no means question-begging. For it might be question-begging when examining an argument in favour of moral error theory to deem it false on the grounds that it excludes *moral* reasons. This is not exactly the case here. Rather, it is the *intuitiveness* of moral reasons that gives us reason for disputing instrumentalism. And indeed, there are also other *non*-moral counterexamples to the theory give us reason to doubt its truth. A nice example is discussed by Schroeder himself: Aunt Margaret who has a desire to reconstruct her favourite scene of *Martha Stewart Living* catalogue - a desire which may well survive the rationalising process (Schroeder 2007: 84). Instrumentalism seems to be committed to saying that Aunt Margaret has a reason to reconstruct the scene, something that seems intuitively false.

[57] For a detailed discussion see Bliss, Ricki and Trogdon, Kelly, 'Metaphysical Grounding', *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/grounding/>.

the claim: Manson is evil, and the grounding claim: … in virtue of being cruel and inconsiderate to others. The latter attempts to ground the former in that it purports to explain why it is so by reference to a further feature of Manson. When asking for a metaphysical grounding, we are interesting in an answer to the following question: In virtue of what in the world are some of the theory's distinctive claims true? As we will see shortly, an answer must feature an *in virtue of* (or a *because*) that is followed by a statement that connects the theory (or its distinctive claims) to some feature of the ('fabric of the') world.[58]

4.7.1 The Instrumentalist's Claim of Superiority

I will now examine an interesting claim to the effect that the reasons advocated by instrumentalism are grounded whereas those advocated by its rival, non-instrumentalism, are not. The claim is interesting because, as I shall concede, it is initially appealing, but as I shall argue, ultimately misguided. To understand it, we need to introduce the basic structure of reason-claims:

R [F, A, Φ, C]

Where this reads: F is a reason, R, for A to Φ in C — where A refers to an agent, C gives a description of the situation in which the reason-claim applies, Φ refers to an action or type of action and F is the consideration that justifies.[59]

Our instrumentalist begins by making a subtle distinction. On the one side of the distinction, we have the reason-giving features F — or perhaps, the reason *itself*.[60] These are the features, plausibly natural facts about the world, which provides the reason. For example, suppose that by pulling the fire alarm one would save the lives of others. Naturally, we may say: the fact that pulling the fire alarm would save the lives of others *provides* (or perhaps, *is*) a reason for Andy to pull the fire alarm. On the other side of the distinction, the instrumentalist thinks, we have the facts, if any, that

[58] When Mackie considers the grounding claim that *an act is morally wrong <u>because</u> it is a piece of deliberate cruelty*, he famously asks: 'But just what in the world is signified by this "because"?'. In saying so, he expresses doubts as to the metaphysical grounding of such claims (Mackie 1977: 41).

[59] (Sinclair 2016: 1.1.2)

[60] I say 'perhaps' because some would object to F being the reason itself. Notably, Olson makes a distinction between two distinct readings of the relation between F and the reason: what he calls the 'provides-locution' and the 'is-locution'. To say that F *provides* the reason, Olson thinks, is different than to say that F *is* the reason. On his view, the former means that F, a natural fact, *grounds* a normative fact, the reason, whereas the latter wrongly *identifies* F, a natural fact, with a normative fact, the reason (Olson 2009: 169). It is noteworthy that Olson's view is that the 'is-locution' is always extensionally empty, and this makes him an error theorist about all normativity. On his view, a true reason-claim to the effect that there is a reason for an agent to Φ reduces to the claim that the agent desires to Ψ and were she to φ, she would Ψ (177). On my view, this is a mistaken view, but I shall here not discuss it as I assume throughout this thesis that the concept of a reason for action is extensionally non-empty.

ground the reason-giving features: it being the case that *F is a reason for A to* Φ *in C*.[61] Importantly, the natural facts that can do the relevant grounding are distinctively psychological states of the agent (for whom F is a reason): his desires, likings, and the like. Thus we may represent the instrumentalist's understanding of the reason-claim with the following schema:

(Instrumentalist's schema) $\underline{R\ [F, A, \Phi, C]}$
D

D stands for the agent's (for whom F is a reason) desire. The reason relation is now grounded — given the presence of the desire, D — and this is represented by the underlining. Effectively, the instrumentalist has proposed a subtle distinction between the types of things that can be reason-giving features (natural facts), and those types of things that can be grounding features of the reason relation. An indispensable part of the proposal is that the latter can only be desires (and the like) of the agent's (for whom F provides/is a reason). Given the instrumentalist's strict adherence to the dogma that all reasons are somehow connected to a psychological state of the agent (for whom they are reasons), it follows that each and every time the instrumentalist postulates the existence of a reason, there *will* be a relevant psychological state available to do the grounding for the reason.[62] Thus given the proposed distinction, the instrumentalist's story of reasons *is* grounded. What about the non-instrumentalist's?

The non-instrumentalist may well accept the proposed distinction between facts that can be reason-giving and those that can do the relevant grounding. Problems begin for the non-instrumentalist if she accepts that grounding for reasons can only be given in the form of psychological states of the agent (for whom they are reasons). For recall that on the non-instrumentalist's story, reasons need not be connected to any desire of the agent's (for whom they are reasons). Thus some of the reasons that the non-instrumentalist will be disposed to talk about (e.g. Andy's reason to pull the fire alarm) will be in need of grounding that the non-instrumentalist will be unable to offer; e.g. no desire of Andy could ground Andy's alleged reason for pulling the fire alarm.

At this point, the instrumentalist may press that her opponent's view, albeit extensionally adequate, it is nevertheless ungrounded in virtue of being unable to fill in the crucial placeholder D that is thought to ground the reason-relation. But the non-instrumentalist, I submit, has no good reason to accept the proposed distinction between facts about desires and other natural facts, in particular the thought that only the former can do the grounding for reasons. As I will now argue, the distinction is highly unwarranted: one without a difference.

4.7.2 The Non-Instrumentalist's Response

[61] A notable advocate of the distinction is Schroeder (2007), who distinguishes between reasons and *background conditions.* The latter concerns 'what must be cited in a complete explanation of why [a reason-giving feature] is a reason' (34).

[62] Although note that in the case of neo-Humeans, this grounding is indirect. That is, the reason is directly grounded in facts about what the idealised advisor would advice, which is in turn grounded in facts about the agent's present desires.

On my view, the non-instrumentalist has two options. First, to reject the distinction between reason-giving facts and reason-grounding facts. To maintain, that is, that all relevant facts about reasons must be placed in the same placeholder, F, and that the reason-relation is either 'self-grounded' or not grounded: it is either in no need of a grounding fact for it to ascribe an action Φ to an agent A in circumstances C and in virtue of a fact F, or it has no grounding to make such claim. Rejecting *this* distinction, I believe, is an unattractive view which the non-instrumentalist must not accept. This is so because the distinction between a reason-giving fact and a reason-grounding fact is necessary to make sense of familiar ways of thinking about reasons. Consider for example the claim that Andy has a reason to pull the fire alarm because doing so would save the lives of others. Suppose that on his way out of the building, Andy faints and that he is now unable to pull the fire alarm. Does it make sense to say that Andy still has a reason to pull the fire alarm? Perhaps it does make sense to say that the fact that pulling the fire alarm would (or could) save the lives of others *would have been* a reason for Andy to pull the fire alarm *had he been able* to do so. At the same time, to ascribe this reason to Andy seems a misuse of language. This can be explained, it seems, via the distinction between reason-giving features and reason-grounding facts. We can say: there *would have been* a reason for Andy to pull the fire alarm (the fact that doing so would save the lives of others) but Andy does not have that reason in virtue of not being able to do so. In short: there *would have been* a reason for Andy to Φ, *had that reason been grounded*.

The non-instrumentalist must go for a second option, namely to object to the idea that facts about desires and only such facts can do the grounding for reasons. In doing so, she must provide an alternative account of what grounds reasons. This will enable her to hold on to the intuitive thought that there is some fact that grounds reasons: explains somehow why facts about reasons obtain. But as I will explain shortly, there are certain limitations in giving such an account. Before I give the non-instrumentalist account that I favour, I shall consider what seems to motivate the instrumentalist's own account.

The idea's initial appeal

The idea that facts about desires are exclusively what grounds reasons, may seem appealing for two reasons. First, we often treat desires (or the presence thereof) as generating or annihilating the existence of reasons for agents to do things. Consider the following reason-claim: The fact that there will be dancing at the party is a reason for Ronnie to go to the party (Schroeder 2007: 29). Intuitively, and we may all agree on this, Ronnie's reason is at least partly explained by his desire for dancing. In Ronnie's case, were his desire for dancing to go away, the reason would go away with it. Thus it is intuitive to think that Ronnie's reason stands or fall — so to speak — with his desire. In a metaphysically significant sense, then, it is tempting to think that Ronnie's desire for dancing *grounds* Ronnie's reason to go to the party.

Second, we often refer to some fact that provides (or is) the reason (in Ronnie's case, the fact that there will be dancing at the party) without even mentioning Ronnie's desire for dancing. In the above example, Ronnie's desire for dancing is not mentioned as part of his reason for going to the party, though it is clearly assumed: it operates on the background (cf. Schroeder 2007: 30). In other words, we seem to take it for granted that the desire explains (or grounds) the reason that we call attention to (the fact that

there will be dancing at the party), as if we assume that a desire is a prerequisite for there being a reason in the first place.

What the non-instrumentalist must accept/deny

The non-instrumentalist must begin by accepting the intuitive thought that Ronnie's reason (to go to the party) does stand or fall with Ronnie's desire (for dancing). They must also accept that we sometimes fail to make reference to a desire that we think grounds a reason. What they must resist, however, is — first — the thought that all reason-claims are *exactly* like that of Ronnie's: that the instrumentalist's proposed explanation of Ronnie's reason which treats a claim about a desire as a grounding claim, generalises to all reason-claims. And second, the thought that desires are the only items that we sometimes fail to make reference to despite taking them to be an important part of what explains (or grounds) a reason.

Relating to the first point above, the question that the non-instrumentalist must press is: Why should we understand facts about desires as the *distinct* class of natural facts that can ground facts about reasons, and not any other types of facts, such as *modal facts*: (say) the fact that pulling the fire alarm (in order to save the lives of others) is something that Andy (and not, say, Ronnie) is able to do? The modal fact that Andy (and not Ronnie) is able to pull the fire alarm, it seems, may be able to explain why Andy (and not Ronnie) has a reason to pull the fire alarm — just as the fact that Ronnie (and not Andy) likes dancing explains the fact that Ronnie (and not Andy) has a reason to go to the party. In short: just as the facts about desires can generate or annihilate the existence of reasons, so do modal facts about what agents can and what they cannot do given the circumstances. What is so special about desires?

Relating to the second point, the non-instrumentalist must observe that facts about desires are not the only kinds of facts that we sometimes suppress when giving an explanation of a reason. We also suppress modal facts, more often than we do facts about reasons. In Ronnie's case, we seem to take it for granted that Ronnie can go to the party, and in Andy's case, we seem to take it for granted that Andy can pull the fire alarm. Were each of those modal facts to go away, the respective reasons would go away with it.

The non-instrumentalist can dispute the special importance that instrumentalists attribute to facts about desires when it comes to giving explanations of reasons by making a further observation. This is effectively anticipating a potential source of support for the insistence that facts about desires are normatively more relevant than other kinds of facts. The idea may be that there is a certain affinity between facts about reasons *for* an agent and facts about desires *of* an agent. This may impel some to think that desires are uniquely equipped to play the role of grounding reasons: in virtue of being essentially relational facts about the agent. For plausibly enough, the fact that something is a reason is always strictly speaking a relational fact since a reason must be a reason *for* some agent or agents (Olson 2009: 176). But this need not be so, given that other facts (i.e. modal facts) are relational too: e.g. the fact that Andy can pull the fire alarm is one about Andy (and not Ronnie), and so explains why Andy (and not Ronnie) has the reason that we think he does.

The non-instrumentalist's metaphysics of reasons

The non-instrumentalist has disputed the special status that her rival assigns to facts about desires in explanations of reasons: that facts about desires and only facts about desires can ground a reason-relation. The former's case is based on a legitimate question that remains unanswered: what makes facts about desires more normatively relevant than other facts: notably, modal facts? At this point, the non-instrumentalist must accept the schema proposed by her rival representing the metaphysics of reasons (R[F, A, Φ, C]), but propose instead that explanations such that of Ronnie's, in which desires do the grounding, do not overgeneralise. There are cases, the non-instrumentalist must say, where desires do not ground reason-relations. Rather, what really grounds reason-relations are relational modal facts. And these facts, she must continue, are present in every reason-ascription. The presence of a modal fact is what *unifies* all reason-claims.

The last observation above enables the non-instrumentalist to satisfy a desideratum that typically motivates instrumentalist theories: that of providing a unified explanatory theory of what makes things reasons (Schroeder 2007: 198). The idea is that on a metaphysical level, a theory of reasons must give an account of that which all reason-relations have in common. The proposal that all reasons are grounded on desires (of the agent's for whom they are reasons) may be unfortunate (for reasons given in 4.4-4.6), but at least tries to fulfill this aspiration. The non-instrumentalist, by contrast, proposes a theory which is both fortunate enough as to deliver the right extensional results, and point to what unifies all reason-relations: the fact or the supposition that the agent (for whom something is a reason) can act on her reason. Indeed, this is a feature that all reasons seem to have in common. Recall Ronnie's and Andy's cases. Ronnie's reason to go to the party may be grounded on the fact that he can go to the party, just as Andy's reason to pull the fire alarm may be grounded on the fact that he can do so. Absent those facts, both reasons go away — as we may put it, they both lose their grounding.

Necessary and Sufficient conditions

The instrumentalist may press at this point that whereas modal facts are necessary for grounding a reason, facts about desires are advocated by her as both necessary and *sufficient* conditions (more on this shortly). Given that her account does more explanatory work than her rivals, this may mean that it is also superior. And in case her rival cannot provide an account which specifies also sufficient conditions for when a reason obtains, then instrumentalism must be deemed superior than non-instrumentalist with respect to *metaphysical grounding*.

Indeed, we must note at this point that the non-instrumentalist's account of grounding only specifies a necessary condition for reasons — what all reasons have in common and without which no reason obtains: the modal fact that the agent (for whom something is a reason *can* perform the action in question). Were the non-instrumentalist to advocate this condition as necessary *and sufficient*, that would generate a kind of too many reasons objection: agents would have reasons to perform *any* action that they are able to, and of course that would be absurd. By contrast, the instrumentalist's account not only gives us a condition which — on her view — is necessary for a reason to obtain (i.e. a desire of the agent), but it also purports to advance this condition as a sufficient one: one under which a reason *certainly* obtains.

In giving us a sufficient condition for when a reason obtains, the instrumentalist's account can be used to *predict* when agents have reasons; in case we are wondering whether an agent has a reason to perform any action, all we need to examine is whether the agent has some desire the object of which is served by that action. If there is some such desire, then the agent has at least a pro tanto reason for performing the action. Relatedly, the account also enables us to provide an explanation of why an agent has a reason to perform one action X, and not another Y, which is more informative than her rival's. To see this, consider the question 'Why do I have a reason to pull the fire alarm, and not a reason to refrain from doing so?'. The non-instrumentalist may say: because doing it would save the lives of others, whereas not doing it would lead to their death. Consider now her answer to this question: Why does the former, and not the latter, fact provide me with a reason? At this point, the non-instrumentalist seems to have run out of explanations, as there is nothing more than she could say: what she is missing is a sufficient condition of when a reason obtains, which generalises to all cases and explains each time why a reason obtains. The instrumentalist, by contrast, who would only be inclined to say that an agent has a reason for performing an action only when there is a desire of the agent that is served by doing so, could point to the presence of a desire in favouring one reason over another. For example, she could say, Andy does not have a reason to pull the fire alarm, even if doing so would save the lives of others, simply because he does not have any desire that is served by doing so.

We thus arrive at a challenge proposed by the instrumentalist: to give a theory of grounding which also specifies the sufficient conditions under which a reason obtains. Though I appreciate the value of such an account, as it would enable us to *predict* when a reason obtains, it is nevertheless not clear to me why a theory of *grounding* should specify such conditions. As I see it, what is really required from theory of grounding is that it must be able to take each and every case where a reason obtains, and explain in virtue of what in the world this is so. And the non-instrumentalist's story can do just that. For example, the questioner begins by asking: Why does Andy have a reason to pull the fire alarm? The non-instrumentalist's answer is: Because doing so would save the lives of others. This is the reason-giving fact *F*. The questioner may then proceeds by asking: But why is this fact a reason for Andy to pull the fire alarm? The non-instrumentalist's answer would then be: Because Andy *can* save the lives of others by pulling the fire alarm. She may also say: Had he been unable to do so, *F* would not have been a reason for Andy to pull the fire alarm. In short: Andy's *being able* to pull the fire alarm is what grounds his reason.

Of course, the questioner may also ask: But why is the combination of *those* two facts (F and the modal fact) reason-giving for Andy, and not the combination of any other two facts? For example, the fact F* that by not pulling the fire alarm would lead to the death of others, and the fact that Andy can bring this about, may be thought to ground a reason for Andy to not pull the fire alarm. At this point, the non-instrumentalist must realise that the question may not be a good one. First, it seems that explanations must stop somewhere, and her own explanation of Andy's reason does two things to suggest that it stops at a pretty good place: it points to a a fact *F* that *intuitively* seems reason-giving, and to a modal fact that affirms F's status as reason-giving. What else do we need from a theory of grounding? Second, we may realise that the instrumentalist's own account must also stop somewhere. For suppose we ask

her: But why is the presence of *these* two facts reason-giving for Andy: the fact that by not pulling the fire alarm would lead to the death of others, and the fact that Andy wants to bring about their death? If the instrumentalist's answer is that facts about desires ground reasons, we may then ask: But why do facts about desire ground reasons? At this point, the instrumentalist may point out that 'they simply do': explanations must stop somewhere.

I conclude that in some sense, the non-instrumentalist can provide grounding for reasons. It may not be a *complete* grounding in that it does not specify sufficient conditions about when a reason obtains, but it is a *necessary* grounding insofar as it provides a condition without which a reason would not be there. I do not think the non-instrumentalist can provide a complete grounding in the sense specified — and neither do I think anyone else can. Instrumentalists attempt to do that, but the theory that follows from their account of grounding is incompatible with the extensionally right results about reasons. I take this to be an indication that a theory of grounding for reasons need not involve sufficient conditions, as the presence of the latter seems to inevitably lead to overgeneralisations about when a reason obtains.

4.8 The Instrumentalist's Mistake: A Diagnosis

In 4.6, we saw that instrumentalism fares worse than its rival with respect to the first criterion on a theory of reason, and in 4.7, I argued that contrary to the claims of some instrumentalists, it is far from clear how their theory fares better with respect to the second. As a result, I am far from convinced that instrumentalism, the theory upon which the error theorist's contention that there are no moral reasons is based, is true. Much to the contrary, the above dialectic impels me to think that instrumentalism is false. As I explain below, this is consistent with holding that the theory gets *something* right. In my view, instrumentalism may be plausible as a theory of *rationality in a certain sense*, though it is false as a theory that understands reasons as being necessarily connected to the agent's (for whom they are reasons) ends. In short, my diagnosis is going to be that instrumentalism centres on a narrow conception of rationality, thereby excluding perfectly good reasons that may be associated with the broader conception of rationality which instrumentalism overlooks. This section is not supposed to put forward an argument against instrumentalism, but rather a diagnosis of its errors which is intended only as supplementary to the considerations of the previous section against instrumentalism.

4.8.1 Rationality: Narrow vs. Broad[63]

We must start by noticing that the term 'rational' (and 'rationality') is polysemous: it has more than one meanings — it may be used to express more than one concept.[64] And though I have thus far been using the term as if there is a single, clear sense attached to it — this does not seem to be the case.[65] As Garner notes, in one sense, 'rationality' is often identified with the calculating selfishness:

---

[63] There are of course other non-practical conceptions of rationality (e.g. relating to what is rational to believe). I shall here speak only of uses of the term relating to the practical domain.
[64] There may be more than two concepts that the term picks out, but I shall here restrict myself to discussing two crucial ones for our purposes.
[65] This is an observation also made by Joyce in (2001: 66).

> One is rational when one's choices are directed to maximising one's own interests, or the interests of a narrow circle of friends and relatives (Garner 1994: 67).

Clearly, something like the above conception of 'rationality' is what lies at the heart of the instrumentalist dogma. But this is not to say that the above description of rationality does perfect justice to the neo-Humean instrumentalist's view. In particular, the neo-Humean has the resources to deem certain (so called) *other-regarding* concerns of the agent as rational — despite the fact that in caring for others the agent need not be acting selfishly in the standard sense of the word. When discussing instrumentalism (under the label of *proceduralism*), Hooker and Streumer explain:

> If Laura has greater natural benevolent concern than Emily does, then it can be rational for Laura to make greater sacrifices for the benefit of others than it would be rational for Emily to make (Hooker & Streumer 2003: 7).

The important thing here is that, were Laura to lose her natural benevolent concern for others, the instrumentalist would be naturally inclined to withdraw the judgement that in benefiting others Laura is being rational. For the instrumentalist who makes the judgement that Laura's actions are rational is employing a particular conception of rationality — which is closely associated with the agent's ability to further her own interests and concerns. Thus even though the current conception of rationality — which I contend lies at the heart of instrumentalism — is not strictly related to selfishness, it is still meant to appraise only those actions that further the agent's own concerns: be they selfish or not.

Call something like the above conception of rationality — the first of the two to be called attention to here — *rationality in the narrow sense*.[66] This is the conception that underwrites and renders intelligible a large number of ordinary statements featuring the word 'rational'. Recall the earlier example of Andy, who is exiting a burning building.

It seems that, *in a certain sense,* Andy is acting rationally; for he is doing perfect justice to his own goals, giving away as little energy as possible which may be used for the advancement of his own ends. This is the *narrow* sense of rationality, and it is precisely in regard to this sense that someone who designs to the last detail (and executes with equal rigour) an intricate plan to murder those and only those citizens who have logged out of their e-mail accounts at precisely 12pm noon, may be regarded as rational. As it transpires, this narrow sense of rationality (apart from its relation to furthering one's ends) relates also to qualities such as the ability for accurate calculations, foreseeing the long-term consequences of one's actions, possibly among

[66] The account of rationality in the narrow sense sketched here comes close to what Scanlon refers to by the same name in (Scanlon 1998: 25). But my account is not identical to Scanlon's, insofar as he allows that irrationality in the narrow sense 'occurs when a person recognises something as a reason but fails to be affected by it' (ibid). On my account, such an individual may very well be rational in the narrow sense, insofar as he takes the optimal means to his ends: i.e. whether or not he is able to recognise (or appreciate) good practical reasons.

other things, in that someone who is good at exercising it scores relatively high with respect to the latter. Recall the previous example of the poker player: this is certainly the case of someone who can be regarded as failing to be rational in this narrow sense; she fails to act in accordance with her own ultimate desires, or to do what is in her own best interest.

The above notion of rationality may even also be the *standard* one: in the sense that it is most often used by philosophers and non-philosophers alike. But it certainly is not — I submit — the only sense. For at the very same time that we may be inclined to think of Andy as rational (in the above specified sense), we may well also think of him as utterly *insane*. This brings us to the second sense of rationality to which I wish to draw attention here: rationality in the *broad* sense. Unlike its narrow counterpart, this is not the sense strictly associated with one's ability (or charisma, if you like) to see and take the optimal means to one's own goals. It is rather the conception associated with one's ability (or charisma) to 'see' and follow good practical reasons: be it instrumental or non-instrumental.

4.8.2 The Broad Conception of Rationality Is that of Reasons

It is rationality in the narrow sense that Broome seems to have in mind when he rightly observes that:

> Just because rationality is conceptually connected with the faculty of reason, it does not follow that it is conceptually connected with reasons (Broome 2007: 166).

Broome's point brings us to my second observation: the rationality in the narrow sense — the one instrumentalism tells us about — is not the rationality of reasons.[67] That is, it is not part of our concept of rationality in the narrow sense that: whenever one rationally $\text{ought}_{\text{in the narrow sense}}$ to Φ, one has necessarily a reason to Φ. But it is precisely our concept of rationality in the broad sense that is associated with the rationality of reasons. This is the conception of rationality that Kolodny seems to have in mind when he rightly observes that:

> There is a common use of rational — we might call it the 'objective' use — in which the phrase 'the rational thing for one to do' simply means … what one has most reason to do (Kolodny 2005: 510).

The above observations are perfectly consistent with the fact that in its generic sense, rationality is intimately connected with the notion of justification. To be rational *in some sense*, it is to be justified in *that* sense. Thus if I am rational in the narrow sense, what this means is that my actions are justified by (or in relation to) rationality in the narrow sense. By contrast, to be rational in the broad sense, it is to be justified

[67] In saying this, I seem to be in agreement with Kolodny: '[T]he normativity of rationality is not straightforwardly the normativity of reasons' (2005: 513). It is clear by the rest of what he says in the article, that by 'rationality' he means something like the narrow conception specified here.

— *period*. Rationality in the broad sense, I submit, is the rationality of reasons.[68] It is the conception of rationality associated with the notion of a practical ought as such presented in 1.3.2. This is why it is fitting to regard the case of someone who is rational in this broad sense as (to use Scanlon's phrase) *ideally* rational (Scanlon 1998: 30).

4.8.3 Making Use of our Resources

As an account of the one conception of rationality (rationality in the narrow sense), instrumentalism may well be true. But as an account of the other (rationality in the broad sense), it is not. The two conceptions of rationality that I have called attention to are equally coherent and important; they are coherent in that both can be easily accounted for, and they are important in that they both play a crucial role in understanding many of the things we ordinarily say.

The above observations enable us to identify the mistake in the instrumentalist's reasoning. Even though moral failure is not rational failure in the narrow sense, it may well be rational failure in the broad sense.[69] In terms of reasons, even though moral reasons cannot be made compatible with the requirements of instrumental rationality (or rationality in the narrow sense), they may well be made compatible with the requirements of rationality in the broad sense. Alternatively, we may put the point this way: the instrumentalist[70] starts by noting a certain sense of rationality: rationality in the narrow sense. She then consults what may be regarded as a basic common intuition, namely that what is rational for one to do may well be what one has reason to do. She then arrives at the following conviction:

> Reasons and rationality go hand-in-hand (also, see 6.3 for a similar line of thought).

The mistake that the instrumentalist makes, however, is concentrating on a narrow conception of rationality — thereby neglecting its broad counterpart. And though it is the latter conception that is most intuitively associated with practical reasons, the instrumentalist — a victim of her own negligence — attaches reasons to the former. Were instrumentalists to recognise and appreciate the broad conception of rationality, they should have no difficulty accepting the possibility of non-instrumental reasons. It is thus tempting to accuse the instrumentalist of being narrow-minded in a rather technical sense: that of focusing solely on one area of rationality — neglecting the rest.

---

[68] Matthew Bedke is another philosopher who seems to recognise this, when he does not 'deny that rationality is sometimes used in a broader, substantive sense that aligns with good reasons' (2008: 97).

[69] Joyce notes that: 'the moral rationalist's position boils down to a simple and surprising claim: that moral failure is rational failure' (2001: 91). The claim is indeed surprising if the relevant notion of rational failure is here taken to correspond to rationality in the narrow sense.

[70] It must be noted that not all instrumentalists are susceptible to this diagnosis. Schroeder (2007) proposes a version of the neo-Humean variety that departs from popular trends. But as I have argued in 4.6.1, Schroeder's version is based on a rosy view about the possibility of desire-based reasons that every agent (for whom they are reasons) has, and which is in dire need of support.

4.9 Explaining Away the Apparent Appeal of Instrumentalism

I shall close this chapter by addressing a potentially unsettling realisation for any opponent of instrumentalism. This relates to the fact that, outside philosophy, instrumentalism becomes increasingly popular. As Hubin accurately observes:[71]

> Humeanism, it is fair to say, is the theory to beat; perhaps it is even accurate to think of it as the default position … [some form of it] seems to underlie the theories of rationality that predominate in the social and behavioural sciences as well as those employed by decision— and game-theorists (1999: 30).

Indeed, it may also seem reasonable to think that any attempt to show that instrumentalism is false must also explain *away* the apparent appeal that the theory has to such theorists. I shall accept the challenge and concede that instrumentalism is indeed popular within and outside philosophical circles. But I shall press that it is not (at least not primarily) popular as a theory of good practical *reasons*. Rather, it is first and foremost popular as a theory of rationality *in the narrow sense*. Economists, for instance, have as their focus and are primarily interested in what Blackburn calls *the economic man*: a creature who always acts with his or her own interest in mind (1998: 134). It would thus be only natural that instrumentalism, a theory that understands reasons as favouring actions that lead to the furtherance of the agent's interests, was appealing to economists. But instrumentalism's success as a theory of rationality in the narrow sense must not be conflated with its failure as a theory of rationality in the broad sense and, by extension, as a theory of practical reasons. Insofar as this is left concealed, instrumentalism can become a dangerous theory, one that threatens to convince philosophers and non-philosophers alike that (say) the inconsiderate person exiting a burning building has neither a reason to pull the fire alarm — nor could he be in any serious sense susceptible to rational criticism for failing to do so. By drawing the crucial — and highly intuitive — distinction between rationality in the broad sense and rationality in the narrow sense, we are in a strong position to expose the importance of the mistake committed by anyone who thinks either that the narrow conception of rationality is the most basic one or (let alone) the only one there is.

Now, of course, economists and game theorists may well continue to be solely interested in the notion of rationality in the narrow sense — the one that instrumentalism is all about. For part of those theorists' objective is to account for expected human behaviour, and humans have a natural and reliable tendency to operate in accordance with their own interests — i.e. in ways that tend to be justified by rationality in the narrow sense. But just as concentrating on that notion is not the economist's fault, it is nevertheless the philosopher's fault to fail to properly disambiguate between rationality in the broad sense and its reasons, on the one hand, and rationality in the narrow sense, on the other.

4.10 Conclusion

[71] He makes clear that by 'humeanism' he refers to theories that are true to the spirit of instrumentalism: 'that reasons are based, ultimately, on subjective, contingent, conative states of the agent' (ibid).

At the beginning of this chapter, instrumentalism was identified as a possible source of support for the error theorist's contention that there are no moral reasons — and I motivated the view that the latter pretty much stands or falls with the former. Given this, I turned my attention to considering whether instrumentalism is true — or at least the most plausible theory of reasons. To this end, I identified non-instrumentalism as the theory's main rival. Following the consideration of how well both theories fare with respect to the two criteria on a theory of reasons, I concluded that we have little reason to think that instrumentalism is true. I then offered a diagnosis of the instrumentalist's mistake — one that puts forward an analysis of the ordinary concepts of *rationality* and *reasons*, including an alleged link between them. What all this means is that the moral error theorist's contention remains unsupported. In the next three chapters, we look for independent arguments in favour of instrumentalism and, by extension, the error theorist's contention.

5. Reasons Internalism

5.1 What Is Internalism, and Why Are We Interested?

Do moral norms[72] invariably supply agents with sufficient normative reasons? Does the dedicated immoralist necessarily have adequate reason to act morally? For there to be sufficient normative reasons invariably supplied by moral norms, *moral reasons*, there must be at least some reasons that are non-instrumental: reasons that do not necessarily favour any of the agent's (for whom they are reasons) ends. We know this because we have a sound understanding of the content of moral requirements: they need not prescribe actions that are fulfilling to the agent's (to whom they apply) ends (see 2.4). If all reasons are instrumental, then it is possible that a moral norm applies to an agent without it being the case that the agent has reason to act morally. In this context, the answer to the initial question turns on the plausibility of practical instrumentalism: the theory on which all normative reasons (pro tanto or normative) are instrumental. If instrumentalism is true, there are no reasons invariably supplied by moral norms. The previous chapter has considered the initial plausibility of instrumentalism, and it was far from clear why we should prefer the theory over its rival, non-instrumentalism: the theory that normative reasons need not be instrumental.

Note that the thesis question is formulated in terms of sufficient normative reasons, but we shall be here interested in reasons unequivocally: pro tanto or sufficient. As I explain in *1.1.2*, in being *normative*, a sufficient normative reason contrasts with motivating and explanatory reasons: a reason *for which* someone does something, and a reason *why* someone does something, respectively. A normative reason for action is a reason there *is* for doing something: a consideration that justifies or supports a prospective course of action. A sufficient normative reason is one that outweighs all other relevant considerations that are at play, thereby determining what the justified course of action is: the one that is most reason supported. In being sufficient, it contrasts with *pro tanto* normative reasons: a consideration that justifies an action but which may be overruled by better ones. If all reasons, pro tanto or sufficient, are instrumental, then the answer to the thesis question is negative.

In this chapter we look at *reasons internalism*: a family of arguments all designed to lend support to instrumentalism. Internalism specifies a criterion or restriction on what could count as a normative reason for action (pro tanto or sufficient): something is a reason only if it can potentially motivate the agent (for whom it is a reason) (Joyce 2001: 108). If internalism is true, then there are only reasons for agents that could potentially motivate them. This gives us good reason to think two things: first, that instrumentalism is true, and that there are no moral reasons: reasons invariably supplied by moral norms. To see the first, consider this question: what could motivate one to do something if not one of her ends? To see the second, consider the possibility of the dedicated immoralist who remains entirely unmoved in the face of a moral norm that applies to him: he remains unmotivated by the supposed reason that he has, thus on internalism's view, he does not really have a reason to act morally. Internalism may thus be understood as a premise in an argument for

[72] In this thesis, I assume throughout that the moral norms that I am concerned with express moral obligations or requirements (Singer 1972).

instrumentalism (Schroeder 2007: 7). Indeed, I shall here grant to my opponent that if internalism comes out true, then instrumentalism is true and there are no moral reasons (in the sense already specified). The dialectic of this chapter must wait, as it would be helpful to gain a good understaning of internalism first.

5.2 Introduction to Internalism

5.2.1 Motivation, Deliberation and Deliberative Routes

I like beer, but I am not always motivated to drink it. Once it hits my lips, I want more; in that moment, I am highly motivated to drink it. But the more I think about the negative effects it can have on my body and mental health, the less motivated I become to drink it. In particular, during daytime and while I am sitting with friends having lunch, I am mostly *un*motivated to drink beer, whereas during night-time and when I am in the pub watching the Merseyside derby, I am mostly motivated to drink it. Examples can be multiplied; having finished exercising, I feel like talking about philosophy, whereas when I have been sitting on the couch for hours and in an awkward bodily position, I do not feel like doing anything. In general, my motivation towards any particular action seems to depend on certain states that I find myself into.

The active and spontaneous types may always do what they feel like doing — that is, without conscious or careful thought — but I occasionally *think* about what to do. We may call the process of thinking about what to do — which may involve weighing relevant considerations — *deliberation*. Suppose while drinking beer at the pub, I come to think that I should stop, and following this, I do so. My doing so — we may say — is the result of my deliberation.

Moreover, we may call the particular state in which one finds oneself deliberating, a *deliberative route*. In the previous example, I find myself in the state of being a little tipsy, having most relevant beliefs true, possibly among other things. We may add the state of thinking about the negative effects of beer drinking as part of my deliberative route — as it can either decisively shape the outcome of my deliberation or alter what I end up doing. At other times, mostly during daytime, I find myself in the state of being sober and focused, and having most relevant beliefs true. This time, my thinking about the negative effects of beer drinking may be more influential to either the outcome of my deliberation or what I end up doing.

5.2.2 Reasons According to Internalists

Deliberating often involves choosing between conflicting courses of action. Whenever that happens, we may cite *reasons* in favour of the one or the other action. Consider Mary and George who argue on whether they should attend their favourite team's game tonight or watch it instead on television. George mentions the fact that going to the stadium will be more pleasurable because of the atmosphere created by the fans as a reason in favour of going, while Mary mentions the fact that the tickets are expensive as a reason in favour of watching it on television. Presumably, George likes the atmosphere created in such events more than he dislikes paying large amounts for a ticket, while Mary dislikes paying large amounts for a ticket more than she likes the stadium atmosphere. In this example, both Mary and George seem to take their pull

towards each action as evidence that they have reasons to act in certain ways. Internalists think that *all* reasons — with no exception — are like that: *related* to the agent's (for whom they are reasons) *motivation*. In particular, internalists adhere to the following idea:

> Whenever an agent has a reason to Φ, that reason bears some *relation* (R) to the agent's *motivation* (M) towards Φing.

Within the internalist camp, internalists disagree on how they understand the relevant *relation* and the agent's *motivation*. In particular, the disagreement is between two basic ways of understanding each. Depending on the particular conception, and the particular combination of the two, we get *types* of internalism. Beginning with *motivation*, the two alternative conceptions are:

> *(M1*) in terms of the agent being motivated towards an action.
> (*M2*) in terms of the agent having a motivational attitude towards an action.

Internalists who accept M1 will say that an agent can have a reason to Φ if the agent has at the very least an actual intention towards Φing. Internalists who accept M2 will say that an agent can have a reason towards Φing if and only if the agent has at the very least a motivational attitude (e.g. desire, emotion, liking, admiration etc) towards Φing. M2 is weaker than M1 in that it does not require that the agent have any actual intention, only an attitude which has the potential of giving rise to an intention towards the action for which the agent can be said to have a reason (Sinclair 2017: 210).

Likewise, there are two prominent ways that the relation *R* between the reason and the agent's (for whom it is a reason) motivation *M* can be cashed out:

> (*R1*) in terms of *actual* motivation
> (*R2*) in terms of *counterfactual* motivation

Internalists who accept R1 will say that an agent can have a reason to Φ only if and only if the agent either *actually* is motivated to Φ, or *actually* has a motivational attitude towards Φing. We may call the former combination M1R1, and the latter M2R1. Internalists who accept R2 will say that an agent can have a reason to Φ if and only if the agent either *would be* motivated towards Φing, or *would obtain* a motivational attitude towards Φing, *under certain conditions*. We may call the former combination M1R2, and the latter M2R2. Note that R2-combinations are, other things being equal, weaker than their R1-counterparts in that they do not require that the agent actually be motivated or actually have a motivational attitude towards the action for which she has a reason.

The division between R1 and R2-type of internalisms is thought to be of great importance within the internalist camp. R1-Type of internalism is evidently more straightforward than its counterpart, as it does not require the presence of certain conditions. This so called *sub-Humean model* is considered by prominent internalists to be too simple a view to merit acceptance (Williams 1981: 102). Due to its simplicity, it is unable to rule out adverse conditions, such as not thinking straight or being under

the influence of alcohol — conditions which can assist the development of *irrational* desires, such as the drunk person's present desire to set his car on fire (Lillehammer 2000: 507). Internalism, the idea goes, must avoid tying the agent's reasons to desires or motivation formed under such adverse conditions. In order to do so, it must proceed by adopting a counterfactual model in which the motivation results *under conditions of a particular kind*.

Internalists typically fill in the counterfactual conditions in terms of a process of *rationally sound deliberation*. This is supposed to be a particular conception of a deliberative route — such as the state of being at the pub thinking about alcohol, only this time the proposed conditions are such that any resulting motivation in the agent would be indicative of the agent's reasons — or so these internalists think. Consider the following generic proposal:

> Necessarily, for all agents A, circumstances C, and actions Φ, A in C has a normative reason to Φ only if A in C can rationally access the state of being motivated to Φ from existing motivations (Sinclair 2017: 209).

To rationally access the state of being motivated to Φ is to either become motivated or come to have a motivational attitude towards Φing through the process (or under conditions) of a rationally sound deliberative route. 'Existing motivations' refers to what Williams calls the agent's *subjective motivational set* (Williams 1981: 102): desires, concerns, needs, aspirations, commitments and the like. Internalists who subscribe to a counterfactual model disagree over the particular conception of the rationally sound deliberative route. From Williams, we get the suggestion that an agent deliberates in a rationally sound way when the agent is fully reflective, imaginative and has all relevant beliefs true[73] (Williams 1981: 104-105). Note at this point that it is not important for our purposes to choose between competing conceptions of the sound deliberative route. Of course, different formulations may yield substantially different results (e.g. subtract the condition of having all relevant beliefs true, and the person who is now craving pizza is no longer motivated to go to the pizzeria because he believes it is closed). But our task here is not to examine the plausibility of any such proposal, rather it is to examine the plausibility of the internalist's *doctrine:*[74]

> Internalist's Doctrine: There is an interesting *relation* (R) between the agent's reasons, on the one hand, and the agent's motivation which results from a process of rationally sound deliberation and which is based on the agent's existing motivations, on the other.

Finally, it is noteworthy that on some conceptions of the relevant process of deliberation, the relation (R) between the agent's reasons and her existing motivations turns out to be too weak to be interesting. On those conceptions, the process of rational

[73] For a list of prominent such proposals see: Finlay, Stephen and Schroeder, Mark, 'Reasons for Action: Internal vs. External', *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/reasons-internal-external/>.

[74] I borrow the term from Lillehammer (2000), who considers the adoption of a counterfactual model as integral to the internalist dogma.

deliberation is defined such that it necessarily instils motivation in the agent to act in accordance with her reasons independently of her existing motivations. The upshot is that on such views the agent's reasons turn out to have a distinctly externalist flavour: there is no interesting connection between them and the agent's existing motivations (Lillehammer 2000: 512). Here, I shall be concerned with counterfactual models of internalism that define the process of rational deliberation such that it preserves an interesting connection between the agent's existing motivations and her reasons.

In what follows, we look at ways that internalists may wish to support their view. These take the forms of distinct — albeit related — arguments all in support of the view that all normative reasons are internal: potentially motivating to the agent for whom they are reasons.

5.3 The Argument from Alienation

The internalist begins by mounting a double charge against externalism: that the latter *alienates* the agent from her reasons by failing to restrict them according to the agent's own concerns, and that it *harasses* the agent by insisting that she act on reasons that leave her entirely unmotivated. Having made this alleged observation, the internalist proceeds as follows: On the assumption that (i) both aforementioned attributes are highly uncharacteristic of the true theory of practical reasons, and provided that (ii) internalism is devoid of both, we have good reason to think that internalism is true and externalism false. I begin by clarifying the double charge, before I question both (i) and (ii) — though my response to the argument hinges only on the rejection of (ii).

As we have seen, on the internalist's story an agent can have a reason for an action only if the agent can draw some motivation for that action after the process of rational deliberation. In the internalist’s story, it is impossible for the agent to have a reason to perform an action and not be motivated towards that action with the completion of the deliberative process. There thus seems to be a consonance between the internalist’s advice about reasons and the agent’s motivation. To see the consonance, consider the case of John who finds himself in a practical dilemma — torn between two possible courses of action: to save a drowning child that is in front of him (an action which will possibly ruin his suit), or go to the party to which he is heading. Let us also grant that John has all relevant true beliefs about the predicament in which he finds himself: that the child is a complete stranger to him, that the child is in life-threatening danger and that by failing to help it, the child will most likely drown. Moreover, and quite crucially, suppose that John pauses for a minute or so to think about what he ought to do, and manages to come up with *no* motivation whatever for saving the child. He thus feels compelled to ignore the child and head on to the party — and he even goes as far as to suppose that he has no reason to save the child, and every reason in the world to neglect it and go to the party; e.g. saving the child would ruin his suit and delay his arriving.

Now, we may acknowledge the fact that internalism — via its distinctive commitment — must concede that John has no reason to save the child and has every reason to go to the party without further ado. In an act of pure intellectual honesty, the internalist will advise John: ‘Despite the fact that you are a heartless egoist, you really have no reason for saving the child, so I should advise you not to’. And John would likely welcome the advice.

Unsurprisingly, things are quite different with respect to the externalist's advice to John. Recall that on the externalist's story an agent can have a reason for an action even if she remains unmotivated for that action after undergoing the process of rational deliberation. It becomes apparent that the externalist need not embrace John's supposition that he has no reason to save the child — even though she may well understand that John is not the kind of person that could be moved by the suffering and misfortunes of others. In fact, it seems increasingly likely that the externalist would respond to John that despite what he thinks and how he feels, he nevertheless has a reason to save the child. In doing so, the externalist would enter a disagreement with John about what *he* (as opposed to externalists, or others sympathetic to their view) has reason to do. Naturally, we would expect John to retort 'So what if *you* think I have a reason to save the child?'. Indeed, that alleged reason (i.e. the fact that helping the child could save it) is something that John does not identify with. In short: by insisting that John has a reason that he does not identify with, the externalist alienates John from his reasons.

Moreover, by insisting that John act on that reason, the externalist is effectively *harassing* John.[75] Insofar as the qualities of alienation and harassment thus construed are deemed unfitting for a true theory of practical reasons, the internalist may object that their rivals must be mistaken. In a similar context, Joyce claims:

> Any adequate theory of normative reasons must make out reasons to be precisely those things that forestall a 'So what?' response. Some theories of reasons threaten to violate this constraint – to 'alienate' an agent from his reasons (Joyce 2001: 81).

Given that (i) the qualities of being alienating and harassing are uncharacteristic of the true theory of reasons, and on the assumption that (ii) internalism is devoid of both whereas externalism is not, internalism must be more plausible than externalism. In responding to the internalist's argument, it must be noted that both assumptions (i) and (ii) can be disputed. Beginning with the latter, internalists can *also* be accused of alienating the agent's from what she regards as here reasons — at least on some occasions. Consider the case of a heroin addict, who despite wanting to get off the stuff once and for all, she *now* nevertheless has a strong desire for an injection. Granted, the agent has not yet undergone the deliberative process favoured by internalism, and hence she need not be motivated to act on her reasons — at least the ones that internalism recognises. Yet he presently identifies himself with the action of taking the injection, and upon acknowledging the internalist's insistence that she has a reason to refrain from it, he may well respond: 'So what if *you* think I have such reason?' As far as he can now see, he has no such reason. The upshot is that even internalism can occasionally alienate the agent from (what the agent takes to be) his reasons. Further, by insisting that the addict — on that present moment — act on the alleged reason, the internalist can also be harassing the agent. Just as this would not prompt internalists to regard their theory as false, why should it prompt externalism to dismiss theirs?[76]

---

[75] Others refer to this as 'browbeating' (Shafer-Landau 2003: 176).
[76] For a similar response, see (Shafer-Landau 2003: 177).

The internalist may now retort that there can be degrees of alienation, and that her theory is evidently *less* alienating than the externalist's. For internalism does not insist on reasons that the agent cannot be motivated for *given her existing motivations*. Even counterfactual versions of internalism, which insist on a process of deliberation that can provide a critical distance between the agent's present desires and her reasons, do link the agent's reasons with some instances of her motivation. Externalism, on the other hand, insists on reasons that the agent may well be unable to be motivated for — again, given her existing motivations. On this score, externalism may be considered *more* alienating than internalism.

In response, the externalist may sensibly wonder: What is it about a process of rational deliberation as such that makes any resulting motivation the agent's own? In other words, what is the inherent feature of the conditions of being fully reflective, imaginative and having all relevant beliefs true that make the resulting motivation the agent's own? My hunch is that — at least on some occasions — the relevant process may be just another *corrupting* activity in the sense that having undergone it, the agent ends up motivated to act in ways that do not represent her true identity. I am tempted to compare the process to *indoctrinating* or even *brainwashing*, both of which are meant to build an unquestioning loyalty in their subjects which would enable them to identify with (or be motivated towards) acting in certain ways. That said, it may be true that some people truly identify with what the indoctrination or brainwashing is about — but the important thing is that not everyone need do so. Similarly, with respect to the process of rational deliberation, it may be true that some people truly identify with the resulting motivation, though it need be true that everyone does. For example, consider the paradigmatic spontaneous individual who harbours a strong commitment to living life to the fullest and constantly at the edge: by thinking clearly about potential risks, our friend may end up doing the things that go directly against his nature as a devoted risk-taker. It seems that there is nothing inherent to the process of rational deliberation that enables the agent to do the things that she truly identifies with. Given that to identify truly with your motivation is what really counts towards non-alienation, internalism may be just as alienating as externalism (that insists on reasons that need have nothing to do with the agent's true identity).

It is not, however, the above criticism that I find particularly appealing as a response to the internalist's first argument. For I do not think that the qualities of being alienating and intimidating thus construed should count as a cost to a theory of practical reasons that has them. This brings me to the second response — this time *contra* (i).

The first thing to note here is that having to be alienated or harassed for your reasons may well be a negative quality *of the agent*. That is, an agent who is such that she cannot be motivated to act in accordance with her reasons is in some sense inferior to an agent who is such that she can. Conversely, not having to be alienated or harassed for your reasons may well be a positive quality of the agent. Mistaking a defeasible advantage of the agent for a defeasible advantage of a theory of reasons is the source of the current misconception — as I take the proposal to be.

Yet at the same time, to regard the above qualities as criteria (or restrictions) on a theory of reasons is clearly false. This may be due to a number of reasons, though I shall here mention two related ones. First, notice how the following two issues are

conceptually independent: a theory being alienating and harassing, and it telling the truth about one's reasons. John may be protesting that we are harassing him when we categorically insist that he get immediately out of the house and come join us in the street, and we may well be doing so — but that does not change the fact that his house is on fire and he is about to burn to death very soon. Upon reminding him of his predicament and telling him that he has every reason to rush downstairs, John may still retort 'So what?' — but it is still far from clear that our judgement about his reasons has gone astray.

Second, and quite relatedly, it seems to be a proper function of a normative reason that it can 'provide critical distance from agents' motivations (both as they are now, and as they would be after sound practical deliberation)' (Sinclair 2017: 214). Indeed, despite it being the case that not every reason-claim need be critical of the agent's present or more refined motivations, it seems highly plausible that at least some reason-claims must be. When one is unable to find any motivation to save a drowning child, even though the scene in front of his eyes deeply troubles him, considerations pertaining to reasons for action may emerge as potential liberators. By reflecting on the circumstances he finds himself in, facts of his situation and potential courses of action, an agent can eventually grasp the normative relation at play: that by helping the child he could save its life at minimal cost for himself. He could then examine his initial stance — that he had no reason to help the child — with a critical eye, and he can then appreciate the particular theory of reasons that insisted all along that he did have such a reason.

5.4 The Argument from the Explanation of Action

The second argument in favour of internalism centres on another alleged advantage of the theory over its rival: that it understands reasons as things that can both *explain action* and *be acted upon* by the agent for whom they are reasons. Williams is often credited with the current line of argument:

> If there are reasons for action, it must be that people sometimes act for those reasons, and if they do, their reasons must figure in some correct explanation of their action (Williams 1981: 102).

The idea is that reasons must be able to explain action, and that it is hence a requirement for any theory of reasons that it must make out reasons to be action-explanatory. It may not be clear, however, at least on first reading, how internalism gains an advantage over its rival in relation to this supposed criterion. Thus, there seems to be at least three questions that must concern us here: (i) What exactly does the explanatory requirement amount to? (ii) Is the explanatory requirement plausible when directed at a theory of reasons? (iii) Is it true that only internal reasons can fulfil it? I shall illustrate (i) by means of an example — and on the assumption that our answer to (ii) is affirmative, I will address (iii). Finally, I shall return to question (ii).

Imagine John who has not yet been corrupted by the tenets of either reasons internalism or externalism, and who, as a result, is unlikely to be influenced by anything those philosophers say. John is a mean-spirited person who has no sympathy whatever for earthly creatures, and has a particular dislike for cats. Occasionally, John

thinks before acting — though the process takes hardly more than a few seconds; when faced with a decision, John consults his relevant beliefs and allows — as it were — his own desires and concerns to work their magic from the background. By utilising this instantaneous and spontaneous process, John typically arrives at an action that is genuinely his: he always does what he wants or cares about. Suppose he is now on the verge of hitting a cat with his car while riding at full speed on the highway. He sees the cat, but does not slow down, and kills it.

In asking 'Why did John act as he did?' we are looking for an explanation of his action. Provided that there is one, i.e. that he did not act in an genuinely mechanical manner, and given what we know about him, one natural explanation of John's action would be this: John's awareness of the fact that by not slowing down or changing direction he will kill the cat, coupled with his desire to kill it, explains why he did not slow down or change direction. This is a possible explanation of his action, and it may or may not be the correct one. A reason (be it internal or external) is explanatory just in case it can feature in the correct explanation of the agent's action — whatever that may be. And if it is a requirement that a reason must feature in such explanations, then a theory is doomed just in case it allows for reasons that cannot feature in any such correct explanation. This answers (i).

Now, suppose that the explanatory requirement is plausible when directed at a theory of reasons — and that it can hence legitimately rule out a theory just in case the latter allows for reasons that cannot be explanatory. The question now is: Does internalism gain an advantage over its rival in relation to this supposed requirement? The answer, I believe, is *no*. To see why, let us revisit the example of John, and consider this time what each side — internalism and externalism — would have to say about John's reasons. Given everything we know about John and his circumstances, with emphasis on his atrocious attitude towards earthly creatures, internalists must say that John has a reason to keep driving in the same direction and at full speed. Externalists, on the other hand, may well beg to differ and say that John does have a reason to either slow down or change direction; for on their view, John's complete lack of motivation for slowing down or changing direction need not be telling as to whether he has a reason to do so. At this point, it may seem as though only the internalist's reason can feature in the correct explanation of John's action — and if this result is generalisable to other cases, then it seems that only internalism can fulfil the supposed criterion. But thinking this would be a mistake.

To see why, notice that the explanatory requirement is not that any given reason that a theory allows must be able to explain the agent's (for whom it is thought to be a reason) *actual* action. Rather, the requirement is that any given reason that a theory allows must be able to explain *some* action — in John's case, the alternative action of slowing down or changing direction. This seems to follow from Williams' own formulation: 'If there are reasons for action, it must be that people *sometimes* act for those reasons' (ibid, emphasis added). On that particular day, John chose not to act for his reason (i.e. to slow down), perhaps on another day, John will be more considerate of his reasons. To put it another way, there is nothing preventing the externalist from saying that even though John did not *actually* act for his reason (i.e. to slow down), he *could* nevertheless *have* done so. He could have done so had he (say) considered the suffering that he was about to cause to the cat. Had he done so, John would have been indeed motivated to slow down or change direction — and his action would have been

in alignment with his true reasons. This would have had the exquisite result of allowing us to explain his action by reference to nothing less than the normative reasons that he had. In short: John's external reason to slow down or change direction has at least the potential to being *explanatory* of John's action.

The internalist may insist at this point that the above interpretation of the proposed requirement is unacceptably weak. The internalist may instead choose to hold on to a stronger interpretation on which any given reason that a theory allows must be able to explain some action towards which the agent can be motivated *given her existing motivations*. Given that, in the previous example, John has no desire whatever that can motivate him to slow down, the internalist's argument goes, he cannot be said to have any external reason to do so. For an external reason is *ex hypothesi* not explanatory in the restricted sense we are now considering: that of being able to explain an action that can be justified by reference to the agent's existing motivations. And if we hold on to this stronger interpretation, this seems to be entirely generalisable: whenever an agent lacks motivation for a course of action, he cannot be said to have an external reason — or, for that matter, *any* reason — for that action. The upshot here is that, given this stronger interpretation of the explanatory requirement, the internalist's view is vindicated almost *ex hypothesi*.

In response, the externalist must reject her rival's proposed interpretation of the explanatory requirement — not just for the sake of her theory — but most importantly for the sake of maintaining here a unified discussion. For remember that the topic of discussion has been *normative* reasons: the *real deal*, considerations that actually do count in favour of an agent's action in certain circumstances (Bedke 2008: 94, cf. Shafer-Landau 2003: 178). Yet on the above interpretation the focus shifts to *motivating* reasons: considerations that can explain why an agent does things while remaining neutral as to whether the agent is justified in doing so (ibid: 93). Normative reasons have to do with *justification* — not necessarily *motivation*. A nazi may be motivated to promote the interests of the white race, but he need not be also justified in doing so — at any rate, the latter must be preserved as a conceptual possibility subject to independent argumentation. In introducing the proposed interpretation of 'reason', the internalist commits the fallacy of equivocation: equates motivating reasons with the ones (i.e. normative reasons) she initially set out herself to explain. In insisting that normative reasons just are motivating reasons, she is begging the question against her rival — for what we need here is not a restatement of her theory, but rather an independent argument in its favour.

We are thus left with a weak interpretation of the explanatory requirement — one on which a reason that a theory allows must be able to explain *some* action whether or not the agent can be actually motivated to pursue it given her psychological makeup. In closing this section, we may address (ii): Is the explanatory requirement (thus construed) plausible when directed at a theory of reasons? It seems to me that the answer is 'yes'. For to say that a normative reason cannot explain *any* possible action of the agent (for whom it is thought to be a reason) is on par with saying that it is *impossible* for the agent to ever become motivated for that action. And this seems highly uncharacteristic of what a normative reason is supposed to be. Remember that a normative reason is the kind of thing that *justifies* an action, as opposed to the kind of thing that *arbitrarily* favours one. Whether or not the agent can be motivated given her psychological makeup to Φ, it must be at least possible that she could be motivated

to Φ were she come to acknowledge the independent considerations that favour her Φing. Crucially for our purposes, however, on this weak interpretation of the requirement, internalism gains no advantage.

5.5 Two Direct Attacks at External Reasons

Building on the argument from explanation, Bernard Williams develops a much discussed argument against the plausibility of external reasons. Williams' argument is rather head-on, as it aims to argue directly that externalism is false — as opposed to stating a requirement that the theory may or may not fulfil. Williams' argument is often labelled as one about the *content* of external reason-claims (cf. Hooker & Streumer 2003), but I want to resist this label — at least for the time being. That is because I can see two distinct — albeit somewhat related — interpretations of William's thought, and it would be helpful to try to unpack them both into distinct arguments. The first argument that I can discern is about the *epistemology* of external reason-beliefs, and the second is about the *content* of the relevant propositions (cf. Finlay 2009). In what follows, I consider both.[77]

5.5.1 The Epistemology of External Reason Beliefs

The first argument we get from Williams attacks the epistemological prospects of external reason beliefs. It effectively asks the externalists to provide a rough account of the conditions under which an agent comes to believe that she has an external reason for action. Importantly, those conditions need better be compatible with the externalist's account, otherwise they may entail the truth of internalism. This is the relevant passage:

> What is it that one comes to believe when he comes to believe that there is reason for him to Φ, if it is not the proposition, or something that entails the proposition, *that if he deliberated rationally, he would be motivated to act appropriately*? (Williams 1981: 109, emphasis added).

The above passage both lays out a challenge for the externalist (underlined) and also gives a flavour of the internalist's own account of the epistemology of beliefs about internal reasons (italics). It is indeed easy to see how — on the internalist's story — an agent can come to believe that she has an internal reason for action. Suppose John (from the previous example) is on the roof of a building hiding, and he is about to throw eggs to some innocent by-passers just for the kick of it. He knows that the chance of negative repercussions of such action to him is minimal — if not non-existent, and so he is suitably impelled to do it. In fact, John's relative sophistication enables him to even entertain the thought that he has a reason for throwing eggs on this occasion. Suppose that internalism is true, and that John really has a reason for throwing eggs. The internalist's explanation of the belief-acquiring process is quite straightforward: John *deliberates rationally* by considering what his reflectively, imaginatively and knowledgeably idealised self would advice him to do on this occasion, and he concludes that he has a reason to throw eggs. John's idealised self

[77] Note that I will be operating on the assumption that Williams was aware of both dimensions of his argument, and so I may occasionally refer to both as *his* arguments.

believes that John has a reason to throw eggs (partly, but crucially) in virtue of his awareness of John's ever-present *desire* for hurting or humiliating others. Were John to somehow lose that desire completely, John's idealised self would have to reconsider (presumably jettison) his belief that John has a general reason for promoting the misfortune of others.

It may be unclear, however, how the externalist's explanation of the external reason belief-acquiring process would go. The internalist is now eager to exploit this apparent explanatory difficulty in order to show that externalists are committed to weird epistemological capacities — or at least a rational belief-forming processes. The internalist may also warn the externalist that not any explanation of such process will be acceptable. For an explanation on which the acquiring of the external reason belief turns out to be merely the result of non-rational persuasion would simply not do. Suppose that I tell John that he has no reason to throw eggs, and that in employing my persuasion skills, I manage to convince him so that he now mistakenly believes that he has no such reason. It is easy to see here what has happened: John comes to believe the external reason claim in a manner that is illegitimate by rational standards. For not only I have failed to call attention to independent features of his situation that make it true that he has an external reason to refrain from throwing eggs, but on the assumption that internalism is true, they may be no such features to begin with. In making him believe the external reason claim, I have engaged in *goading* him: merely provoked a cognitive response on his part, as opposed to *guiding* him: let the independent considerations (provided there are any) to speak for themselves (Falk 1953: 151). If she is to give an account of the belief-acquiring process, the externalist need better rule out goading-like explanations. Williams is rather explicit with respect to the aforementioned requirement:

> [On the externalist's account] the agent should acquire the motivation *because* he comes to believe the [external] reason statement, and … moreover, because, in some way, he is *considering the matter aright* (Williams 1981: 108, emphasis added).

So how does the externalist explain the relevant process? One further thing that the externalist cannot say — or at least on my view — is that an agent has an external reason because a supernatural deity says so. I shall rule out the possibility of any such mystical explanation. Yet the externalist may reasonably ground the relevant beliefs about external reasons on perfectly natural considerations, such the fact that by throwing eggs to innocent by-passers, John will elicit in them a feeling of humiliation. She may say: John fails to believe what he has reason to do (i.e. the external reason-statement) *because* he mishandles an important piece of information relating to his situation, namely the fact that by throwing eggs he will cause humiliation. Compare: I may have at my disposal the belief that outside is raining heavily, and at the same time fail to believe that the weather is currently atrocious. That the weather is currently atrocious is arguably a plain fact — i.e. we can make sense of it being in crucial qualitative respects different than the weather currently in south Greece. Just as in failing to handle the information about the weather properly, I am not thinking the matter aright, in failing to handle the information about the ramifications of his egg-throwing properly, John is not thinking matters aright either (cf. Gibbard 2003: 293-294).

A final thing that the externalist must stress is that even agents who are aware of their reasons may lack the motivation to act for their reasons. But again, in the externalist's account, motivation (or the absence of it) need not be indicative of one's reasons. For the internalist at this point to claim otherwise, would be question-begging against her rival, as opposed to giving an independent argument against her rival's view.

5.5.2 The Semantics of External Reason Statements

The second of William's arguments may be seen as a reaction to the externalist's account of the belief-acquiring process. The internalist might still be wondering what the content of such external reasons beliefs is — or what it could be. Granted, she may begin, external reason claims and beliefs may well have something to do with facts that are in some sense nothing to the agent yet still significant to other people — or even to most people; e.g. the pain of others may not mean anything to the inconsiderate individual, even though the pain of anyone may mean a lot to the empathetic Christian. At the same time, she may continue, there is still no sense to be made of the claim that *the inconsiderate* individual (as opposed to other people) *has a reason* to care about others. In this second argument, details of which are explained below, the internalist is challenging the *intelligibility* of external reason claims.

Indeed, here seems to be an interesting question: What exactly is the meaning of the claim that John is practically unjustified (i.e. has no reason to do what he intends to do) *if not* that he is 'inconsiderate, or cruel, or selfish, or imprudent' (ibid: 110)? For suppose that at least on some occasions (i) to claim that an agent is inconsiderate, cruel and the like *just is* to claim that he fails to take the suffering of others into account in deliberation. Suppose also that (ii) to claim that an agent is negligent of his external reason to (say) refraining from hurting others *just is* to claim that he fails to take the suffering of others into account in deliberation. In case both (i) and (ii) are true, it seems to follow that external reason claims (e.g. John has a reason to refraining from egg-throwing) are devoid of content. For there is nothing above and beyond the claim that the agent is inconsiderate, cruel and the like with which to underwrite the claim that *in addition to* being inconsiderate, cruel and the like, the agent is also negligent of her external reason to (say) refrain from egg-throwing. External reason talk may, for all we know, be empty rhetoric: bluff (Williams 1981: 111).

The internalist may also add: Of course things would be in some sense better, more preferable, more humane and the like were inconsiderate agents to change their ways, but this does not subtract from the fact that such agents do not have actual reason to change their ways. In insisting that they do *because they have reason to do so*, we are being deceitful. The internalist's aim, of course, need not be convincing us that being truthful is always better or practically more useful than being deceitful.[78] Rather, her view must be understood as simply the claim that despite it being in a practical sense bad advice, it is nevertheless *true* to hold that inconsiderate persons have no reason (at least for all we know) to change their inconsiderate ways.

[78] Thus the internalist need not be making a pragmatic claim similar to Garner's, according to which there are pragmatic benefits to being truthful (1994: 341-361).

Upon first reading, the internalist's argument may seem troubling for externalism — but externalists can respond. The externalist must begin by acknowledging the peculiar nature of terms such as inconsiderate, cruel, coward, brave, virtuous and the like. What these so-called *thick terms* have in common is their carrying evaluative connotations or meaning. To call an act cruel is not just to describe; it is also to condemn it. That is to say that thick terms have both a descriptive and an evaluative meaning. For example, *cruel* means: (descriptive) such as to inflict pain or suffering, (evaluative) such as to be in some sense negative or reprehensible. Thick terms stand in contrast to purely descriptive terms (e.g. table, chair) and purely evaluative terms (e.g. good, bad). This much should be obvious — but what may not be so obvious is how such observations can lend support to externalism.

The next step in the externalist's response is resist one claim and assert another. The claim the externalist must resist is the internalist's suggestion that thick evaluative terms such as *inconsiderate* and *cruel* exhaust the meaning of what is to be intelligibly said about (say) Attila the Hun *while remaining silent on whether such individuals have reason to change their ways*. In particular, it is the italicised part that the externalist must resist. The externalist must instead assert that it is an indispensable part of the meaning of such thick terms that such individuals are in negligence or violation of their practical reasons. For example, part of what it means to call an individual cruel is that in causing pain or suffering, she is acting against her practical reasons. In effect, the internalist's proposed analysis of thick concepts identifies the evaluative dimension of the relevant concepts (e.g. cruel) with the presence or the absence of a practical reason for the agent to whom the concept applies. If successful, the externalist's response accomplishes a vital task: it explains how claims about (say) cruelty render claims about external reasons superfluous without rendering them at the same time unintelligible. We can put it this way: the former can be seen as talking on the latter's behalf.

Some effort must be taken to motivate the externalist's response. The first thing to observe here is that not each and every use of a thick term makes reference to the normative dimension of reasons. For instance, there is a genuinely positive sense in which someone or something can be *different*: as opposed to *weird*. But the aforementioned concepts, albeit thick, need not always be about reasons (e.g. I may be weird in virtue of not liking pizza: different in a negative sense, though I do not necessarily have reason to change that). Yet some thick terms are in their use typically or canonically about reasons. Consider again the term c*ruel* — at least when applied to people. It sounds increasingly odd to claim that someone was being *cruel* while implying at the same time that she had every reason to be so as in 'The Nazi officer was *justifiably* appalled by the presence of the Jewish woman' — or conversely, that one failed to be cruel when she had every reason to be so as in 'He *shamelessly* refrained from setting the cat on fire'. Linguistic conventions are, of course, a fuzzy affair and it may at times seem unclear whether a relevant term applies or not. For example, it may be cruel to take an intruder's eye off while in self-defence, even though you have every reason to defend against his attack. But the cruelty does not consist in defending yourself, it rather consists in *overdoing* it: you only have reason to take as much action as to protect yourself, and anything beyond that seems conceptually unjustified.

Indeed, the externalist must welcome talk involving thick evaluative concepts. For claims about cruelty and inconsiderateness turn out to be not only informative about one's reasons, they also give us part of the explanation why the agent is negligent or in violation of her reasons. In being cruel, one goes against her reasons *in virtue of* inflicting unnecessary suffering to others. The externalist may also welcome claims that are strictly about external reasons and which are intended as *supplementary* to their more informative counterparts. For to add the purely normative judgement about reasons ('His act was practically unjustified') to the corresponding judgement about cruelty ('The nazi officer was cruel') is to lay emphasis on the normative dimension: to pinpoint the fact that the individual was either in violation or negligence about her reasons.

At this point, the internalist may retort that, *still*, claims about cruelty and inconsiderateness are not about reasons at all because for an agent to have a reason to Φ, it must be possible for that agent to be motivated for Φing given her existing motivations. In short: that such claims cannot be about reasons because they are clearly not about *internal* reasons. To claim this, however, is for the internalist to beg the question against her opponent insofar as she is insisting on a conception of *reason* on which only internal reasons count. What we need here is an independent argument in favour of internalism, not a restatement of the view.

5.6 Arguments for Externalism

Defensive work can be meritorious, yet the externalist can go to offensive and argue directly for her theory. I can see two basic ways that the externalist can lend direct support to her theory. The first concerns the theory's ability to get the *right* (or intuitive) results, while the second concerns the theory's ability to account for the intuitive distinction between normativity and psychology.

5.6.1 The Argument from Examples and Counterexamples

The externalist may appeal to examples of cases in which agents seem to have reasons that can only be explained as external. By appealing to such examples, the externalist can assert that her theory is more intuitive – or *extensionally adequate* (see 4.4) — than her rival's in that it can account for a wider range of ordinary intuitions regarding reasons. This, she may add, provides defeasible grounds for thinking that externalism is true and internalism false.

Consider a calm and collected individual who reasons logically and has all relevant empirical beliefs true, and who nevertheless wants to have a particular plastic bag that is now floating in the air. Crucially, we must suppose that this person neither wants the plastic bag as a means towards creating or achieving something else, nor because it happens to be symbolic of something.[79] He just wants the plastic bag for no further reason that it is a plastic bag — or even more astonishingly, *that* plastic bag. Suppose further that he is not willing to put any effort into grabbing it, nor does he confesses his desire for having it to anyone. These stipulations may turn out to be

[79] Another notable example is that of an individual who despite reasoning well wants a saucer of mud, and that of someone with a disposition to turn on radios whenever possible (Hooker & Streumer 2003: 67).

important because such pursuing could potentially harm the agent in some way that could matter to him. For if the agent also highly dislikes being humiliated in public, and by pursuing the plastic bag he would bring that about, the internalist may be able to explain why he does not have a reason to have that desire by reference to such disliking on his part.

Examples like the above seem particularly troubling for the internalist because they involve desires for things that seem intrinsically unworthy of being desired as ends in their own right (Hooker & Streumer 2003: 68). Moreover, by merely having such desires, as opposed to any intention to actually pursue the object of his desires, the agent is not even at risk of sacrificing anything that might get in the way of bringing about his desired state (as I explained above). Having mentioned this, I find it useful at this point to alter the above archetypal example so that it becomes at least possible for the internalist to explain the relevant intuitions by reference to some of the agent's existing motivations. In doing this, I will have hopefully made the argument from counterexamples even more convincing.

So suppose this time that someone else catches the plastic bag and makes our friend an offer: to give him a large portion of his life savings in order to have it. Our friend does not see the offer as an obstacle, and immediately agrees. For as he sees it, his desire for obtaining the plastic bag is insuperable by anything else he may desire. Again, it seems that our friend — who reasons logically and has all relevant empirical beliefs true — is nevertheless motivated to spend a large portion of his savings for something *that* ridiculous. At the same time, however, it seems highly unintuitive that he has a reason to do that. If anything, it seems, our friend has a reason to discard his desire for the plastic bag — and that reason seems on first reading to be *external*.

But the internalist may retort at this point that the reason that our friend has to discard his desire for the plastic bag is *internal* after all. She may claim that what such ridiculous desires have in common is that they conflict with basic desires that every agent has. For example, a desire that is common to all agents and which will be frustrated should the agent spend a large portion of his savings to acquire a plastic bag is the desire for a decent life; e.g. having the financial resources to support a lifestyle which includes eating good food, enjoying pay-per-view television shows, vacations etc. This response, however, is clearly ruled out by the supposition that the agent who is being motivated towards acquiring a plastic bag has undergone a process of deliberation that includes *high reflection, imagination and relevant true beliefs*. And indeed, it is at least possible to imagine such an agent who remains motivated towards things no matter how ridiculous they are for the simple reason that she lacks any antecedent motivation that could get in the way. Diogenes the Cynic, to give a notable example, used to live and sleep in a large jar in the marketplace and had no desire for anything *decent* by conventional standards. Still, he may well have had a strong desire for having a bag that the internalist would be unable to explain.

But perhaps this is too quick. The internalist could respond that even if the agent does not have *actual* desires that conflict with her ridiculous ones and hence give her reason to discard the latter, she may still have *counterfactual* ones. Granted, she may say, the desire for the plastic bag is in perfect alignment with the agent's motivational set, but by pursuing such desires — or even by just having them — agents deprive themselves of worthwhile activities; crucially, ones that they would *enjoy* and

hence come to *desire* were they to engage in them. This counterfactual stipulation is meant to show that agents do have reasons to discard ridiculous desires (or pursue ridiculous things) and that those reasons are *internal* ones (cf. Joyce 2001: 113). Back to the previous example and let us clarify that the agent is not the kind of person that likes concerts, and that he is instead the kind of person that likes totally dull activities such as looking at plastic bags. As a result, he spends all his mental energy desiring and wishing that he had *that* plastic bag, and he is even willing to spend his savings to acquire it. In doing all this, the internalist claims, he deprives himself from enjoying concerts and the financial means to attend them. The internalist can say that a fully informed and imaginative counterpart of the agent would advise him to discard her desire for plastic bags and substitute it for a desire for (say) concerts.

Let us grant that the person of the previous example would indeed enjoy concerts more than he enjoys plastic bags were she to go to one. On the assumption that were the agent to give concerts a chance she would prefer them over plastic bags, Elijah Millgram and Richard Joyce seem to disagree over the nature of the aforementioned reason, with the former thinking that it counts as an external reason and the latter insisting that it counts as internal.[80] Even though I tend to agree with Millgram that the relevant reason is external, I do not consider the issue as having too much significance for our purposes. First, I think the reason counts as external given that the idealised version of the agent would have to recommend the discarding of the desire for plastic bags without having anything from the agent's actual motivational set to back up the recommendation. But even if we can imagine the actual agent possessing some motivation to back up the recommendation, we can always imagine an agent that both lacks the required motivation *and* who (intuitively) has a reason to discard the desire for plastic bags. This brings me to the second point. For both Millgram and Joyce seem to be operating on the assumption that the actual agent would have a generic (or second-order) desire for enjoyable activities. But suppose that this assumption does not hold, and that our agent has no such generic desire for doing what would bring him enjoyment. Rather, he just wants to engage in prolonged, meaningless activities devoid of any enjoyment. In relation to such agents, we would still be inclined to say that they have no reason to have the ridiculous desires that they do, and reasons to discard them — and if they do have such reasons, the reasons must clearly be external.

Were the internalist at this point to claim that the idealised version of *any* agent would advice her against having desires for (say) plastic bags, her view would sound a lot more like externalism than internalism. For to claim that any actual agent has a reason to refrain from desiring plastic bags even though a desire to the contrary is absent in her motivational set, it is to suggest that agents have necessarily reasons for pursuing enjoyable and worthwhile activities irrespective of their motivational sets. In short: it is to insist that agents have *external* reasons to pursue such activities. As I understand the dialectic, internalists seem to have two options here: either to bite the bullet and concede that people can occasionally have reasons for plastic bags or any other ridiculous such thing they may desire, or to become externalists and thereby

[80] The two discuss the example of Archie who deprives himself of fulfilling relationships. I take it that the example used here is in all important respects similar to Archie's (Joyce 2001: 112-115, Millgram 1996).

provide themselves with the resources to accommodate the intuitive results that they so much strive to explain.

5.6.2 The Distinction Between Normativity and Psychology

Externalists are able to resist the claim that agents have reasons to have or pursue ridiculous desires, and hence their theory answers to some common intuitions. There is yet a further intuition that externalists are poised to respect, and that is what I shall call *the distinction between normativity and psychology*. In what follows, I explain what I take the distinction to be and how I think it lends further support to externalism.

The starting point of the current argument is what we may call the *reasonable and moderate individual*: someone committed to acting in accordance with her good practical reasons, and whose mind is characterised by an openness to more than one views about a given matter. As I understand it, the sensible individual stands in stark contrast to what we may call the *unreasonable and self-absorbed individual*: someone with complete disregard to her good practical reasons, and who is interested solely in what she thinks about any given matter. Now, we are here interested in the former, and we may start by noting that in the life of such an individual there will be moments in which she will find herself wondering:

> *Question*: 'Do I really have reason to do the things *I think* I have reason to do?'[81]

As a matter of fact, most people tend to be more like the reasonable and moderate and less like the unreasonable and self-absorbed individual. Perhaps that is why the question *makes perfect sense*: it seems to be a genuine, real question that when uttered under the right circumstances calls for a definite answer. And here the right circumstances are no other than the conditions of rational deliberation specified in the internalist's formulation: having true relevant beliefs, full reflection and imagination. If it is true that the question is perfectly intelligible — as I think it clearly is, a good theory of practical reasons must be able to explain why it is. And here, I think, externalism gains an edge over its rival. To see this, let us start by considering the following *question*:

> *Meta-question*: Why does *the question* make sense?

To answer the *meta-question*, the externalist must begin by noting that there are two possible forms that any given deliberation may take. The first corresponds to those cases in which the motivation that results from conditions of full information, reflection and imagination is indicative of one's true reasons, and the second corresponds to those cases in which the very same motivation is not indicative of one's true reasons. Call the first D1, and the latter D2. An example of D1 is the following: John, who likes watching movies and who enjoys the suffering of others, is between two practical choices: to watch the horror flick *Saw II*, or to watch the romantic *Sleepless in Seattle*. By carefully considering his liking for both movies and the

[81] I borrow the formulation from Bedke who makes a similar point (2008: 105).

suffering of others, John becomes motivated to watch *Saw II*, as this will satisfy his desire for watching movies and he may also get the occasional kick by watching fictional characters suffer. Thus he infers that he has no reason to watch *Sleepless in Seattle* — at least no more than he has to watch *Saw II*. Indeed, John seems to be right about this one — as intuitively, it seems the right judgement about his true reasons.[82] Now, we may consider an example of D2: John, who likes watching movies and who enjoys the suffering of others, is once again between two practical choices: to watch the horror flick *Saw II*, or to delay doing so in order to call the emergency line at the request of his dying neighbour. By carefully considering his liking for both movies and the suffering of others, John becomes motivated to watch *Saw II* right away — as he infers that by doing so he will be satisfying two of his desires at the same time: watching movies and enjoying the sufferings of others. But it seems that this time, John has got things wrong: for he seems to have every reason to call the emergency even if it means delaying the start of the movie for a few minutes.

Indeed, the externalist will notice, consulting our motivational tendencies that are the result of conditions of full information, reflection and imagination is ordinarily perceived as a prominent way for thinking about our reasons, to such an extent that we may be tempted to identify what we are motivated to do under such condition with what we may have actual reason to do. Now, by using the above mentioned possibilities (D1 and D2), the externalist can now explain why the *question* seems intelligible: occasionally, the agent may become confused as to whether she finds herself in D1-type of deliberation or D2-type of deliberation. That is to say that the agent may be occasionally unclear as to whether what she is motivated to do under conditions of rational deliberation is what she has actual reason to do. As Bedke puts it:

> Even our ideal selves can sensibly ask 'I think I have reason to Φ, but do I really have reason to Φ? (ibid)

By contrast, it seems somewhat unclear how the internalist can make sense of the *question*. For on their view, D2 seems to collapse into D1: what the agent has actual reason to do *just is* what the agent would be motivated to do under conditions of full information, reflection and imagination. But if this is so, then the *reasonable and moderate* individual who often exemplifies her moderation by entertaining the thought that what she thinks she has reason to do may not be after all what she really has reason to do, must be suffering from some kind of conceptual confusion. For in the internalist's story, it is simply conceptually impossible to distinguish between what the agent is ideally motivated to do and what she has actual reason to do. In short: despite appearances to the contrary, the *question* does not really make sense.

To be fair to the internalist, it is open to her to claim at this point that the conceptual possibility expressed by the *question* is only licensed by intuitions which do not stand up to philosophical scrutiny. And it is open to her to utilise substantive

[82] There is, however, a complication here that I am tempted to disregard. For it might seem that John must try to eliminate his liking for the suffering of others, and I tend to agree that he has a reason to do so. And if he can start doing it right away by watching *Sleepless in Seattle*, he may have a reason to do just that. However, for our purposes here, we may disregard this objection.

arguments against those intuitions — i.e. to argue against the plausibility of the theory that shelters them: externalism. But as I have hopefully shown in this chapter, some of the strongest arguments traditionally put forward in favour of internalism can be persuasively rebutted. This brings us to the conclusion of this chapter.

5.7 Conclusion

The aim of this chapter was to consider additional support for the instrumentalist's contention that the only reasons there are are those in favour of actions that are instrumental to the agent's (for whom they are reasons) ends. The additional support was to come in the form of a family of arguments in favour of *reasons internalism*: a view which for all intents and purposes is congenial to instrumentalism. Having rejected four arguments in favour of internalism, and embraced two in favour of externalism, I can say with some confidence that internalism is false. What this means for the bigger picture is that the error theorist's claim that moral norms do not invariably supply agents with sufficient normative reasons remains unsupported. If the dedicated immoralist has no such reason to act morally, this cannot be true because of internalism — or at least in virtue of what has been advanced here in favour of it.

6. A conceptual route to a substantive conclusion

6.1 Our Investigation into a Substantive Question

We are still interested in the following question: does the dedicated immoralist — who has only ends towards immoral actions, and can get away with pursuing such actions — necessarily have reason to act morally? More specifically, does he have sufficient normative reason to do so? In general, do moral norms invariably supply agents with such reasons? As I explain in *1.1.2*, in being *normative*, a sufficient normative reason contrasts with motivating and explanatory reasons: a reason *for which* someone does something, and a reason *why* someone does something, respectively. A normative reason for action is a reason there *is* for doing something: a consideration that justifies or supports a prospective course of action. A sufficient normative reason is one that outweighs all other relevant considerations that are at play, thereby determining what the justified course of action is: the one that is most reason supported. In being sufficient, it contrasts with *pro tanto* normative reasons: a consideration that justifies an action but which may be overruled by better ones. If all reasons, pro tanto or sufficient, are instrumental, then the answer to the thesis question is negative.

The error theorist of this context assumes — as we do — a certain understanding of the content of moral requirements: they need not prescribe actions that are fulfilling to the agent's (to whom they apply) ends (see 2.4). This means that a moral norm[83] – expressing an alleged moral requirement - may prescribe that an agent perform an action, despite it being true that the agent has no end served by doing so. The error theorist thinks that dedicated immoralists do not necessarily have sufficient normative reason to act morally because she thinks that all reasons *unequivocally* (pro tanto or sufficient) are instrumental: favouring actions that are instrumental to the agent's (to whom they are reasons) ends. Our journey in examining the error theorist's grounds for thinking that there are no moral reasons (sufficient normative reasons invariably supplied by moral norms) carries on in this chapter.

In this chapter, we turn our attention to an argument that may be advanced in favour of the view that all normative reasons are instrumental (6.4). The argument has a strong conceptual flavour, as it attempts to establish a conceptual link between the instrumentalist's conception of rationality (or rationality *in the narrow sense*, see 4.8.1) and normative reasons. Though conceptual in nature, the argument nevertheless promises to fulfil the substantive aspirations of the error theorist of this context, as the alleged conceptual connection may potentially license the conclusion that there are no moral reasons (in the sense already specified).

---

[83] It may be tempting to call the rules or norms of morality *requirements*. In this thesis, I take it that moral norms (or a considerable subset thereof) *express* requirements, and leave it open whether there really are any such requirements. As I argued in chapters 2 and 3, moral requirements are supposed to invariably supply agents with reasons, whereas moral norms merely express those alleged requirements. Here, my use of 'moral norm' will be elliptical for 'moral norms that express alleged moral requirements'.

6.2 Scepticism about the Justification of Morality

In *Humean Doubts*, James Dreier expresses scepticism about the justification of morality — which he takes to be the claim that morality lacks practical justification (Dreier 1997: 82). Establishing some key terminology is necessary to understand Dreier's claim. We may start by following Dreier in thinking of morality as a set of rules, or as I shall prefer *norms*. Together, this set of norms make up a *normative framework*: a set of norms that tell us what to do. This terminology enables us to discern the unquestionable similarities between morality, on the one hand, and any other set of norms on the other. In general, any set of (practical) norms may be regarded as a distinct normative framework. Examples may include anything from games such as football to well-thought-out legal frameworks such as the laws of the United Kingdom. This much should be uncontroversial.

Now, for any normative framework, we may ask for its *justification*. The notion of justification can be vague, so let me make it clearer. Suppose you are walking inside a shopping centre and a stranger approaches and asks you: 'Why are you walking on the right side of the lane, and not on the left?'. You respond to him with a hypothesis: 'You seem to be suggesting that I *should* walk on the right side of the lane'. Before you wonder: 'Am I *required* to do so, and *why*?'.

Let us pause for a moment, and observe that your response to the stranger's question seems to be perfectly in line. First, the question seems to presuppose that there is a set of norms directed at you and which require you to walk on the left side of the lane. Second, for any set of norms, we may sensibly ask for *justification*. Consider now the stranger's response: 'You should walk on the left side of the lane because that is the conventional way of doing so'.

The stranger's response is certainly a step in the right direction, as it serves to elucidate the norm by revealing its origin. It effectively locates the norm in the normative framework of *walking-inside-the-mall*. This first kind of justification is concerned with the *content* of a normative framework. It purports to show that *these* norms (e.g. you ought not kill) do constitute *this* normative framework (e.g. morality). Against it, one could hold that (say) the norm to walk on the left side of the lane is not really part of the conventional way of doing so; e.g. that he does not remember things correctly. We may call this first kind of justification *content justification*.

But there also seems something missing. Granted: we know what the norm(s) of *walking-inside-the-mall* is, *why* should we *do as* this norm says? In raising this question, we are asking for a different type of justification. We are interested in what *reason* do we, or anyone, have to follow these norms? Following Dreier, a request for (this kind of) justification is a request for *reasons* (83). This is the kind of justification that we are here interested in, and we may follow Dreier in calling it *practical justification* — henceforth, sometimes just *justification (82).*

In the previous example, the stranger may justify the norm to walk on the left side of the lane while in the mall by citing prudential reasons: when visitors heading to a single direction walk on a single lane (left or right), the circulation is facilitated, especially on busy days. This explanation will be a step in the right direction towards providing a justification for the norm, but one may still sensibly retort: 'Still, why

should *I* do as it says?'. It seems sensible to say that the aforementioned justification depends on our having a certain desire, say, to walk inside the mall with minimum friction. In terms of reasons, it may sometimes be that an agent has a reason to follow a norm, while on other times he does not. For example, the mall's manager may have a reason to walk on the left side of the lane on busy days, provided that he wants to walk with minimum friction, whereas he has no reason to do so when the mall is closed, provided that there are no people to interfere with his walking.[84]

As it may transpire, the notion of justification as used here is not meant to apply to the agent — or the agent's action — as when the agent may be thought to be justified or not justified in doing something. Rather, justification here is meant to apply to the norms themselves: to denote whether a norm is justified or not when directed at an agent. Thus a practical justification, as understood here, is one *for* a normative framework. The question we are going to be interested in here is whether morality — understood as a set of norms — is *practically justified* — whether there is a practical justification *for* morality. But the question may seem to make little sense if whether a normative framework is justified is always a matter of circumstance (an agent having a certain desire to comply with it, or finding herself in such-and-such circumstances). To avoid begging any questions, we shall regard as a possibility that a normative framework may be practically justified in the following sense:

> Necessarily, for all normative frameworks N, agents A, circumstances C, and actions Φ, a normative framework N is practically justified iff it is true that whenever N, and because N, requires of an A that she Φ in C, A has a sufficient normative reason to Φ in C, where that reason may be desire-based or not.

To put it another way: For a normative framework to be practically justified in our sense, it must invariably supply an agent with a sufficient normative reason to Φ whenever it requires that the agent Φ. To give an example, even though football requires that players refrain from touching the ball with their hands, it is still not justified in our sense since we can always imagine someone that is required by football to do such-and-such, but who lacks a sufficient normative reason to do so due to (say) having lost all interest in the game. It is worth emphasising here that we are solely concerned with what is *required* — as opposed to what would be *good* to do from the point of view of *this* or *that* framework. Compare: In football, when a team is awarded a penalty kick, the team is *required* to take it, and it would only be *good* (from the point of view of football) if the team also scores a goal in the process. Similarly with morality, one may be thought to be *required* (or *obligated*) to provide food for his own child, though it would arguably be merely *good* (from the point of view of morality) if he also provided food for other children.[85]

---

[84] There is a complication that I am willing to neglect here. It may seem that the norm to walk on the left side of the lane presupposes that the mall is crowded. If this is so, then it may seem that when the mall is closed, there is no norm directed at the person who walks the lane alone to begin with.

[85] In the case of morality, of course, it is arguably the case that what is good from the point of view of morality is also good from the point of view of practical reason. This, however, is precisely what this thesis is devoted to examining.

The sceptic's claim

The sceptic claims that morality is *not practically justified* in the above specified sense. Consistent with the above remarks, if the sceptic's claim is true, then it may be possible that an agent is morally *required* to perform an act without it being the case that she has also a sufficient normative *reason* to do so. Observe how the sceptic's claim addresses the question that this thesis centres upon: do moral norms invariably supply agents with sufficient normative reasons? In short, if morality is not justified in our sense, then there are no *moral* reasons: sufficient normative reasons invariably supplied by moral norms. (Similarly, if football is not justified in our sense, there are no *football*-reasons, and so on).

Of course, consistent with the sceptic's claim, it may be true that on some occasions an agents has a normative reason to be moral (or do as moral norms say), but that need not be the case. For example, on the assumption that I have a reason of self-interest to abide by the moral norm to give money to charity (say, because it would make me look good to others), my reason could hardly be described as *moral*. For one thing, were my interest to go away, the reason would go with it. For another, and more strikingly, my reason to give money to charity in this scenario, is not supplied by (the normative framework of) morality but by a different framework, say, self-interest.

The sceptic of this chapter is an instrumentalist about reasons, who attempts to show that as a matter of conceptual necessity, all normative reasons (pro tanto or sufficient) are instrumental. Her argument for this claim must wait for 6.3. Before we get there, we consider two rather straight-forward attempts at vindicating the claim that moral norms do not invariably supply agents with sufficient normative reasons.[86] These are rather naive attempts, for reasons to be unveiled, but they serve as a nice introduction to the most interesting argumentation to follow. I consider these because they seem to share a continuity with their more interesting counterparts of 6.3: they all suggest a conceptual link between rationality in some sense and normative reasons.

6.3 Two Naive Attempts

6.3.1 An Alleged Link Between Rationality (In Some Sense) and Reasons

Defenders of the view that there are no moral reasons often utilise a familiar line of thought, which starts with the idea that there is a very close link — or even an obvious conceptual connection — between rationality, on the one hand, and normative reasons, on the other (Broome 2005: 327). More specifically, the starting idea is that it cannot be the case that rationality requires an agent to perform an action, without it being also the case that the agent has a reason to perform that action. Consider the following passage by Joyce:

> If it is rational for S to φ, then S cannot just turn around and say something of the form 'But I do not care about φ-ing; I have no reason to do so.' There is a

[86] It must be noted that even though Dreier (1997) never seriously considers the naive attempts as strong support for his sceptical claim, both can emerge as misinterpretations of his own argument.

reason for S to φ, just in virtue of what we mean by 'rational' and 'reason' (Joyce 2000: 468).

Those who take the above idea seriously are often quick to point out that by 'reason', they do not just mean a reason *in some sense*. For example, they do not mean an *institutional* reason: a reason that one has solely in virtue of subscribing to an *institution*, also understood here as a *normative framework* (Joyce 2001: 100). Thus even if legal rules can be thought to generate *legal* reasons, associative obligations *associative* reasons, and moral norms *moral* reasons, we can always ask what reason we have to give reasons of a particular sort any weight (Shafer-Landau 2003: 166). If we have any such reasons, they must be the real deal: *normative* reason. Indeed, it is in this sense of 'reason' that Dreier thinks that 'reasons of rationality … are *ultimate reasons* in some sense' (88, my emphasis), and even more boldly, that 'reasons are *in terms of* rationality' (84, his emphasis).

The next step in the present line of thought, is to observe that it can sometimes be the case that an agent is rationally required to Φ, while she is also morally required not to Φ. Consider, for example, the suggestion that it can sometimes be immoral to believe ill of a person, even if it is rational to do so; as in the case of a parent who in the face of the evidence indicating that her child is 'dumb', she may still be morally required to disbelieve it.[87] In general, there seem to be traits, or requirements, that we associate with rationality (e.g. having true beliefs, being highly calculative and imaginative) and which need have nothing to do with traits, or requirements, that we associate with morality (e.g. being kind or compassionate). Another example is the case of *the Brain*: a highly intelligent planner, who develops complex plans for global domination. In pursuing his plans for global domination, it may be that the Brain abides by rational requirements without it being the case that he also abides by moral ones. In fact, much the opposite: the Brain seems to be acting in a paradigmatically immoral fashion.

Combining the two above steps, we get the following argument:

(1*) Rational requirements *entail* sufficient normative reasons.
(2*) Rational requirements can be incompatible with moral norms.
(A) Sufficient normative reasons cannot conflict.
(C*) Moral norms do not entail sufficient normative reasons in cases where they are incompatible with the requirements of instrumental rationality.

On the assumption (A) that normative reasons cannot conflict, the argument may seem initially plausible, but it need not be so. To see this, observe that the notion of rationality is a slippery one (Shafer-Landau 2003: 167), and for the purposes of the argument ambiguous in a crucial respect. In one sense, rationality may be associated with traits such as being highly calculative, imaginative, responsive to evidence and, crucially, the intention to take the necessary means to one's ends. In 4.5, I called this *rationality in the narrow sense*. Yet in another sense, rationality may be associated with the capacity to be appropriately responsive in one's attitudes and behaviour to reasons, where what counts as reasons need not be determined by the idea of rationality

[87] (Kelly & Roedder 2008: 529)

itself but by a substantive theory of reasons (Shafer-Landau 2003: 168, Kolodny 2005: 510). In 4.5, I called this *rationality in the broad sense*.

For this first argument to go through, the sceptic about moral reasons must insist on the first conception (rationality in the narrow sense), and hence build it into (1*). For if we go with the second conception (rationality in the broad sense), the truth of (2*) is severely threatened, as it is far from clear how the rational requirement to be responsive to reasons is incompatible with moral norms. For example, suppose that I have a moral reason to help you despite having no desire or concern for doing so, and that rationality requires of me that I act on this reason of mine. The upshot here is not that the availability of the two conceptions of 'rationality' *shows* that this first argument is false, rather it is that the pressure is on the defender of the argument to insist on the particular conception of rationality that can make (1*) come out true. For such an attempt, we have to look at the next argument.

At this point, a terminological point is in order. From now on, I shall exchange the label *rationality in the narrow sens*e, which denotes the conception of rationality that is congenial to the instrumentalist's argument, with *instrumentalist conception of rationality*. The latter characterisation is fitting as it concentrates on the particular requirement that is implicit in the conception of rationality in the narrow sense and which is thought to be potentially incompatible with moral requirements: (roughly) to take the necessary means to your ends.[88]

6.3.2 Insisting on the Primacy of the Instrumentalist Conception of Rationality

The second argument builds on the work of the previous one, while implementing a crucial addition. That is, it maintains that the requirements of rationality (in some sense) are connected to reasons, and that the requirements of rationality (in some sense) can be incompatible with moral norms. But in order to avoid the previous problem of ambiguity, this argument needs to insist on a particular conception of rationality, and unsurprisingly, that conception must be the instrumentalist conception. This is how I read this second argument:

> (1**) The requirements of instrumental rationality entail sufficient normative reasons.
> (2**) The requirements of instrumental rationality can be incompatible with moral norms.[89]
> (A) Sufficient Normative reasons cannot conflict.
> (C**) Moral norms do not entail normative reasons in cases where they are incompatible with the requirements of instrumental rationality.

[88] Less roughly, the requirement that: if you intend to G, and if you believe your Fing is a necessary means to your Ging, and if you believe you will not F unless you intend to F, then you intend to F (Broome 2005: 322).
[89] This premise has been defended in *2.3*, *2.4*.

As with the previous one, this argument looks initially plausible. Things are relatively straightforward with the second premise (2**), for reasons to be shortly explained. Moreover, the combination of premises (1**) and (2**) does validate the conclusion (C**). The most complex matter, however, is motivating premise (1**). In particular, what needs motivating is the insistence on (1**) over the more innocuous (1*) of the previous section.

(1*) Rational requirements entail sufficient normative reasons.

Before I elaborate, let me explain my reasons for conceding the truth of (2**), which I have already given in chapter 2: Moral norms are diametrically opposed to the requirements of instrumental rationality, in that only the latter (and not the former) depend on the agent's (to whom they apply) contingent ends for their validity. Put simply, what instrumental rationality requires an agent to do may be entirely independent — if not contradictory — to what morality requires an agent to do (e.g. morality may require that Hannibal refrain from eating his friend, but Hannibal may have a strong desire for human flesh). Following Shafer-Landau, 'it [would] be nice if that were wrong, and there were a perfect coincidence of self-interest and morality', but it would be too much to assume such an optimistic view at the outset (Shafer-Landau 2003: 166). Furthermore, even if it was true that, as a matter of perfect coincidence, each and every time a moral requirement applies to an agent, the agent *also happens* to have a sufficient normative reason of self-interest to abide by that moral norm, that would not entail that there are *moral* reasons. That is, the such reasons would not be *supplied* by moral norms — insofar as agents would have them *not* in virtue of a moral norm that obtains. Rather, such reasons would be best described as *reasons of self-interest for agents to be moral*.

Now, the task of motivating (1**) boils down to the task of motivating the insistence on the particular conception of rationality featuring in the first premise (1**). How can the instrumentalist do this? From Dreier, we get a hint towards a possible way of motivating the insistence:

> [T]here is something special about exactly the kind of norms of rationality that Humeans accept (Dreier 1997: 92).

The suggestion seems to be that the instrumentalist conception of rationality is in some sense *primary* — or at least *more basic* than any other competing one such as the conception of rationality in the broad sense.[90] Given its primacy, the current idea goes, this conception must be the one conceptually connected to reasons (or justification). The underlying idea here seems to be that the sense in which the instrumentalist conception is primary is *relevant* to its status as being conceptually connected to reasons. The current suggestion provides at least a promise that the instrumentalist may be able to license her insistence, thereby vindicating the crucial premise (1**) for the conclusion that there are no moral reasons.

[90] Note that Dreier gives an argument for the above claim, which will be presented and addressed in 6.4.2. Until I reach that point, I shall first examine other possible routes for licensing the crucial insistence on the instrumentalist conception of rationality.

At this point, I cannot help but express my awe for the task awaiting the instrumentalist. For I think that showing that one conception of rationality is primary — let alone in the relevant sense — is a very hard enterprise. To see this, consider that both conceptions of rationality considered thus far (*instrumentalist* or *narrow*, and *broad*) are pervasive and seem equally indispensable in making sense of a large number of ordinary statements regarding the rationality of actions. From my own experience with philosophy first-year undergrad students, I have noticed that people tend to acknowledge both notions in their thought experiments. For example, one student once mentioned that even though Adolf Hitler was *in a sense* rational in taking the required steps to further his goals, he was still irrational *in the particular sense* that he was neglecting his moral reasons. Reflective statements such as these coming from philosophy students — people who have not yet been completely 'corrupted' by philosophical theory — demonstrate the pervasiveness of both conceptions. The upshot here is that the task of demonstrating the primacy of one conception over the other need not be an easy one.

Let me now reject, and get out of the way towards a good answer, a few implausible proposals. The first is the idea that the instrumentalist conception is primary in virtue of being the *most popular and frequently used* by philosophers and non-philosophers alike. That is, this particular conception features in our assessment of other peoples' behaviour or intentions most of the time, and this is perhaps evidence that it is also primary in the relevant sense of being conceptually connected to reasons. Though there seems to be some correlation between popularity and truth, at least with respect to some topic, this reasoning seems obviously flawed. Compare: The fact that *Foucault* shows up far more times in Google than *Chomsky* may indicate that the former is more popular than the latter in mainstream philosophical circles, but it need not show that the former was right in their disagreement on the *modernism*/*post-modernism* debate.

A second, related proposal may be that the instrumentalist conception is primary in virtue of being the one most closely associated with human nature itself. That is, provided that humans have a natural and reliable tendency to operate in accordance with their own interests (i.e. in ways that tend to be accord with the instrumentalist conception), the instrumentalist conception may be deemed 'truer' of human nature. Be that as it may, however, this does not show that the instrumentalist conception is primary in the relevant sense: that of being 'truer' of reasons — i.e. the one truly associated with reasons. This criticism echoes a previous one targeted at internalist conceptions of reasons, namely the conflation of normativity with psychology (5.5.2).

6.4 Two Sophisticated Attempts

Let me briefly summarise the dialectic. The instrumentalist has given an argument which is valid and, if true, can license the conclusion that there are no moral reasons. However, she finds herself in a certain predicament. In order for the argument to go through, she needs to motivate the premise (1**) that *the requirements of the instrumentalist conception of rationality* entail sufficient normative reasons. Unfortunately for her, to insist that the instrumentalist conception of rationality is primary *in some sense*, is not enough to make us accept the premise. Rather, it must be shown that the conception is primary in the relevant sense: in that it is conceptually

connected to reasons. As I said in the beginning, the instrumentalist's argument of this chapter will have a strong conceptual flavour, but if successful, it can have substantive results. That is, through an argument establishing a conceptual link between instrumental rationality and normative reasons, the instrumentalist may be able to license the conclusion that all normative reasons (*pro tanto* or sufficient) are instrumental — i.e. rougly, instructing agents to take the means to their ends — and hence that there are no moral reasons.

6.4.1 Insisting on the Primacy of the Reasons of Instrumental Rationality

How can the instrumentalist argue for the claim that the instrumentalist conception of rationality is conceptually connected to reasons? I can see one way that this can be done, namely to show that the instrumentalist conception of rationality is primary in the sense that the reasons associated with it — i.e. instrumental reasons: reasons that favour actions that are instrumental to the agent's ends — are *conceptually basic* (or primary). A good starting point for understanding the notion of conceptual primacy is offered in the following passage:

> To hold that a given concept is the fundamental concept within a region of thought is to make two claims: first, that it cannot be analysed in terms of other concepts from that region; second, that those other concepts can be analysed in terms of it (Robertson 2009: 6).

Let me try to illustrate the notion of conceptual primacy that I have here in mind by using a different example. Suppose someone who is not at all familiar with football asks us what football is: what is the concept FOOTBALL? We say that football is a game in which eleven players compete against … — at this point, our friend rapidly stops us: what is a *game*? We say that a game is an activity whose point is … — once again, our friend stops us: what is an *activity*? We feel at loss, as we seem to be running out of ways to explain to our friend what football is. What has gone wrong? I think there is a prominent — and highly relevant for our purposes — way to explain this. An analysis of our concept FOOTBALL must make reference to the concept GAME which, in turn, must make reference to the concept ACTIVITY. For football is a game, and a game is an activity. Neglecting any complications, we may say that in the domain of thought that we find our concept FOOTBALL, the concept ACTIVITY is *conceptually basic:* an analysis of either FOOTBALL or GAME must include it or make reference to it. Moreover, it may be helpful to think of the relation between these concepts in terms of *conceptual priority*: our concept ACTIVITY is conceptually prior to both our concepts GAME and FOOTBALL, while our concept GAME is conceptually prior to our concept FOOTBALL. In general, a concept X is conceptually prior to another in case an analysis of the latter requires reference to the former (but not *vice versa*).

The instrumentalist's claim here, as I understand it, must be that in the domain of thinking about normative reasons for action, the concept of the reason associated with the instrumentalist conception of rationality — i.e. the concept INSTRUMENTAL REASON — is conceptually *basic*: an analysis of any other kind of reason must make reference to it. One thing that this means is that for us to give anything close to a complete conceptual analysis of *any* concept of a normative reason for action (e.g. MORAL REASON), it is necessary that we make reference to our concept INSTRUMENTAL

REASON at least at some level of the analysis. For as I understand the claim, there is no sense to be made of the concept of a normative reason that resists analysis in terms of the concept of an instrumental reason. In a nutshell: the concept of an instrumental reason *just is* the concept of a normative reason for action. A related way to understand the claim is that to *grasp* the concept of any kind of reason — or understand any reason-claim — it is necessary that one first grasps the concept of an instrumental reason. Consider the following claim by Dreier:

> [T]here is no sense at all to be made of the question whether we have a reason to follow the rules of [instrumental] rationality. To think that there might be reasons for following the rules of rationality, I would say, is to *misunderstand* what reasons are. Reasons are in terms of the rules of [instrumental] rationality (Dreier 1997: 84, my emphasis).

I attribute to Dreier — the prominent instrumentalist in this debate — the claim that the concept of an instrumental reason is conceptually basic in the previously specified sense. On the assumption that he thinks this, it is easy to understand why he also thinks that there is no sense at all to be made of the question whether we have a reason to follow the rules of instrumental rationality. For on his view, reasons *are in terms of* instrumental rationality: reasons *just are* instrumental reasons. To ask whether you have a reason to be instrumentally rational, on the instrumentalist's analysis of normative reasons, is to exhibit a failure to understand what reasons are, in a similar way that asking whether the four-legged object that is in front of you is a chair is to exhibit a failure to understand what chairs are.

Now, by supplementing the argument of the previous section with the premise about the conceptual primacy (or basicness) of instrumental reasons, we seem to be validating it. In particular, the first premise (1**) would be true as a matter of definitional necessity. For those — such as myself — who do not 'see' the truth of the premise, we would have to add this further, suppressed premise:

(0**) Reasons *just are* reasons of instrumental rationality.

This would then license the premise with which we began:

(1**) The requirements of instrumental rationality entail reasons.

The addition of the conceptual premise (0**) could make the argument work, except it does not. For, *intuitively*, it does not *seem* true. Given that the premise is essentially conceptual, intuitions here may be good informants. Dreier thinks that there is no sense to be made of the question whether we have reason to do as the requirements of instrumental rationality say, but other people (including myself) insist that the question does make sense. To give an example, John Broome states his firm belief that 'the question makes good sense' (Broome 2005: 331). Thus we have, on the one hand, those who regard the question sensible, and on the other, those who do not. Before we deem the disagreement a stalemate, we must reflect on it in an attempt to understand it better and see which side, if any, has the burden of proof for their claim.

There is, I believe, a way to ascertain which side has the burden of proof here. To see this, let us begin by considering the following reason-claim:

> Andy, who has no concern whatever for the well-being of the people inside the burning building which he is now exiting, has nevertheless a normative reason to pull the fire alarm and save their lives (Bedke 2008: 107).

Indeed, a lot of people — philosophers and non-philosophers alike — would think that the above reason-claim is true, and it is here meant to be a paradigm case of a considerable portion of reason-claims that seem to call attention to non-instrumental reasons. Now, suppose that the reasons associated with the instrumentalist conception of rationality are indeed conceptually basic (as the instrumentalist would have it). This would have the astonishing result that in order to make sense of statements such as the above, we would have to analyse them in terms of reasons of instrumental rationality. That is, we would have to posit some desire or concern of Andy which makes it true that he has a reason to pull the alarm — or else deny the reason-claim. But, crucially, since we are taking it for granted that Andy has no desire or concern for the well-being of the people inside the building, and given that there is no other desire or concern of his served by his pulling the alarm, reason-talk such as the above would have to be deemed false, if not *unintelligible*. That is, the latter seems to follow from the instrumentalist's contention that an analysis in terms of instrumental reasons is ultimately the only way to make sense of a reason claim. So effectively, for the instrumentalist's contention here to be true, it would have to be the case that people who find such reason-talk intelligible lack competence with respect to the very concept of a normative reason.

The above is indeed a very hard pill to swallow for non-instrumentalists, especially considering the fact that the converse is not true. Consider this time the following reason-claim:

> Andy, who has a strong liking for ice cream and is now craving some, has a normative reason to enter the ice cream shop that is in front of him.

As with before, a lot of people — philosophers and non-philosophers alike — would think that the above reason-claim is true, and it is here meant to be a paradigm case of a considerable portion of reason-claims that seem to call attention to instrumental reasons. The non-instrumentalist of this section, despite denying the conceptual basicness of the reasons of instrumental rationality, nevertheless would — could, and should — find the reason-claim intelligible. In short: the non-instrumentalist need not reject any intuitions about reasons. The same, however, is not true of the instrumentalist — as we have seen above.

The upshot here is that from the two sides, it is only the side of the instrumentalist who insists on denying intuitions held by the opposing side, such as those exemplified by the reason-claim about Andy who is exiting the burning building. This observation prompts me to locate the burden of proof at the instrumentalist's side. Fortunately for the instrumentalist, this preliminary conclusion need not be final in this dialectic. James Dreier, the prominent instrumentalist in this discussion, develops a promising argument in an attempt to discharge this apparent burden.

6.4.2 An Argument in Favour of the Alleged Primacy of the Reasons of Instrumental Rationality

We finally arrive at Dreier's argument in favour of the view that the reasons of instrumental rationality (or the reasons associated with the instrumentalist conception of rationality) are conceptually basic. Dreier's argument is long and careful, but I shall take effort to unpack it and present each step of it in clear and orderly way. For the purposes of this thesis, understanding Dreier's argument is important as it enables us to pinpoint a glaring flaw in some of the instrumentalist's finest pieces of reasoning.

As a preliminary, we are invited to see the instrumentalist conception of rationality as a set of rules, or *requirements*, all of which are structured around the *means/ends rule* (93-94):

> (M/E) *Ceteris paribus*, if you desire to Ψ, and believing that by Φ-ing you will Ψ, then you ought to Φ.

This depiction of the instrumentalist conception of rationality as a set of rules may strike some as far-fetched, given (for one thing) that ordinary people would not refer to rationality as a set of rules akin to (say) football or even see it that way. Nevertheless, this depiction is innocuous for our purposes as it need not distort any conclusion regarding any conceptual connection of rationality to reasons. Moreover, as I already said, it allows us to see the unquestioned similarities between rationality and other normative frameworks.

'What is Ann missing?'

We are invited to imagine the disobedient Ann, who has a general tendency to disobey rules whenever those apply to her. Dreier considers (what he takes to be) two contrasting — in the explanation of their origin — instances of Ann's disobedience, and he attempts to draw some interesting results from the contrast in question. In the first instance, Ann is told that she ought to Φ — and when she shrugs, she is told that there are some rules, say, the rules of *good manners*, that require her to do so. Even after acknowledging this normative fact — that she is required to Φ by this set of rules — Ann remains unmotivated to act in accordance with it. We even go as far as to explain how culturally significant our system of *good manners* is, and that its rules are predicated on the standards of good taste. Despite all this, Ann remains unmotivated. These rules have no grip on her (92).

Quite sensibly, Dreier asks: What is Ann missing? He is quite confident, and we may be with him, that we have enough information about Ann's case to infer that what she is missing is a desire.[91] As he eloquently puts it: 'Since she believes that were she to Φ, she would comply with R, and we want a state that gets her from that belief to the motivation to Φ, we know that the state is the desire to comply with R' (92). And the same can be inferred with respect to Ann's shrugging-off of any other rules that

[91] Following Dreier, we may understand a desire here as term of art covering 'what is covered in ordinary language by the notions of desire proper, wanting, valuing, having a goal, preferring, and probably many other things' (87).

may be placed upon her. For example, tell her to act for a moral rule (or norm), and if she shrugs, we may reasonable infer that she lacks a desire to do as the rule says. In general, Dreier proposes, we may establish the following methodological axiom (MA):

> (Axiom): Whatever one is missing when she has a belief that she is required to Φ, and lacks motivation to do so, must be a desire (of some sort) (93).

With the above case-cracker at our disposal, we are now asked to imagine a second instance of Ann's disobedience. Suppose that we know for a fact that Ann wants to raise her chances of getting into a competitive law school — and that this much she openly admits. Upon acknowledging this end of hers, we tell her that she ought to Φ: go to a prep school. More carefully, we tell her that *given* her end of getting into a competitive law school, Φing is absolutely necessary. In response, Ann acknowledges that in order to get into a competitive law school, she must go to a prep school: that Φing is indispensable towards achieving her objective. But for some reason, she remains unmotivated to do so. As a first reaction, we cite the rule of the instrumentalist conception of rationality, *M/E*, and we remind her that *M/E* requires her to go to Φ. Still, Ann remains unmotivated: this rule has no grip on her. The question that now invites itself is: can we use the above methodological axiom to explain Ann's disobedience — at least in the same way we did with respect to the first instance? That is to say, is it what Ann is now missing a desire (of some sort)?

Dreier thinks *not* — and for the purposes of understanding his argument, it is important to see why he thinks this. There is indeed something special about *M/E* and the rules of instrumental rationality: they are always attached to, or predicated upon, some end or desire of the agent to whom they apply. This is not true of most if not all other normative frameworks; morality, etiquette and whatever the case may be. When we ask Ann to do as morality or etiquette say, we should take it as a genuine possibility that the rule to which we are calling attention and Ann's desires may be calling for different or even conflicting actions. For example, etiquette may require that Ann eat peas with a fork, but Ann cannot stand forks and prefers to eat peas with a spoon. In this example, supplementing Ann with a desire to eat peas with a fork may be just all we need to get her to comply with etiquette (or even with a general desire to comply with etiquette). In general, it is precisely the potential discrepancy between the object of Ann's desires and that of those rules, along with the homely thought that motivation requires a desire (of some sort), that allows us to infer that the agent is missing a desire (of some sort) when she acknowledges that she is required to Φ and remains unmotivated towards Φing. And these two observations enables us also to know what it takes to instill motivation in Ann to comply with our rules (say) of etiquette: supplement Ann with a relevant desire.

But things are much different with respect to *M/E*: all it asks, or requires, is that the agent act on her desires. That is, whatever is that Ann wants to do, *M/E* would have Ann do. Suppose that Ann wants to become a pilot: *M/E* would require Ann to take the necessary means towards becoming a pilot; Suppose that Ann changes her mind and now wants to become a veterinarian: *M/E* would require Ann to take the necessary means towards becoming a veterinarian, and so on. In general, were we to supplement Ann with a different desire, *M/E* would follow us in requiring Ann to pursue that desire. It thus seems, at least at first approximation, misplaced to say that

a desire is what Ann is missing when she remains unmotivated to comply with *M/E*. This is at least what Dreier thinks:

> Once you have (accept) the means/ends rule, what you need to get you to acceptance of other rules is one or another desire. But no desire will get you to the means/ends rule itself (94)

So what, if anything, is Ann missing when she remains unmotivated to comply with *M/E*? Dreier's diagnosis is:

> *[Ann] fails to be motivated by the acknowledged means to her desired ends*. So adding a desired end does not in her bring about the motivation to perform the acknowledged means to that end (94, emphasis added).

We may follow Dreier in granting that the instrumentalist conception of rationality is *special* in the following way. Whereas a desire is necessary for an agent to be motivated towards acting in accordance with any other rule, a desire is neither necessary nor sufficient to act in accordance with *M/E* of the instrumentalist conception of rationality. Rather, what is required is *acceptance* of it — whatever this means. In short: a desire brings motivation in the agent to *R* (where R is complying with any other rule), but a desire does not bring motivation in the agent to *M/E*. In a nutshell: the motivation to *M/E* is not dependant on desires. We may call this *special status*.

Before we move on to consider what may follow from *special status*, it is noteworthy pointing out that it is not clear that no desire can bring one to comply with *M/E*: that *M/E* is special in the above sense. Granted, Ann desires to increase her chances of getting in a competitive Law school; so she is certainly not missing *that* desire. But we must note here that whenever *M/E* applies to an agent, there can be two distinct levels of desiring: the *specific* and the *general*. On the specific level, the agent may or may not desire to comply with M/E *in pursing a given end* (e.g. to increase the chances of getting into Law School). On the general level, the agent may or may not desire to comply with *M/E in pursing any given end*. Dreier disguises the general level of desiring by saying that the agent may or may not *accept* the rule. But isn't 'accepting' here just another species of desiring? In fact, in an earlier passage, Dreier himself seems to admit this much: 'I am using "accepts" for whatever it takes actually to be motivated by the kind of reason in question' (89). Thus if *accepting* here may be understood as another species of desiring, it is not clear why we have to accept that no desire in and of itself can make one act in accordance with *M/E*. For the sake of argument, however, I shall grant that M/E is special in the above specified sense — for at least this is definitely the case with respect to the *specific* level of desiring: if one lacks motivation to comply with *M/E* in pursuing a given end, no specific desire (regarding that end) will result in motivation in the agent to pursue that end.

Having established that 'means/ends rationality' is special in the above specified way, we may sensibly ask whether that special status is *normatively* relevant: can it license an interesting conclusion about the status of the reasons associated with the instrumentalist conception of rationality, that is, instrumental reasons? Dreier thinks it can — and as I shall explain shortly, we may well interpret him as thinking

that the special status of ‘means/ends rationality’ suggests that instrumental reasons are conceptually basic in the sense specified previously.

Can we draw a normative conclusion?

Though it is not perfectly clear to me why Dreier thinks that special status is normatively relevant, I can discern the considerations that drive his thinking on this. Suppose, we tell Ann that she ought to Φ — and she is asking for a reason to Φ. We then explain to her that she is required to Φ by some normative framework, say, *the laws of Neverland*. Effectively, what we are saying to her is this: the fact that *the laws of Neverland* require to Φ is a reason for you to Φ. Ann is not impressed by our answer, and she holds on to her initial question: ‘But what reason do I have to Φ?’. Dreier admits — as I believe we all should — that Ann's question is perfectly in order: still, why should she do as we or *the laws of Neverland* say — what reason does she have to?

Now, suppose this time, we tell Ann that she ought to Φ — and she is asking for a reason to Φ. This time, we explain to her that she is required to Φ by *M/E*. In other words, that she is required to Φ by the normative framework that defines her reasons in terms of the desires that *she* has. Effectively, what we are saying to her is this: the fact that you desire to Φ is a reason for you to Φ. Ann — being Ann — is not impressed by our answer, and she holds on to her initial question: ‘But what reason do I have to Φ?’. Dreier now thinks we are at loss:

> Not merely at a loss to provide a compelling answer, but at a loss to know what to think of such a person. What would *count* as a reason, by her lights? (95).

We need not downplay the importance of the above claim. We are here considering someone who knows that she is required by *M/E* to Φ and who keeps on asking for a reason to Φ. That is, someone who knows that were she to comply with *M/E* she would have to Φ, but who is nevertheless unclear as to whether she has any reason to Φ. Dreier seems to be effectively saying that such a person cannot be considered competent with regards to the concept of a normative reason for action. For *we* have a concept of a reason, and *she* uses the term 'reason' to refer to one such concept when asking for a reason, but (Dreier seems to think) we are pretty certain that she is not referring to our concept of a reason. This makes us wonder what, if anything, would count as a reason for her. In short, the suggestion is that such a person does not know what *our* or *the* concept REASON is. Given this, whatever concept, if any, comes up in that person's mind when she uses the term 'reason' cannot be *our* or *the* concept of a reason. And if we grant this much, does it not follow that the reasons associated with the instrumentalist conception of rationality are conceptually basic (in our sense)? I think the answer is yes. For if a concept *X* is deemed unqualifiedly to deserve the label 'reason' solely in virtue of the fact that it resists analysis in terms of the concept of an instrumental reason, it seems to follow that a reference to the concept of an instrumental reason is required in any analysis of a concept deserving that label. In other words, that the concept of an instrumental reason is conceptually basic in the domain of thinking about reasons.

So the question is: Is Dreier right in supposing that anyone confronted with a rule of the instrumentalist conception of rationality (M/E), and who keeps asking for a reason to comply with it, cannot be considered competent with respect to our concept of a practical reason? Before I answer this question, I shall first grant that there is indeed something odd about anyone who is asking for a reason when confronted with M/E. But crucially for our purposes, the oddness pertains not to what reasons are. That is, *in a sense*, to ask for a reason when confronted with *M/E* is unintelligible, but this is not the relevant sense of unintelligibility which has to do with normativity and reasons, and which concerns us here.

Let me explain. There seem to be (at least) two things that the question 'What reason do I have to Φ?' may be asking or implying. First, it may be asking for a *motivating* reason to Φ, or implying that the agent to whom a rule applies lacks any motivation to act in accordance with it. Second, it may be asking for a *normative* reason to Φ, or implying that the agent to whom a rule applies lacks or sees no reason to Φ. Now, suppose someone is asking for a *motivating* reason when confronted with *M/E*: we may well interpret her statement (and herself) as unintelligible — that is, unintelligible *with respect to motivating reasons*. For a motivating reason to Φ is a consideration that actually does motivate the agent to Φ, and a motivation to Φ — we are happy here to establish — follows reliably our desiring to Φ. Someone who has a desire to Φ, and who still inquires about what motivating reason he has to Φ, simply does not know what a motivating reason is.

But crucially, it seems far from clear that the same charge of unintelligibility applies to someone who is asking for a *normative* reason when confronted with M/E. For one thing, it seems intuitive to think that a normative reason to Φ is a consideration that justifies one's Φing — though not necessarily one that actually motivates one towards Φing. For another, the question 'What *normative* reason do I have to Φ?' seems perfectly sensible when coming from someone who is confronted with *M/E* — that is, someone who is told that she ought to Φ in virtue of having a strong desire towards Φing. We have no shortage of examples confirming this intuition. Consider someone who has a strong desire for harming others — and who can do so while easily getting away with it. To question whether he has a *motivating* reason to harm others would certainly be revealing of an incapacity to understand what such reasons are. But to question whether he has a *normative* reason to harm others would not only be insufficient to lead us to the same conclusion, but it would in fact be highly indicative of that person's capacity to understand what normative reasons are — or so I submit.

Thinking the above is consistent with an observation that I think we are licensed to make, namely that *M/E* is *motivationally* special — not necessarily *normatively* so. That is, for all his efforts, the instrumentalist here manages only to show that there is a close, undeniable connection between an agent being motivated towards Φing, on the one hand, and an agent being required to Φ by *M/E*, on the other. What the instrumentalist who aspires to show that the concept of an instrumental reason is conceptually basic must show is that there is a similar close, undeniable connection between an agent having a normative reason towards Φing, on the one hand, and agent being required to Φ by *M/E*, on the other. This, I submit, is no easy task. As a first approximation, in inferring that anyone who is asking for a reason to follow *M/E* is unintelligible in the crucial, normative sense, the instrumentalist seems

to be committing the fallacy of confusing normative considerations with motivational ones.

The flaw in the Instrumentalist's reasoning

In closing this chapter, I want to call attention to what I take to be the erroneous assumption underlying the instrumentalist's reasoning. In particular, this is an erroneous supposition regarding — what I will follow Dreier in calling — the *fundamental practical inference*. Consider the following passage:

> We can give you practical reasons by finding things you want, and some things you believe, and getting you to draw *practical inferences*. If you can't draw *the practical inference*, not even the *fundamental*, M/E kind, then nothing counts as a reason for you (96, emphasis added).

Let us ask: What is it meant here by a *fundamental practical inference*? An inference is a step in reasoning — such as the move from a premise or a number of premises to a conclusion. I understand a practical inference as a reasoning aiming to draw a practical conclusion: about what one *ought* to do, or what one has *sufficient reason* to do. An example is this: I know you like ice cream, I want you to be happy, there is an ice cream in the area, I conclude (or infer) that I *ought* to buy you some ice cream. An inference (practical or otherwise) may be valid or invalid: valid when it is logical or rational (in a broader than merely *practical* sense), invalid when it is not. This much will do for our purposes.

Now, when Dreier calls the *M/E* kind of inference (roughly, you ought to take the necessary means to your ends) the *fundamental practical* inference, and based on other things he says, I can interpret him as saying that this is the inference that *underwrites* or *validates* all practical inferences. How does it do that? The answer is: by being implicit in any instance of practical reasoning, not as a premise but as an *inference rule*. An instance of practical reasoning or inference that does not abide by the M/E inference rule fails to be valid in the same way that an instance of theoretical reasoning or inference that does not abide by modus ponens does.[92] To illuminate the idea further, consider an instance of invalid theoretical reasoning:

(P1) If I go to the party, I will drink beer.
(P2) I will go to the party.
(C) I will not drink beer.

Insofar as modus ponens is indeed a fundamental practical inference governing all instances of theoretical reasoning, the inference from (P1)-(P2) is invalid in virtue of not respecting modus ponens. Now, consider the following instance of practical reasoning:

(P1) I have the end of hurting John.

[92] This purported analogy between practical and theoretical reasoning is used by Dreier as an argument, which he calls the Tortoise Argument (94). I come back explaining why the analogy is a bad one shortly.

(P2) I can hurt John by calling him nasty names.[93]
(C) I ought *not* to call John nasty names.

Were we to grant the instrumentalist's contention that the M/E inference rule is the fundamental practical inference — the one that validates any instance of practical reasoning — we should deem the inference from (P1)-(P2) to the conclusion (C) invalid in virtue of not respecting that inference. In general, the instrumentalist of this chapter would have us think, any instance of practical inference that does not respect the M/E inference rule fails to be valid. To the instrumentalist, this is an important part of the explanation why instrumental reasons (those favouring actions that are instrumental to the agent's ends) are conceptually basic: a non-instrumental reason is not a *valid* reason — in short, it fails to be a reason — in virtue of resisting analysis in terms of instrumental relations between the agent's means and her ends.

But now it is clear to me what the instrumentalist fallacy *par excellence* is: conflating two possible readings of the means-ends inference, and erroneously thinking that the stronger one is the fundamental practical inference. The one reading of M/E inference — which has been already mentioned — is this:

(M/E) *Ceteris paribus*, if you desire to Ψ, and believing that by Φ-ing you will Ψ, then you <u>*ought*</u> to Φ.

Consider now the weaker reading:

(M/E*) Ceteris paribus, if Φ-ing is necessary towards Ψ-ing, then in order to Ψ, one <u>*must*</u> Φ.

In the strong reading (M/E), the emphasis is on the 'ought' — because this is what differentiates this reading from its weaker counterpart. This reading is strong in that it incorporates normative considerations: oughts and reasons. Note that this 'ought' is supposed to be the *practical ought as such*: the one telling you what you have most reason to do. I formulated the conclusion of the weak reading (M/E*) in terms of 'must' (instead of 'ought') in order to avoid conflating two distinct ideas: that of *natural* necessity with that of *rational* or *normative* necessity. For the weaker reading is just reporting a natural fact: the necessary causal relation between some necessary means to a supposed end. The weaker reading, it seems obvious to me, should be uncontroversial and as such accepted by everyone as *a* (though not necessarily the only) fundamental principle of practical reasoning. Hence I find it also true that anyone who does not accept (M/E*) — by which I mean either fails to see its truth or respect it in an instance of practical inference — fails to qualify as rational. That is, someone who denies (M/E*) — who denies a (say) necessary relation between means and ends — fails to qualify as rational in a broader sense: she is *insane* in a similar way that the wildly irrational defender of '2+2=5' is. With respect to such a person, we may indeed wonder what to make of her. But note that the unintelligibility of such a person is not restricted to her capacity for practical reasoning — though the latter may supervene on the former. That is, someone who fails to draw basic rational inferences such as

[93] Of course, there may be plenty of ways to hurt John, but for our current purposes, suppose that this is the only way I can do so.

(M/E*) or '2+2=4' may well be unintelligible with respect to practical inferences *too,* but that would be the result of her being unintelligible more generally.

By contrast, it seems far from clear to me that anyone who does not accept the strong reading (M/E) fails to qualify as intelligible or rational *in any sense*. What exactly is unintelligible about the person who is aware that Φ-ing is necessary towards Y-ing, who wants to Y, but who fails to draw the conclusion that she ought (meaning that she has sufficient reason) to Φ? For example, what exactly is unintelligible about Ann who knows that calling John a nasty name is necessary towards her end of hurting him, but who nevertheless thinks that she ought not to call him a nasty name? At risk of repetition, Ann would indeed be unintelligible in some fundamental sense where she to resist the conclusion that were she to Y she would have to Φ, when she knew that Φ-ing is necessary towards Y-ing. My hunch is that the instrumentalist who thinks that instrumental reasons are conceptually basic, erroneously arrives at this conclusion by (first) failing to distinguish the above two readings, and (second) by taking the strong reading (M/E) as denoting the fundamental practical inference.

Finally, we now have the resources to dispute the famous tortoise analogy (Carroll 1895). In the domain of thinking about theoretical reasons, (we may well accept that) the inference rule modus ponens really is fundamental, and so anyone who does not accept it fails to qualify as intelligible — at least in that domain. As a result of this, we may be at loss and sensibly wonder what could count as a (theoretical) reason for anyone who does not see a reason to believe a conclusion validated by modus ponens. But in the domain of thinking about practical reasons, it is far from clear (if not utterly implausible) that the inference rule (M/E) is fundamental to begin with. Thus we are not necessarily justified in thinking that anyone who does not accept it as such fails to qualify as intelligible in that domain. Relatedly, we must grant as a major possibility that anyone who does not accept it as such may well be competent with respect to our concept of a practical reason.

At this point, you may sensible wondering what *is* the fundamental practical inference, after all. I submit that the only fundamental practical inference there is is the following:

If you are practically required to Φ, then you ought to Φ.

What determines what you are practically required to do is an interaction of all relevant practical considerations at play. I realise that this may seem a tautology, and as such not very informative in the way of a substantive principle, but it seems the only valid conclusion that we can draw *conceptually*: by means of conducting our intuitions on practical reasons.

6.5 Conclusion

In fairness to the instrumentalist of this chapter, her task was always going to be difficult. For her argument against the existence of moral reasons was always going to be hinging on an alleged link between normative reasons, on the one hand, and rationality, on the other. By contrast, her opponent — the non-instrumentalist — could remain neutral on the alleged link, maintaining with some modesty that reasons may well be conceptually distinct from rationality (e.g. I may have a normative reason to

help you, though in *some* sense it would be rational if I do not do so) (cf. Scanlon 1998). Moreover, in order to make her argument go through, the instrumentalist had to insist that reasons are conceptually connected to a *particular conception* (namely, the instrumentalist conception) of the slippery notion of 'rationality'. But showing this, I may have hopefully suggested, is far from easy. The upshot is that the crucial claim in the error theorist's armory, that all normative reasons are instrumenal, remains unsupported. And if we cannot rule out the possibility of non-instrumental normative reasons, why think that moral norms — that have little to do with the furtherance of our ends — do not invariably supply agents with sufficient normative reasons?

7. A Constitutivist Argument

7.1 What Is Constitutivism, and Why Are We Interested?

What is constitutivism?

Constitutivism is as an attempt to derive philosophical insights from what is constitutive of agency. In particular, constitutivists are interested in finding out what agents *as such* ought to do. The attempt is motivated by an intuitive idea: in order to know what chess players as such ought to do, we need to understand what chess is, what makes up a correct chess move, what makes up a successful chess game, and so on. (And of course the same goes for other activities, e.g. essay-writing.) In general, in order to know what someone who engages in an activity *X* ought to do qua engaging in *X*, we need to understand the nature of the activity (Enoch 2006: 170). The attempt is premised on a purported analogy between activities such as essay-writing and chess-playing, on the one hand, and the activity most characteristic of agents, *action*, on the other. More carefully, the analogy is between: agents and chess-players; agency and chess-playing; action and chess. Thus in order to know what an agent ought to do *qua* agent, all we need to do is understand what action or agency consists in — or so the idea goes.

Constitutivism operates on two levels: the normative and the metanormative (Enoch 2006: 170). On the normative level, constitutivists attempt to establish a connection between what is constitutive of agency, on the one hand, and norms: what agents as such ought to do, on the other. Given that the norms in question are thought to be grounded on what is constitutive of action, they can be said to be constitutive norms (more on this notion shortly). The important thing for now is that this alleged basis of the norms in question, which qualifies them from mere (or *generic*) norms to constitutive norms (see 7.2.1), is exploited on the metanormative level. On this level, the constitutivist attempts to show that normative conclusions pertaining to what agents as such ought to do are based on as solid a ground as normative conclusions pertaining to what essayist as such ought to do: the very standards that are constitutive of the activity in question. If successful on the metanormative level, constitutivism may be able to sustain a robust normative conclusion.

Why are we interested?

Recall that the question that this thesis focuses upon relates to the possibility of moral reasons. If there are moral reasons, moral norms must invariably supply agents with sufficient normative reasons for action. It must be true, that is, that the dedicated immoralist who has no concern whatever for the wellbeing of others, but only ends served by wreaking misery to them, has nevertheless adequate reason to refrain from hurting them. What is more, for there to be *moral* reasons, that is, for moral norms to invariably supply agents with reasons, it must be true that dedicated immoralists always have such reasons *and* in virtue of a moral norm that obtains — i.e. not in virtue of any other fact about them or their situation. The error theorist of this context denies that dedicated immoralists have such reasons. In particular, our interest lies in resisting the error theorist's claim that moral norms do *not* invariably

supply agents with reasons. Her argument hinges crucially on a substantive premise: that all reasons — *pro tanto* or *sufficient* — are instrumental.

As I explain in *1.1.2*, in being *normative*, a sufficient normative reason contrasts with motivating and explanatory reasons: a reason *for which* someone does something, and a reason *why* someone does something, respectively. A normative reason for action is a reason there *is* for doing something: a consideration that justifies or supports a prospective course of action. A sufficient normative reason is one that outweighs all other relevant considerations that are at play, thereby determining what the justified course of action is: the one that is most reason supported. In being sufficient, it contrasts with *pro tanto* normative reasons: a consideration that justifies an action but which may be overruled by better ones.

The error theorist's claim that there are no moral reasons hinges on the premise that all reasons (pro tanto or sufficient) are instrumental. This is partly because of a presupposition defended in chapters 2 and 3 of this thesis relating to the content of moral requirements. A moral requirement need not, and very often does not, have anything to do with the satisfaction of the agent's (to whom it applies) ends. Moral norms are thought to express such requirements. But if all reasons are instrumental, it cannot be true that moral norms invariably supply agents with reasons. It cannot even be true that moral norms *sometimes* supply agents with reasons. For if agents have reasons *invariably* in virtue of instrumental considerations (ones pertaining to the satisfaction of their ends), it follows that it is never true that agents have reasons (solely) in virtue of a moral norm that obtains. In the previous three chapters, we sought support for the crucial claim that all reasons are instrumental. This is also the task of this chapter.

We are interested in constitutivism because it may have the resources to vindicate certain types of reasons. It can do so by understanding the nature of action; what makes up a correct action; what makes up a successful action, and then based on these findings, infer what agents as such ought to do. Agents as such, they may be able to say, ought to be aiming at performing a positive net balance of such-and-such actions. The error theorist of this context, who embraced instrumentalism, and now embraces constitutivism, is interested in a particular conclusion: that agents as such ought to be aiming at performing a positive net balance of actions that favour the satisfaction of their ends. If this much can be shown, then in a natural sense she may be able to infer that agents as such have reason to be aiming at performing such actions (cf. Katsafanas 2013: 1).

And if agents *as such* have reason to be aiming at performing instrumental actions, and that no other type of action is in this way linked to the nature of agency, it may follow that instrumental reasons (reasons for agents to perform actions that are instrumental to their ends) are the only types of reasons that agents as such have. In short: that, as the error theorist's crucial claim has it, all (genuine) reasons are instrumental.

I am so interested in the question of whether all reasons are instrumental that I am here willing to make important concessions to the error theorist (converted constitutivist). I will assume that action has a constitutive aim (to be explained), and importantly, I will assume a particular version of constitutivism on which instrumental

actions, and by extension reasons, are truly thought to be linked to the nature of agency. This will enable us to examine whether the conclusion that all reasons are instrumental can be said to follow. In other words, I will assume the success of the normative part of a particular constitutive account, importantly one that suits the purposes of the error theorist, with an eye for examining the plausibility of its metanormative counterpart. My aim is to argue that, despite the constitutivist's efforts and resources, the conclusion that all reasons are instrumental does not follow. Before we reach that point, we need to gain a better understanding of the constitutivist project.

7.2 Constitutive Resources

The aim of this section is to explicate core constitutive ideas. These form the backbone of any constitutive argument. The first is the notion of a constitutive norm, and the second is one particular way that constitutivists may understand the nature of an activity — and importantly, agency.

7.2.1 Constitutive Norms

Suppose you are playing chess, and that at some point I tell you that you must touch your nose while playing any chess move. You quite reasonably wonder why you should do that, and I cite a norm:[94]

> *Touch Nose Move*: You must touch your nose while playing any chess move.

I playfully elaborate: *Touch Nose Move* is a norm about chess just as the norm that tells you to move your rook in either a straight line, horizontally or vertically is:

> *Rook Move*: You must move your rook in either a straight line, horizontally or vertically.

As I really want to annoy you, I carry on with this a bit more: A norm is *merely* an idea about how one should behave — and there is no interesting, *inherent* difference between *Rook Move,* that you so willingly follow, and *Touch Nose Move*, that you are so keen to neglect. If this is so, what explains the difference in the way you are treating them?

You are quick to point out that the answer is: *chess*. The reason you are willing to follow *Rook Move*, but not *Touch Nose Move*, is because you are presently interested in just playing chess. And *Rook Move is* chess — or at least obeying it is an indispensable part of playing chess. At this point, I remind you that in a sense *Touch Nose Move* is also a norm *about* chess. This is clearly true, as it makes direct reference to chess, and one can obey or disobey it only when playing chess.

You are baffled for a moment, and then you point out: though it may be true that the *Touch Nose Move* is a norm *about* chess, it is still not a norm *of* chess. The

[94] (McPherson, forthcoming: 9)

former is merely an *addition to* chess,[95] the latter is an inherent part of the activity. *Rook Move* defines or even constitutes chess; it flows from the very understanding of it. To see this, we may ask: What is *chess*? A game in which two players aim at checkmating their opponent through a series of *regular* moves, one of which is expressed by *Rook Move*. That is, a player cannot win a game of chess by (say) arbitrarily moving her pieces in a favourable way. That would be an *irregular* move, and it seems intuitive to say that such a move would not count as a *chess* move to begin with. Rather, in order for a move to count as a chess move it needs to be in alignment with norms such as *Rook Move*: ones that make up *regular chess* moves. It seems that, in a sense, *Rook Move is* chess — where 'is' seems to be about *constitution* (see 7.2.2).

We can tie up the above observations by distinguishing between *generic* and *constitutive* norms. Let us focus on chess, and call norms such as *Rook Move constitutive* (for chess), and norms such as *Touch Nose Move generic*. What both kinds of norms have in common is that they are *norms*: ideas about how one should behave. Where they differ is in their relation to the activity for which they are norms. That is to say: there is no interesting, intrinsic difference between a constitutive (*Rook Move*) and a generic norm (*Touch Nose move*) in and of themselves, though there is one concerning how each relates to the activity for which they are norms. A constitutive norm is a norm that in some sense defines or constitutes the activity for which it is a norm; it flows from the very understanding of the activity for which it is a norm.

Constitutive norms have authority that generic norms lack. In particular, a constitutive norm for an activity has authority over anyone who *willingly* engages *(or clearheadedly intends to engage)*[96] with that activity. For there is a natural sense in which anyone who willingly engages with an activity *ought* to follow the constitutive norms of that activity. To see this, suppose that Mary is playing chess, and we point out to her that she cannot move her rook diagonally. She asks 'Why?', and we cite *Rook Move*. Mary shrugs, and we seem to be at loss. She is either not interested in playing chess, or she does not know what chess is — we think. For anyone playing chess, must be concerned with or acknowledge the authority of *Rook Move*. Following McPherson, 'it would be puzzling for me to say that I am playing chess, but I simply don't care about playing according to its rules' (McPherson forthcoming: 10). By contrast, things would be different if we advised Mary to touch her nose while playing chess, and upon her asking 'Why?', we cited *Touch Nose Move*. For suppose she shrugs: why should she not? As someone who plays chess, Mary need not be concerned with *Touch Nose Move:* the norm is nothing to the activity that she engages in. This is why it makes perfect sense for Mary to say that she does not care about *Touch Nose Move* — 'without [causing] any confusion' (ibid).

Note that the point above is *not* that it does not make any sense for a chess player to shrug off some rules of chess. Rather, the intuitive idea is that it does not make any sense for a chess player *who clearheadedly intends to continue engaging with chess* to do so. For it makes perfect sense for one to cease to bother with chess and its rules upon (say) having become increasingly frustrated after a series of lost games. The present clarification brings us to the following intuitive result concerning

[95] Korsgaard sensibly calls norms that can be thought of as additions to already existing frameworks *external* (Korsgaard 2009: 32).
[96] (McPherson forthocoming: 10)

the authority of constitutive norms: insofar as someone willingly engages with an activity (and clearheadedly intends to keep doing so), one ought to follow the constitutive norms of that activity.

7.2.2 Constitutive Aims

> If you understand the nature of a game, such as chess, you thereby understand a host of normative claims that regulate chess players. For example, part of what it is to play chess is to *aim* at checkmating your opponent (Katsafanas 2013: 1).

The reason you managed to come up with good answers in the previous section is because you were concentrating on the *nature* of the relevant activity, chess. That is, your explanation of the intuitive contrast between *Rook Move* and *Touch Nose Move* was accurate because it was informed by considerations pertaining to the very nature of that activity. *Rook Move* is chess, you said, whereas *Touch Nose Move* is not, because you know what chess *is*! Constitutivists have the resources to go a bit further and tell us something more informative about the nature of an activity — thereby demonstrating a finer understanding of it. In particular, they can talk of a structural feature of an activity, an *aim* that it can be thought of as having.[97]

In the chess example, we may understand chess as an activity which *aims* at check-mating one's opponents through a series of regular moves. Consistent with the understanding of the constitutive relation given earlier, we may say: *aiming* at check-mating your opponent through a series of regular moves *is* chess. That is, the aim of check-mating one's opponents through a series of regular moves is *constitutive* of chess. It is constitutive in that it must be present in order for a series of movements to count as an episode of chess-playing (ibid). That is to say: you fail to engage in chess-playing just in case you fail to have this aim. On this understanding, a constitutive norm differs from a generic norm in that it is derived from a certain structural feature of the activity for which it is a norm (Katsafanas 2013: 1). The structural feature in question is here understood as an *aim* that in a certain sense *constitutes* the activity.[98]

At this point, it seems sensible to ask: What, if any, is the connection between the *attainment* of *X* (that which an activity must aim at), on the one hand, and actually engaging in the activity, on the other? In answering this question, it will be enough to distinguish between *correct* engagement and *successful* engagement in an activity. Via an example, one engages *correctly* in chess, and thereby her behaviour qualifies as chess-playing, in case one aims at check-mating. By contrast, one engages *successfully* in chess, in case one not only aims at check-mating but also manages to check-mate

[97] Aims are not the only structural features that constitutivists talk about. Others, notably Korsgaard, have suggested that the notion of a *function* may be just as useful in understanding the structural features of an activity (Korsgaard 2009). Though for our purposes it is not important taking sides on which particular feature best describes the nature of an activity, I will here stick with aims. For a discussion that does justice to both aims and functions, see (Silverstein 2016).

[98] I will here stick with constitutive aims, though others have suggested that the notion of a constitutive function may be just as useful in understanding the structural features of various activity. See (Korsgaard 2009).

(cf. McPherson forthcoming: 9). For the very possibility of *bad* chess players (those who play but lose) seems to presuppose something like this distinction. In other words, it seems perfectly sensibly to say that one can engage in a game of chess without ever attaining the (constitutive) aim of check-mating her opponents: all that is required is that she aims at doing just that.

We now have the resources to formulate what it is for action to have an aim: What makes a bit of behaviour an action is its being regulated or driven by an aim (whether or not the agent is consciously aware of that aim). Since being thus regulated or driven by an aim is what all actions have in common — the aim in question must also be *constitutive* of action. The basic structure of any such explanation of action and agency in terms of a constitutive aim is the following: Action is activity driven by the aim to *X*, and agency is one's capacity to engage in action thus construed (Silverstein 2016: 17).

7.2.3 Action and Constitution

Despite their pure philosophical interest in what it is for an activity in general to have a constitutive norm, what constitutivists are really interested in is whether action — the activity most characteristic of agents — has any constitutive norms. For any constitutivist project, it is thus imperative to have a sound conception of action. To do this, they may consult the philosophy of action. The good news is, we here do not need to adjudicate between competing conceptions of action. For the purposes of this chapter, we only need a minimal — somewhat uncontroversial — conception of action. Recall that my interest here lies specifically in the metanormative part of constitutivism, and for good reason. For the purposes of this thesis, I am only interested in showing that even if we grant that instrumental norms or reasons enjoy a privileged connection to the nature of agency, it still does not follow that all normative reasons are instrumental. In other words, that even if the normative part of such argument is successful, its metanormative part need not be so. In order to do this, I only need to present a conception of action that allows us to make the former claim, and assume that it is more or less accurate.

Action here can be contrasted with mere *activity* or *behaviour*. The distinction seems highly intuitive, and to see this, let us begin with an example. Suppose you are sitting on a chair, and an automaton is sitting on a chair next to you, and that there are two glasses of water on the table. At a certain point, both you and the robot begin moving towards the table until each grabs one glass. To the bare eye, both activities seem indistinguishable — but intuitively, there has to be a difference. For one thing, you are a being with intentions and as such you may well be interpreted as acting on them, whereas the robot is merely an automation programmed to perform certain movements. To do justice to such intuitive differences, we may reserve the term *action* to refer to the activity of beings like yourself, and use the unqualified term *activity* to refer to the doings of things like the robot. In the context of constitutivism, action is thought to be exclusive to *agents*: beings with conscious awareness of their activity, who have control over it and shape it through their intentions,[99] whereas mere activity

[99] In saying this, I am subscribing to a standard conception of agency, such as the one spelled out in: Schlosser, Markus, 'Agency', The Stanford Encyclopedia of Philosophy (Fall 2015

is thought to be mostly associated with non-agents: things like automations or the (philosophical) zombie who lack the capacity for such subtle behaviour.

Depending on the particular element that is deemed constitutive of agency, we get different versions of constitutivism. Korsgaard (2009) has it that the constitutive function of action is self-constitution, and hence that agents as such ought to self-constitute themselves. Velleman (1989) once thought that the constitutive aim of agency was self-knowledge, and hence that as an agent, you ought to aim at knowing thyself. Katsafanas, interestingly enough, identifies two constitutive aims of agency, *agential activity* and *power*, and holds that agents as such ought to aim at both.

In making the necessary and intuitive observations on constitutive norms and aims, we only needed to employ a loose — somewhat natural — notion of constitution. At this point, it would do no harm to be a bit clearer on that.[100] For the purposes of this chapter, I need only go with a standard conception of constitution on which constitution is understood not as identity but rather as an asymmetric and dependence relation: to say that *X* is constitutive of *Y* is to say that *X* is a precondition of *Y*, such that by losing *X* you necessarily lose *Y*, though not vice versa.[101] To put the relation in reverse, we may say that *Y consists in X*: is *not something over and above X* (Wiggins 1968: 91). In our case, to say that *X* (where *X* is an aim, function or whatever the case may be) is constitutive of agency, is to say that *X* is a precondition of agency. For example, if *X* is deemed to be an aim, to say that aiming at *X* is constitutive of agency is to say that aiming at *X* is a precondition for being an agent. As the discussion progresses, I will note a particular difficulty for constitutivists which relates to the conception of constitution. But as my criticism of the view will not hinge on such difficulties, I shall not take sides.

7.3 The Analogy with Chess, Revisited

Constitutivists, who ultimately want to say that agents as such ought to perform such-and-such actions, are guided by the analogy with chess. Just as norms of chess apply to chess players as such, the idea goes, some norms of practical reason are thought to apply to agents as such. The end result — they hope — may be a metanormative vindication of the latter group of norms which is modelled on the analogy with chess, and which is not available for any other set of norms that does not enjoy such a privileged connection to agency. As it transpires, constitutivists hold dear to an analogy between activities such as chess, on the one hand, and agency or action, on the other. They want to say that: just as chess has a constitutive aim from which we can extract constitutive norms, so does agency. But in all this, they must note where the analogy is likely to break down. Here, I mention two potential disanalogies between chess and action. The first concerns the possibility that action might not have a single aim, and the second concerns the intuitive realisation that unlike chess, action may be regarded as inescapable.

Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2015/entries/agency/>

[100] For a more detailed explication, see https://plato.stanford.edu/entries/material-constitution/

[101] For a finer piece of explication, see Wasserman, Ryan, 'Material Constitution', The Stanford Encyclopedia of Philosophy (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/material-constitution/>.

The first concern

Let us begin by noting that it is true that *chess* has a constitutive aim: something that all chess-acts have in common: to check-mate one's opponent (through a series of regular moves). By contrast, and even if it is true that all actions aim at *something*, it need not be true that there is *a single aim* that all actions have in common. Compare: it may be true that each and every person on earth loves *someone*, but it need not be true that there is *a single person* loved by everyone. Not everyone loves Peter Andre, for example.

The above observation is important insofar as constitutivists want to derive a non-trivial normative claim from what is constitutive of agency. For if it turns out that aiming at *something* is what constitutes action, and that there is no restriction as to what 'something' can be, it will not be possible to extract an *actual* or *particular* constitutive norm from this. To see this, suppose that it is true that all actions aim at something, but that something can be anything from self-constitution, self-knowledge, or the fulfilment of what one takes to be her ends, among other things. From this, we may be only able to infer the following constitutive norm: agents as such ought to aim at either *X*, *Y*, *Z*, and so on. This is another way of saying: there is *no single thing* that agents as such ought to aim at, as long as they aim at *something*. This sounds more like a declaration that as far as practical reason is concerned, *anything goes,* than an actual norm of practical reason.

It is thus an additional challenge for any constitutivist theory that aspires to derive a substantial normative claim to convince us that all genuine actions share a *single* aim, in the same sense that all chess-acts do (Katsafanas 2013: 2). Yet many are optimistic that agency can be thought of as having an aim.[102] I shall herein take it for granted that agency does have an aim, as my interest lies not in the plausibility of this claim, but rather in the whether a constitutive account that surpasses this challenge can derive robust normative conclusions from what is constitutive of agency.

The second concern

The second point that the analogy with chess breaks down is far less controversial. Plausibly, we may regard chess an escapable activity: one may opt out from playing chess at any given time (despite of course the protestations of her opponent who is probably winning). And as mentioned already, it makes perfect sense for a chess player to disobey the constitutive norms of chess insofar as she chooses to end her participation in any given game of chess. Unlike chess, however, agency seems to be an inescapable activity — or at least not 'as escapable' as chess: agents cannot opt out from it — at least not without suffering severe costs to their personhood. It is not our present concern to examine what the implications of this disanalogy are for the constitutivist's aspirations. For now, we shall only note that the present disanalogy *may* enable the constitutivist to provide a metanormative vindication of some norms of practical reasons which is in a certain respect more robust than the metanormative

[102] Notable examples include constitutivists: Velleman (2000), Katsafanas (2013), but also their critics: Enoch (2006), Silverstein (2016).

vindication of chess norms: namely, it may not be possible to shrug it off by escaping the activity for which it is a vindication

7.4 Towards our Constitutive Argument

We can now begin building our argument: one which will be intended to address the question of whether all reasons are instrumental. Effectively, we shall develop a constitutive argument for the claim that the norms of instrumental rationality are constitutive of agency and thereby reason supported: agents as such have reason to follow them. In 7.4.1, I sketch a parallel argument about believing, for reasons to be shortly explained. In 7.4.2, I introduce an account of the constitutive aim of agency on which action aims at the satisfaction of some of the agent's ends. This is effectively a crucial premise for the constitutive argument that follows in 7.4.3 on which agents as such ought to aim at the satisfaction of some of their ends. As I am interested in the metanormative implications of the constitutive argument to be considered, rather than the conception of agency on which it is premised, I shall merely assume the account of 7.4.2.[103]

7.4.1 A Parallel with Believing

In his seminal paper *On the Hypothetical and Non-Hypothetical*, Peter Railton sketches an argument which purports to vindicate the norms governing belief-formation, and as such it makes use of a premise on what may be constitutive of *believing*.[104] For our purposes, this is an important sketch, as it provides a relevant example which is in crucial respects similar to the constitutive account that we are after; both accounts appeal to what is constitutive of agency with the hope of inferring normative principles governing believing, in the one case, and acting, in the other. I find it useful to consider this argument as it illustrates how a constitutivist argument may go using resources from the theoretical domain that may be considered by some as forming more solid a ground for such an argument.

Railton is interested in a constitutive argument for the conclusion that 'two agents in the same epistemic situation (same evidence, same background beliefs) would have the same reasons for believing any given proposition, regardless of possible differences in their personal goals' (Railton 1997: 53). In other words, he is after a metanormative vindication of some (theoretical) reasons or requirements. The particular requirement that he sets out to justify is this: you ought to increase your belief in *P* as evidence for *P* increases. As expected, he hopes to premise the justification on what is constitutive of believing — and in particular, what is the constitutive *aim* of believing. Moreover, given that believing is thought to be a precondition of agency, what is constitutive of believing must also be constitutive of agency (58-59).

So let us ask: What must be true of someone in order to *believe* anything? Granted the hypothesis that believing aims at something *X,* and that this *X* is

[103] Though I do think that the account of the constitutive aim of agency introduced here is minimal enough to be independently plausible, yet substantial enough to yield a constitutive aim (Katsafanas 2013: 2).
[104] For a similar attempt, see (Velleman 2000: 183-189).

constitutive of believing, what must someone be aiming at (or for) in order for her attitude to qualify as believing? The *prima facie* intuitive proposal that Railton considers is that the constitutive aim *X* of believing is *truth*:

> [A] belief that h 'aims at' the truth of h. A belief that h necessarily 'misses its target' when h is false (56).

That is, for a state to qualify as belief it must be necessarily regulated by the aim to attain truth. When it fails to attain truth, it ends up being a false belief: a *bad* belief, but a belief *nevertheless*. Note that on the present understanding of a constitutive aim, the believer need not have truth as *her* aim; our believer may be aspiring to believe only falsehoods. Railton correctly acknowledges that this possibility is consistent with the understanding of believing that he proposes. As he eloquently puts it:

> None of this argues against the possibility of belief that is … unresponsive to evidence or truth. One could … have some success in coming to believe certain convenient falsehoods through a suitable programme of self-imposed indoctrination. What the argument above purports to show is not the impossibility of such a programme but a *design constraint* upon it: if one is to succeed, one must somehow contrive to veil the programme's true nature from oneself (57).

That is, there is a sense in which you can aim at believing a falsehood. For example, you may choose to take a pill which would make you whole-heartedly believe that you performed live at Woodstock in ’69. You may take measures to ensure that your newly formed beliefs remain intact: e.g. entrust professionals to insert you into the original live footage. But in order to succeed at believing your falsehoods, and maintain the relevant beliefs, you must somehow conceal the fact that they are falsehoods. You must remember that your attitude of believing cannot help but be guided by what you take to be evidence: indicators of truth. You may be aiming at believing falsehoods, but your attitude of believing necessarily aims at believing the truth. For this is what believing is.

In my view, everything said thus far sounds sensible as an account of what is constitutive of believing. But at the same time, and based on the rest of what Railton says, it is not clear to me how the conclusion that an agent as such ought to (or has reason to) believe what is true or well-supported by evidence can be said to follow. Granted: it may well be true that believing necessarily aims at truth, and part of what ‘aiming at truth’ means is that when one is *fully aware* that *P* is true, one cannot be also indifferent to the truth of *P* in one’s belief-formation. As Railton puts it: one must ‘accord some deference to what he takes to be truth and evidence’ in his belief-formation (57). But whence does it follow that a believer as such ought to (in a normatively robust sense) believe the truth: believe that *P* when he is aware that *P* is true or well-supported by the evidence? It may well be impossible to do so — given our constitutive understanding of believing — but impossibility does not seem to entail normativity (in 7.5.2 I elaborate on this idea) (cf. Enoch 2006: 187-190).

It seems that, for all Railton has said, the 'deference' that a believer as such must accord to truth may just be similar (if not identical) to the respect that a prisoner must show to his captors: acknowledge that it is impossible for him to escape, without having to embrace the fact that he is thus constrained. Say, he may grant them that they are powerful, but he may still be in a deep sense justified to burn down their prison and escape. By analogy, it may be impossible for an agent to form any beliefs without according some deference to what she takes to be truth, but this does not entail that the agent has to embrace the truth. As a believer, the agent must embrace the truth, but as an agent, she need not do so. In fairness to Railton, he admits that 'not very much normative epistemology can be wrung from this', as his primary aim is to establish *prima facie* grounds for optimism that such a justification of certain epistemic requirements can be linked to an account of what is constitutive of believing (59). In what follows, I proceed to a particular constitutive argument which aims at establishing a vindication of the requirements and reasons of instrumental rationality. Having done that, we may revisit the present worries — for as I shall show, they seem to apply to that argument as well.

7.4.2 Our Conception of Action: The Constitutive Aim of Agency

A constitutivist strategy as such is open to a defender of *any* set of norms (or alleged requirements) purporting to guide action. For example, it may be possible to argue on the basis of what is constitutive of action, that instrumental, non-instrumental (or *what-have-you*) norms are justified: supported by reasons. Moreover, the identification of particular aims as being constitutive of action may well favour different agendas. That is, if action is thought to be aiming at *X*, where thinking that action aims at *X* is more likely to favour the justification of alleged moral requirements than thinking that action aims at *Y*, then the project of vindicating moral requirements via a constitutivist strategy becomes more plausible.

Railton is guided by the parallel with believing when he asks whether there is anything in the realm of action which plays the role that truth plays in the realm of belief (62). He discusses two basic dogmas in the history of constitutivist strategies: *High Brow* and *Low Brow*. According to the former, agents act by putting whatever they take to be good in 'a positive evaluative light, and deem it choice-worthy' (63). In other words, action aims at (the attainment of) the *good*. 'Good' is here understood as a generic term covering various senses of what can be good: moral, aesthetic being among them. Depending on the particular interpretation of good, particular constitutive arguments may follow.

But as Railton rightly notes, despite its somewhat generic character, *High Brow* is non-trivial (64). For it puts forward a substantive claim regarding the nature of agency, one that — crucially — seems unwarranted. Recall that we are here seeking a constitutive feature of action as uncontroversial upon reflection as the connection between belief and truth, and hence the question is: Is the aim of attaining the good (in some acknowledged sense) thus uncontroversial? For reasons to be shortly made explained, I think the answer is no. That is to say that this account of what is constitutive of agency is false. If I am right, this is good news for the instrumentalist who seeks to ground a metanormative justification of the norms of instrumental rationality on what is constitutive of agency. For the instrumentalist's hope here must be to locate what is constitutive of agency in something congenial to the norms that

she hopes to defend, namely the idea that agents ought to pursue the satisfaction of their ends. And what is good (even in this somewhat generic sense) seems to be independent from what agents may wish to adopt as their ends; e.g. I may judge duck hunting to be in some sense bad, but I may still have it as an end to go duck hunting.

The instrumentalist can build a constitutive argument by following closely on *Low Brow*: the view that agents as such necessarily possess and act on ends (65). This is the view that action aims at (the attainment of) acknowledged *ends*. Here, an *acknowledged end* does not have to be something that the agent (for whom it is an end) judges to be in any sense good:

> The tenability of any Low Brow position therefore depends upon the possibility of distinguishing *the possession of ends* from *the making of judgements that certain ends are good* (65).

That is, if the above distinction can be made, then the current proposals gets rid of the notion that agents as such must aim at the good (however that is to be defined). And as already suggested, this distinction is plausible; e.g. one may have the end of going duck hunting despite knowing that shooting innocent creatures for recreation is neither morally good nor good in any other specified sense. We thus seem to have in our hands an account of what is constitutive of action which seems generic enough and could perhaps as a result be considered trivially true. Importantly, the account thus sketched sounds a lot like the instrumentalist who wants to vindicate the norms of instrumental rationality (and the supremacy of instrumental reasons) via a constitutivist strategy would want it.

Indeed, I want to follow Railton in suggesting that the above proposal is indeed highly plausible. The idea is: if you do not aim at the attainment of an end (or goal) of yours, you fail to engage in action; thus you either end up with inertia or behaviour that is essentially a reflex of some sort. The proposal may sound crude upon first reading — but only if, I submit, *aiming at the attainment of an end of yours* is read in a robust, serious enough sense. For if the idea is that action requires the pursuance of anything like a *premeditated* or *well-calculated* goal, then it is hopeless. For example, on my way to the fridge I might notice a slice of pizza that is on the table, and without thinking much about it, I might go for it without really thinking about it. Surely, this seems to be an instance of action despite it being the case that *eating the slice of pizza* cannot be considered 'a goal of mine' in the same sense that *finishing my PhD* can be. Rather, we must understand the *end* in 'aiming at the attainment of one's ends' in a more relaxed way.

Return to the above example, and let us pause at the point when I notice the slice of pizza. There it is me — in a current state of inertia — and a slice of pizza that has now drawn all my attention. A first question here is: Why have I stopped, to begin with? A natural answer is: because upon noticing the slice of pizza, and given some other things that must be true of me (e.g. being hungry), I either formed the intention of going for the pizza, or at least I am really thinking about doing so. Now, as we happen to know the rest of the story, between me and the slice lies a prospective action: that of me moving towards it. For the purposes of our thought experiment, we may say that the action terminates at the point when I grab the slice of pizza. The question now is: What must be true of me or my behaviour in order for the latter to qualify as action

— as opposed to a mere automatic (or robotic) pattern of behaviour? We already know that I have the *intention* to go for it — what can in general be considered the very last step before genuine action. Intuitively, what connects the two points: my forming the intention for the slice, on the one hand, and my grabbing it, on the other, is a behaviour which is in a deep sense *driven* by that intention.

Now, if we interpret the 'forming of an intention to *X*' as 'adopting *X* as my end (in some sense)', we seem to be on the right path towards accepting the proposed Low Brow account of what is constitutive of action and agency. And I believe that the explanation is entirely generalisable: for an instance of behaviour to be an action, it must aim at the fulfilment of an intention (end), and for an agent to be able to engage in such behaviour, she must first form an intention (adopt an end) and then let it regulate her behaviour.

To sum up: I have suggested that *Low Brow* is highly plausible as a generic account of what is constitutive of action. Moreover, for the sake of the argument, I want to grant that the account is true. Recall that our sole aim in this chapter is to examine whether a constitutive argument premised on something like the aforementioned account can lend support to a metanormative vindication of instrumental reasons. Specifically, and on the supposition that aiming at the attainment of one's ends is constitutive of agency, we shall examine whether it follows that agents as such only have reason to follow the norms of instrumental rationality.

7.4.3 A Constitutivist Argument for the Supremacy of Instrumental Reasons

Our constitutive argument for the supremacy of instrumental reasons — the view that all normative reasons are instrumental — is quite simple.[105]

> P1 Aiming at the attainment of one's ends is the single aim that all genuine actions have in common.
> C Agents as such ought to aim at the attainment of their ends.
> C* Agents as such have only normative reason to aim at the attainment of their ends.
> C** All normative reasons for agents to perform actions are instrumental.

There is a certain natural sense that C follows from P1. If aiming at the attainment of one's ends is the *single* aim that all genuine action has in common, such that no behaviour counts as action unless it has that aim, then one must have that aim insofar as one is interested in being an agent. Put simply: the more often you aim at the attainment of your ends, the more of an agent you are. The following may be regarded as a constitutive norm of agency:

> IR (expressed by C) One ought to aim at the attainment of one's ends

[105] I take it that the argument is similar to the one Railton considers (1997: 66-69).

Both conclusions, C* and C**, make the *seemingly* uncontroversial inference from *oughts* to *normative reasons*. If agents as such ought to aim at the attainment of their ends, it seems to follow that agents have reason to do so. Constitutivists are keen to make such inferences. Consider the following passage:

> This aim generates reasons for action: if the player sees that she can achieve checkmate by moving her rook to a certain space, then she has a reason to do so (Katsafanas 2013: 1).

Here, I have already granted the truth of P1. Moreover, I shall now grant that *in a certain sense*, C is also true. But I think that the inference from C to the rest of the conclusions is dodgy — to say the least — in virtue of being highly unclear whether the sense of *reason* expressed by C* and C** corresponds to the sense of *ought* expressed by C. In short, my claim will be: considerations pertaining to what agents as such ought to do (if they are to remain agents) have very little to do with considerations pertaining to what normative reason agents as such have. In light of this, I do not think that neither conclusion follows — and that the claim that all reasons are instrumental, which plays a crucial role in the error theorist's argument (see 2.5), remains unsupported. In the following sections, I go through three objections that the above argument may be said to be vulnerable to. The first objection (7.5.1) will be discharged, whereas the other two will be of greater importance. The objection of 7.5.2 disputes the inference from P1 to C, and the objection of 7.5.3 disputes the inference from C to the rest of the conclusions. It is noteworthy that the nature of both latter objections is general: they apply to any constitutive argument, not just this one. They purport to attack the very foundations, that is, of the constitutive project.

7.5 Objections

I have granted that the *Low Brow* conception of agency is not only plausible but also true. In saying this, I have conceded that the instrumentalist who aspires to be a constitutivist is right with respect to a particular discussion: the philosophy-of-action view on what action and agency consist in. The concession disarms one famous objection (7.3.1), though it leaves untouched its more forceful counterparts (7.3.2, 7.3.3).

7.5.1 Agency, Schmagency?

A discussion addressing the weaknesses of a constitutivist argument must make some reference to the famous *agency-schmagency* objection. Let me make this clear from the start: An agency-schmagency objection does not apply to the particular constitutive argument which we are examining.

Let me explain. A standard way of understanding constitutive arguments of the sort we are here interested in is that they attempt to make constitutive *aims* into constitutive *norms* — or at least force a connection between the two. Recall that with respect to chess, it may seem natural to say that the requirement to check-mate your opponent is a constitutive norm of the activity: one fails to engage in chess when showing complete disregard for it. Therefore, the idea goes, a chess player as such ought to check-mate her opponents. As we saw this norm is grounded on what seems

to be the constitutive aim of the activity: to check-mate your opponent. Similarly, our constitutivist seems to attempt to force a similar connection between the constitutive aim (attainment of one's ends) to an alleged constitutive norm of agency: an agent as such ought to aim at the attainment of her ends.

It is at this point that the critic of a constitutivist argument can exploit the analogy with chess. For suppose I am playing chess with my little brother with the aim of teaching him the basics of how to play the game. At some point, I have the chance to check-mate him, but instead of doing this, I intentionally worsen my strategic position by pursuing a bad move. Garry Kasparov, who happens to be sitting in the cafe next to us and has been sneakingly observing our play, says to me: 'You do not know what chess is — are you sure you are a chess player?'. I explain to him my reasons for playing like this, but Kasparov is not convinced: 'This is still not the right way to do it, you are either teaching the boy chess *by playing chess*, or you are not teaching him *chess* at all'. I have to admit he makes a good point, but I do not care. I recall Enoch's famous objection, and say to him:

> Classify my bodily movements and indeed me as you like. Perhaps I cannot be classified as [a chess player] without aiming to [check-mate my opponent]. But why should I be a chess player? … I am perfectly happy being [a schmess player] …[someone] who is very similar to [chess players] but who lacks the aim (constitutive of [chess] but not of [schmess][106] (Enoch 2006: 179).

Likewise with the 'game' of agency: If there is indeed an alternative activity which is in important respects similar to action but lacks the particular aim identified by our constitutivist as constitutive of action, why can't we all be schmagents instead of agents, and engage in schmaction instead of action? This, of course, will have the additional benefit of escaping the norms linked to agency and action that are imposed on us by the constitutivist; for those norms come with membership in the activity of action, and by unsubscribing to the latter, one may be thought to escape their application. Following Katsafanas, the analogy with chess illustrates a general phenomenon:

> [Y]ou can escape the standards of a practice by engaging in a new practice (Katsafanas 2013: 54).

The first thing to note here is that the present objection is *specific* rather than *general*, as it hinges on the particulars of a constitutivist argument. That is, the objection presupposes that the constitutive argument in question fails to give a *true* account of what is constitutive of agency and action. But at least in principle, it is perfectly possible that the account of what is constitutive of action that the constitutivist proposes gets things right in a crucial sense: it is not possible for any activity not fulfilling the constitutive aim identified to even remotely resemble action (as we intuitively understand it) — at least, not to the extent that schmess resembles chess. The upshot here is that it is possible that an account of agency leaves critics

[106] Of course, Enoch talks about agents and schagents, but the objection has perfect application to our analogous example.

with no real alternative: if they do not engage in action as defined by our constitutive aim, they end up losing all that which distinguishes them from zombie-like creatures and vegetables. Following Velleman, they can of course 'leave the game in a final exercise of agency — say, by taking drugs or jumping off a bridge or just dozing off for a while'[107] (Velleman 2004: 291). But they would then lose all their 'acting privileges'.

The second thing to note here is that I am assuming that the account of what is constitutive of agency which features in the argument of the previous section gets things right in the above sense. That is, I am taking it for granted that the constitutive aim identified in the argument pertains to something that really is inescapable[108] (Katsafanas 2013: 55). Given that schmagency and schmaction are not really alternatives (i.e. one cannot engage in them without losing all of what is significant about their counterparts), Enoch and the like-minded cannot resist the argument's conclusion that agents as such ought to aim at the attainment of their ends by saying things like 'Classify my bodily movements as you like…' — in fact, they cannot resist it by saying *anything*: for by failing to be agents, they lose their capacity to think, say or do thing, to begin with. Critics must respond in a different way, and this is the task we now take on. In what follows, I express two genuine worries for constitutivist arguments in general.

Before moving on, there is a final thing to note. Constitutivist arguments appeal to some sort of necessity (or inescapability): If you want to be an agent, you ought to *X* — where *X* is constitutive of agency. But as Railton notes, the necessity had better be not *merely linguistic* (Railton 1997: 69). That is, the constitutivist cannot fix the meaning of 'action' (or 'agency') in an eccentric way such that it refers to an activity that others would not consider action. Rather, the necessity had better be *substantive*: 'action' (or 'agency') must pick out nothing more and nothing less than the activity that is distinctive of agents — those who differ substantively from (say) robots.[109]

---

[107] Also, see (Enoch 2011: 215).

[108] It is noteworthy that I do not do so just for the sake of the argument, but rather because I really believe the account is true. As already mentioned, the account given is *generic* enough to be plausible. As I understand it, accounts of what is constitutive of agency that tend to be vulnerable to an agency-schmagency type of objection are extravagant in that they make questionable suggestions of what might be constitutive of agency - ones that try to read too much into what is involved in agency and action.

[109] Railton worries that constitutivists who take the necessity to be substantive can be accused of *xenophobia* in case they aspire to criticise those who fail to qualify as agents by the proposed definition (71). I do not think he should be worried. If the constitutivist is right in that agents as such must aim at *X* if they want to be agents, then anyone who does not aim at *X* fails to be an agent. With respect to such creatures (e.g. animals, infants, robots, zombies), we need not say that they are *bad* agents - for they are not agents to begin with. We can *dismiss* them on these grounds - we can say that since you are not one of us, we do not expect you to aim at *X*. This is consistent with the observation that we do not speak of animals in normative terms: we do not say, for example, about a lion that 'it ought not to have killed that deer'. But constitutivists can still be interested in what agents - those who are like us - ought to do *qua* agents. And though I do not think constitutivists can succeed in justifying any such normative conclusion based on what is constitutive of agency, the point here is that there is nothing problematic with dismissing non-agents when examining what agents qua agents have reason to do.

7.5.2 Inescapability Does Not Entail Normativity

I thus begin my criticism of the argument with the concession that the 'game' of agency is in an important sense *inescapable*. Granted: the game of agency differs substantively from chess — and hence I — the critic — cannot rely on analogies with chess to resist the *necessity* with which the constitutive aim of agency applies to me — and everyone else who aspires to be an active creature like myself. In short: whether I want it or not, I may well be a prisoner in this game due to my unconditional want for being an agent. At the same time, however, it is unclear how this type of necessity supports *any* claim regarding what agents — prisoners in the game of agency — *ought* to do. In a nutshell: inescapability does not entail normativity[110].

As I understand it, the present objection is meant to attack the inference from the first premise to C:

> P1 Aiming at the attainment of one's ends is the single aim that all genuine actions have in common.
> C Agents as such ought to aim at the attainment of their ends.

The reason the inference seems initially plausible, I think, is because the first premise may be read in terms of a certain type of necessity — let us call it a *constitutive* necessity. Let us rewrite the first premise in terms of that necessity:

> (P_n) Agents as such *must* aim at the attainment of their ends.

The necessity is constitutive in that it tells us what someone or something must be doing in order to succeed at being an agent. It effectively reports what we have taken here for granted, namely that in order for one to be an agent, one must aim at the attainment of her ends. Compare: a piece of bread *must* (necessary condition) have melted cheese on it in order to succeed at being a pizza. But, crucially, the problem is that the necessity specified in C seems to be a different type of necessity: what Enoch eloquently calls — a *normative* necessity[111] (Enoch 2006: 188). The bread must have melted cheese in order to be a pizza, but it is odd to say that the bread *ought* to be a pizza. Consider a plausible instance in which this necessity features: I ought to abstain from eating meat because Catholicism requires me to do so. In a sense, this necessity may be regarded constitutive: If I am to succeed at being a Catholic, I need to do so. But importantly, there seems to be something above and beyond this particular necessity that the judgement calls attention to: namely, that doing so is necessary in

[110] This claim - coming from this side of the dialectic - is somewhat ironic. In chapter 2, we established a sense in which moral requirements are inescapable: they are thought to apply to agents irrespective of their contingent ends. Yet error theorists and instrumentalists alike (my opponents throughout the thesis) do not take the aforementioned inescapability to entail normativity: they resist the leap from the claim that moral requirements are (in a sense) inescapable, to the claim that moral requirements invariable supply agents with sufficient normative reasons. Yet in this chapter, my opponent seems to be claiming that another kind of inescapability - the one suggested here - entails normativity.

[111] Elsewhere, Enoch puts the point rather bluntly: 'For the move from "You inescapably Φ" to "You should Φ" is no better – not even the tiniest little bit – than the move from "You actually Φ" to "You should Φ."' (2011: 216).

order to align with the norms of Catholicism. At risk of repetition: insofar as the norms of Catholicism go, I ought to do so. By contrast, there seems to be something missing from Pn — at least if it is to be taken to call attention to a normative necessity. In particular, in order for the inference from P1 to C to be validated, there needs to be a suppressed premise:

(P_s) If an agent as such *must* aim at *X*, an agent as such *ought* to aim at *X*.

In the absence of something like Ps, P1 tells us nothing normative. It merely specifies a necessary, constitutive connection between two things: being an agent and aiming at *X*. The constitutivist may attempt to motivate Ps by reminding us that we have already accepted that agency is in an important sense inescapable: someone cannot escape it without severe costs to her personhood. What connects the constitutive necessity expressed by the first clause of Ps with the normative necessity expressed by the second, is the assumption that agents cannot escape being agents – or that non-agency is not a genuine option. That is, insofar as the non-negotiable end of remaining an agent goes, you *ought* to aim at *X* (the attainment of your ends).

Still, Ps seems intuitively false — and if this is so, it is unclear how the inference from P1 to C may be validated. Recall the prisoner analogy which was employed earlier to make a similar point. This is someone who remains confined in a maximum security prison, and who cannot escape it (to avoid getting distracted by normative considerations pertaining to morality, suppose that his captors are evil and that he is not serving any deserved time). Granted: the prisoner finds himself in an inescapable situation: he cannot help but be a prisoner. Moreover, as he values living more than non-living, committing suicide is to him not an option. But whence does it follow from all this that he *ought* to be a prisoner – to stay in the prison? That is, to highlight the analogy, whence does it follow that he ought *to aim at staying* in the prison? Quite the contrary, it seems that the prisoner ought to at least try to escape the prison.

Examples such as the above can be multiplied, and they all point to an intuitive observation: inescapability does not entail normativity. The upshot here is not that the intuitive idea conclusively disproves C or the rest of the conclusions, only that the inference to the latter is still in need of validation. For a more conclusive objection to the constitutivist's argument, we turn to the next section.

7.5.3 Normative Arbitrariness

It is Katsafanas' (2013) response to the previous objection that leads us directly to the current objection — which I take to be the more decisive one against constitutivist strategies at large. We have successfully objected to the claim (or at least the implicit suggestion) that inescapability entails normativity, and it is at this point that Katsafanas complains that the critic has 'misstated the constitutivist strategy' (56). He then makes — what is on his view — a crucial distinction:

> The move is not from 'you inescapably Φ' to 'you should Φ'. Rather, the aim-based version of constitutivism moves from 'you inescapably *aim* at Φing' to 'you should Φ'(ibid).

On Katsafanas' view, it is not mere activities — whether they be inescapable or not — that generate reasons, but rather aims. The prisoner may be doomed to stay at the prison, but he is not doomed to be *aiming* at staying. By contrast, agents are doomed to be *aiming* at *X* (the attainment of their ends). That is, the claim is not that being inescapably bound to do something gives you a reason to keep doing it. Rather, the claim is that being inescapably bound to be *aiming* at doing something gives you reason to *attain* it. On Katsafanas' view, aims generate reasons, and inescapability merely ensures that one keeps having those reason to attain the aim: that one cannot escape having those reason. But how does an aim generate an ought in the first place? On Katsafanas' view, aiming at something makes you (or your behaviour) subject to *standards of success*.

(*Success*) If *X* aims at G, then G is *a standard of success* for *X*.

This principle, he thinks, is all the constitutivist needs (60). And the fact that you are inescapably bound to aim at something means that you cannot escape being subject to the standards of success for attaining it, and by extension, having reason to attain it.

In responding, we need to acknowledge the intuitiveness of *Success*. If I aim at something, I may be said to be successful or unsuccessful depending on how well I do in regards to securing it. A standard of success does seem to follow from an aim; e.g. if my aim is to score a goal, whether or not I actually do so can be used as a basis to infer whether my action was a success or failure (relative to that aim).[112] Insofar as success is something that the very act of aiming somehow presupposes, *Success* can be said to be plausible. The problem, however, is that it is not clear how the robust normative conclusion that agents as such ought to (or have reason to) attain *X* (i.e. be successful in their aiming at *X*) follows. It seems that a lot in the constitutivist's argument turns on the success of *Success* to bridge the aforementioned gap. But sensible as *Success* is in and of itself, it is still dubious whether it can get us from aims to normativity: to get us from the claim that one inescapably aims at *X* to the claim that one has reason to aim at *X*. All it can show, it seems, is that when aiming at *X*, one's activity is regulated by a standard of success. In order to bridge the gap, constitutivists need to show that standards of success generate reasons.

From constitutive aims to normativity: the gap that cannot be bridged

Granted: I must aim at *X* if I am to be an agent, and I am inescapably bound to aim at *X* in virtue of the fact that being an agent is not something that I can opt out (at least not without severe costs). And of course, aims do seem to generate standards of success: When I am aiming at something, whether or not I actually manage to hit the target can be given a further, complimentary meaning: hitting the target means success (in some sense), missing the target means failure (in some sense). Thus I may be said to be a successful agent (or successful qua agent) when I attain *X*, and unsuccessful otherwise. If only this could be the end of the story.

[112] Or as Silverstein puts it: 'Criteria of success follow so straightforwardly from aims because "aim" and "success" are - or at least can be - both teleological terms' (2016: 20).

The interesting question now is: what reason do I have to be a successful agent in the first place — i.e. to keep doing the actions that 'hit the target' and hence qualify me as a successful agent? Whence does it follow that one has reason to be successful at aiming at *X*? The pressure is on the constitutivist to answer this question, but I do not see how this could be done.

Considerations about what we have normative reason to do seem to be *above and beyond* considerations about what constitutes successful *agency*. What considerations about normative reasons may be linked with is considerations about (what may be called) *rational* agency. Whereas a successful agent is someone who exemplifies the aim of agency, a successful *rational* agent is one who not only does that, but who is also responsive to her good practical reasons. The standard of success invoked by the constitutivist is meant to apply to our capacity as agents — crucially not to our capacity as rational agents. A certain connection between the two capacities should be obvious. In order to be a good rational agent, one first needs to be an agent — for being good or bad in any sense presupposes that one first has the fundamental capacity for action: to be an agent. Compare: you can be a good or bad president of France, but being so presupposes being a citizen of France. One can be a generally good *agent* without being a generally good *rational* agent, and *vice versa*.

Suppose in any given period of time, John has *X* amount of ends and manages to fulfil all of them. Paul has also *X* amount of ends and fails badly at fulfilling any of them. Whereas John is a good agent, Paul is not. That is, unlike Paul, John engages much more often in activity that exemplifies the virtues of agency: possessing ends, aiming at their fulfilment, and actually fulfilling them. At the same time, however, John's ends are always self-centred as he fails to care about anything beyond his egotistic desires. For example, on his way to the cinema, he notices a drowning child, but fails to adopt the end of saving it: he never aims at the attainment of the end of saving the child, therefore he never fails in this respect. If our constitutivist is right, and all normative reasons are grounded in what is constitutive of agency, it seems to follow that John never fails as a rational agent. For by definition, a rational agent is one that is responsive to her good practical reasons. But this is unintuitive: John is bad as someone who is supposed to be thus responsive.

Now consider Paul who tends to be a lousy agent. Paul notices the drowning child, sets for himself the end of saving it, and quite miraculously manages to attain it. Overall, it seems, Paul still scores lower than John with respect to the 'game' of agency: he hits the target far fewer times. But crucially, from the two, it is only Paul that can be considered successful as a *rational* agent. The moral seems to be this: you can be a bad agent in the closed sense that you fail, on average, to achieve the aim that constitutes you as an agent: a being capable of action, but at the same time, you can be a rational (or virtuous) one: someone capable of action and who is also responsive to her good practical reasons.

On my view, the mistake that our constitutivist makes is similar to the one that I attributed to the reason internalist in chapter 5. The latter seems to conflate psychology with normativity: considerations pertaining to the conditions under which agents are motivated with considerations pertaining to what agents have reasons to do (see 5.6.2). The former seems to conflate agency with normativity: considerations pertaining to the conditions under which agency is possible (or successful) with

considerations pertaining to what agents have reasons to do. And both types of considerations featured in the first clauses, I submit, are *normatively arbitrary* (cf. Enoch 2006). To borrow a couple of phrases, how can nonnormative facts about an aim close normative questions about what we have reason to do, and what do aims — constitutive or otherwise — have to do with reasons? (Silverstein 2016: 18-21).

7.6 Conclusion

In this chapter, we sought support for the error theorist's crucial claim that all normative reasons are instrumental. Vindicating this premise would mean that the error theorist's conclusion that moral norms do not invariably supply agents with reasons is supported. For if moral norms have little to do with the attainment of the agent's ends, and all reasons are instrumental, it cannot be true that agents invariably have reasons to do as moral norms say which obtain in virtue of those moral norms. The constitutivist argument considered here was meant to show that agents as such have only reason to aim at the attainment of their ends, in virtue of the fact that aiming at the attainment of one's ends is constitutive of agency. The argument puts forward an account of what is constitutive of agency that is generic enough to merit approval. But it fails nevertheless because it employs — and relies crucially on — a much dubious premise: that considerations pertaining to what is constitutive of agency can be informative about what agents have reason to do.

8. Standard and Alternative Error Theories

The first part of this chapter is a synopsis of the dialectic which has taken place in this thesis. I begin by stating the thesis question, before I summarise the case of the error theorists that was unfolded in chapters 4-7. The second part sketches an alternative error theory which purports to vindicate the same conclusion as that of the standard version. I mention some potential source of support for the alternative theory, and briefly state my reasons for resisting it. My primary aim in this chapter is speculative: though the alternative version purports to lend support to the very same conclusion that this thesis has resisted, I do not think the defender of this thesis need be troubled even if the alternative theory is true. Or at least, as I shall explain in the last section, the alternative theory being true would be a lesser pill to swallow for the moralist of this thesis.

8.1 The Standard Route To Error Theory

In the final stages of his plan, just before he gives the final order for the city to dissolve into chaos, Dan - a dedicated immoralist - does not hesitate for a second. He is aware that he has put together every piece in a puzzle designed for the implementation of *his* plan, the satisfaction of *his* peculiarly strange ends. He knows that his actions and ends are not be justified by moral standards, but he is not moved by what morality dictates. Is there anything wrong with this person? Should he care about what morality would have him do?

Let us begin backwards: morality can be understood as a normative framework — a set of norms or requirements. In general, the question of what a normative framework requires of us, and the question of what we *really should do* may seem independent. To see this, suppose that etiquette requires that I do not talk with my mouth full, but that I need to do so in order to warn you that the food you are about to eat contains allergens. Intuitively, it seems that, contrary to what etiquette requires of me, I really should talk with my foul full. But some people think that when it comes to morality, the two questions are not independent. In particular, they think that moral oughts are *real* or *all-things-considered* oughts (Joyce 2000: 465). This is the thought that what morality requires of me is always what I *really* or *all-things-considered* ought to do.

A natural way to understand claims about what agents *really* should do, or what they *all-things-considered* should do, is in terms of a certain type of reasons: sufficient normative reasons (Joyce 2000: 465, Baker 2018: 4). The idea is that what I really or all-things-considered ought to do is what I have sufficient normative reason to do. As I explain in *1.1.2*, in being *normative*, a sufficient normative reason contrasts with motivating and explanatory reasons: a reason *for which* someone does something, and a reason *why* someone does something, respectively. A normative reason for action is a reason there *is* for doing something: a consideration that justifies or supports a prospective course of action. A sufficient normative reason is one that outweighs all

other relevant considerations that are at play, thereby determining what the justified course of action is: the one that is most reason supported. In being sufficient, it contrasts with *pro tanto* normative reasons: a consideration that justifies an action but which may be overruled by better ones.

Some people would think that Dan *really* should care about what morality would have him do, though he may choose to be indifferent to the injunctions of other normative frameworks, say, etiquette. They would seek to elaborate on this claim by pointing to the alleged fact that moral oughts or norms *invariably supply* agents with sufficient normative reasons: that every time an agent is morally required to perform some action, the agent has sufficient normative reason to do so in virtue of the fact that a moral requirement obtains. On those people's view, the fact that Dan chooses to act contrary to what he really should do, what he has sufficient normative reason to do, indicates that there is something wrong with him: he is either unresponsive or lacks appreciation for his practical reasons. In this thesis, we call people like this, *moralists*.

Others, notably error theorists, think otherwise. They begin by noting that albeit intuitive, the moralist's proposal is ultimately false. That is, contrary to what may seem intuitive to some people, and what moralists categorically profess, moral norms do not invariably supply agents with reasons. It is not true, for example, that Dan has sufficient normative reason to do as morality says merely in virtue of the fact that a moral requirement. In thinking this, error theorists find themselves in opposition to not only moralists, but also ordianry thought (see 2.3, 3). For the purposes of this chapter, I shall understand *error theory* as the view that ordinary thought is committed to error, and that the alleged error is the thought that moral norms invariably supply agents with sufficient normative reasons.

The error theorist of this thesis has attempted to vindicate her conclusion through the following argument:

(P1) Moral requirements *are supposed* to apply to agents irrespective of their ends.

(P2) Moral requirements *are supposed* to invariably supply agents with sufficient normative reasons.

(P3) There *are* only instrumental reasons.

(C) Moral norms – which express alleged moral requiments - do *not* invariably supply agents with reasons (or it is *not* the case that: moral norms invariably supply agents with reasons).

The first two premises, P1 and P2, tell us something about the actions that moral requirements tend to favour, whereas the third premise, P3, tells us something about what reasons there are: only ones favouring actions that are instrumental to the agent's ends (see 2.5). The combination of P1-P3 licenses the conclusion, C. For if all reasons are — as she thinks — instrumental, it may be possible that we are sometimes morally required to do an action for which we lack reason. And even if our ends and interests invariably align with moral requirements, and hence we invariably have reasons to do as moral requirements say, it is not true that we have such reason *because* we are morally required. For had our ends been different, we would have lacked reason

to do as moral requirements say. Hence it cannot be true that moral norms invariably *supply* us with reasons.

I have already defended the truth of P1 and P2 (chapters 2,3), and accepted that the conclusion follows from the premises. A lot turned, then, on the truth of the crucial third premise, P3. The majority of this thesis has been devoted to examining the grounds for thinking that P3 is true (see 8.1.1). Here, I shall understand an attempt to vindicate the error theorist's conclusion via the above argument as the *standard route to error theory*. In what follows, I summarise the support for the standard error theory.

8.1.1 Support For the Standard Error Theory

Why think that P3, and by extension the standard error theory, is true? The substantive part of this thesis, chapters 4-7, investigated the grounds for thinking that P3 is true. In chapter 4, I considered the most straightforward support for the crucial premise: a certain theory of reasons, *practical instrumentalism*. Practical instrumentalism holds that there are only reasons favouring actions that are instrumental to the agent's (for whom they are reasons) ends. I identified non-instrumentalism as the theory's main rival. In examining the two theories, I proposed two criteria that a theory of reasons must fulfil, and sought to examine how well each theory fares with respect to them. I concluded that we have little reason to think that instrumentalism is true, given that the theory does not fare better than its rival with respect to the two criteria. In the next three chapters, I looked for independent arguments in favour of instrumentalism and, by extension, P3 of the error theorist's argument.

In chapter 5, I considered a family of arguments that may be advanced in favour of instrumentalism, which go by the name of *reasons internalism*. Internalism specifies a criterion (or restriction) on what could count as a reason, which posits a necessary and interesting connection between the agent's reasons and her motivation. The idea was: whenever an agent has a reason to Φ, that reason must bear some relation to the agent's motivation towards Φing. Given the close connection between the agent's motivation and her own ends, I granted that if true, internalism can vindicate instrumentalism. Internalism, however, is itself a view under pressure, and the particular considerations in its favour were found to be unconvincing. My verdict was that internalists conflate psychology with normativity: the conditions under which agents are motivated, with considerations which justify their actions.

In chapter 6, I considered an argument which attempts to show that instrumental reasons are *conceptually basic*: the concept of an instrumental reason *just is* the concept of a normative reason for action. In effect, this is a conceptual argument purporting to show that as a matter of definitional (or conceptual) necessity, all reasons are instrumental: a conceptual route towards a substantive conclusion. Though the inference from the conceptual premise to the substantive conclusion may strike some as problematic, I was willing to accept the possibility for the sake of the argument. My efforts were instead concentrated on attacking the claim that the means/ends (M/E) inference is fundamental in that anyone who does not accept it cannot be considered competent with the notion of a practical reason:

(M/E) *Ceteris paribus*, if you desire to Ψ, and believing that by Φ-ing you will Ψ, then you ***ought*** to Φ.

By showing that M/E is not fundamental in the above stated sense, I effectively disputed the claim that instrumental reasons are conceptually basic, thereby resisting the alleged inference to the substantive conclusion.

In chapter 7, we sought support for the error theorist's crucial claim that all normative reasons are instrumental. Vindicating this premise would mean that the error theorist's conclusion that moral norms do not invariably supply agents with reasons is supported. For if moral norms have little to do with the attainment of the agent's ends, and all reasons are instrumental, it cannot be true that agents have reasons to do as moral norms say which obtain in virtue of those moral norms. The constitutivist argument considered here was meant to show that agents as such have only reason to aim at the attainment of their ends, in virtue of the fact that aiming at the attainment of one's ends is constitutive of agency. The argument puts forward an account of what is constitutive of agency that is generic enough to merit approval. But it fails nevertheless because it employs — and relies crucially on — a much dubious premise: that considerations pertaining to what is constitutive of agency can be informative about what agents have reason to do.

8.2. An Alternative Route to Error Theory

Following the rejection of the standard route, I shall now sketch an alternative route to error theory. In introducing this alternative route, it is helpful to make explicit a premise that is implicit in the argument of the standard route.

(P1) Moral requirements are supposed to apply to agents irrespective of their ends.
(P2) Moral requirements are supposed to invariably supply agents with sufficient normative reasons.
(P3) There are only instrumental reasons.
(P3implicit) There are *at least some* normative reasons (for agents to perform actions).
(C) Moral norms do not invariably supply agents with sufficient normative reasons.

For there to *be* only instrumental reasons, it follows that there are at least some reasons: *pro tanto* or *sufficient*. (Compare: if there *are* only green crocodiles, it follows that there are *at least some* crocodiles). It is just that moral norms cannot be supplying any of those reasons that there are — given that the latter are mostly in favour of actions that need not have anything to do with the satisfaction of the agent's ends.

The alternative route to error theory denies P3implicit, and by extension, P3. It does not accept the existence of sufficient normative reasons to begin with:

(P1*) There are *no* sufficient normative reasons (for agents to perform actions).

(C*) Moral norms do not invariably supply agents with sufficient normative reasons.

Recall how the error theorist of the standard route needs the conceptual premises, P1 and P2, in order to support the conclusion that there are no sufficient normative reasons supplied by moral norms. The error theorist of the standard route *does not deny* the existence of sufficient normative reasons. She rather accepts that there are sufficient normative reasons, though she insists that all such reasons are instrumental. Now, if there are only sufficient normative reasons, and moral requirements occasionally apply to agents, it is at least possible that moral norms (which represent those alleged requirements) invariably supply agents with sufficient normative reasons. That is, even if all sufficient normative reasons are instrumental, why think that moral norms do not invariably supply agents with reasons? On the hypothesis, for example, that something is a moral requirement if and only if it favours the agent's (to whom it applies) ends, and given that all sufficient normative reasons are instrumental, it certainly does not follow that moral norms do not invariably supply agents with reasons. But the conclusion, C, rules out that possibility. The argument thus needs a good story to tell: a premise which would validate the conclusion ruling out the possibility of moral reasons (sufficient normative reasons supplied by moral norms). It does so by telling us something substantive about the nature of those requirements: that they apply to agents irrespective of what their ends may be. This means that the dedicated immoralist, who makes acting against moral norms her sole primary end, does not have sufficient normative reason to comply with moral norms. She would have had, had moral norms invariably supplied her with sufficient normative reasons.

By contrast, the error theorist of the alternative route does not need any conceptual premises to support her conclusion. She does not accept the existence of any sufficient normative reasons in the first place. If there are not any sufficient normative reasons to begin with, and regardless of the nature of the alleged moral requirements, moral norms do not invariably supply agents with sufficient normative reasons. That is, the conclusion follows regardless of how we understand the nature of moral requirements. For example, even if moral requirements are supposed to favour the agent's (to whom they apply) ends, it still follows that moral norms do not invariably supply agents with sufficient normative reasons: there are no such reasons to begin with.

Skepticism about Sufficient Normative Reasons as such

The error theory of the alternative route is premised on an error theory about sufficient normative reasons *as such* — which is expressed by P1*. By contrast, the error theory of the standard route is premised on an error theory about *non-instrumental* reasons - pro tanto or sufficient. This is expressed by P3 of that argument. This is the kind of scepticism defended by Joyce (2001) and Dreier (1997), both of whom never object to the possibility of sufficient normative reasons as such.

We must be careful at this point not to conflate the scepticism expressed by the alternative error theory with its radical counterpart. The latter is discussed by Köhler and Ridge, who concern themselves with ‘an error theory about all practical normative judgements’ (Köhler & Ridge 2013: 429-430). The scepticism that concerns us here is

one about all *sufficient normative* reasons, as opposed to both *pro tanto and sufficient* normative reasons as such. A proponent of P1* need not object to the possibility of *pro tanto* normative reasons as such, only to the possibility of them adding up and thereby forming a *sufficient* or *adequate* normative reason.

8.2.1 The Merits of the Alternative Error Theory

The alternative error theory sketched above is an attractive view in virtue of its being able to accept some intuitive thoughts that its standard counterpart denies. What seems particularly attractive about the alternative theory is its stance on the possibility of *pro tanto* moral reasons: *pro tanto* normative reasons supplied by moral norms. At the same time, however, the alternative view is unable to accept one particular thought that is intuitive, and which can be accepted by defenders of the standard version. This is the thought that there are at least some sufficient normative reasons for action.

Through its commitments, specified above, the alternative version can hold dear to a thought that is both intuitive and which must be denied by the standard version. The combination of these facts gives the alternative version an edge over its standard counterpart. The important thought is this: moral considerations provide — even invariably — *pro tanto* normative reasons for action. The alternative theory can do this given that it is only committed to a denial of sufficient normative reasons as such, and absolutely need not object to the existence of *pro tanto* normative reasons provided by *this* or *that* normative framework. Given that she is able to maintain that a normative framework can supply pro tanto normative reasons for action, what is stopping her from also thinking that morality can invariably supply agents with such reasons? It seems that our alternative error theorist can consistently maintain both following claims: moral norms do not invariably supply agents with sufficient normative reasons, while they do invariably supply agents with pro tanto normative reasons.

Let us illustrate the above two claims via an example. Recall Andy who is exiting a burning building and has no inclination whatever — nor a desire or an end — for pulling the fire alarm in order to save the lives of others. There are two important considerations (or facts) in the example: the fact that Andy has no desire to save the lives of others, and the fact that by doing so he would save the lives of others. The latter may be called a moral consideration — insofar as it is deemed morally relevant. The standard error theorist would say: given Andy's particular ends, he has no sufficient normative reason to pull the fire alarm, nor any reason (i.e. pro tanto) to do so, and the fact by doing so he would save the lives of others (or that moral norms would have him do it) has no normative significance whatever. Moreover, she would elaborate, Andy has sufficient normative reason to not pull the fire alarm, again given his particuilar ends.

By contrast, our alternative error theorist need not even bother acknowledging Andy's particular ends, for she has no commitment linking the agent's ends with what she has any reason (pro tanto or sufficient) to do. Like her standard counterpart, she can reject the claim that Andy has sufficient normative reason to pull the fire alarm, but also reject the claim that Andy has sufficient normative reason to not pull the fire alarm (or do anything else that is compatible with him not doing so). Importantly, she can say that Andy has pro tanto normative reason to pull the fire alarm: that the moral

consideration that is at play represents a normative reason of sorts and hence has some normative significance.

We may sum up the attractions of the alternative version in the following way. First, it does not maintain that agents have sufficient normative reasons to be selfish, and second, it does not deny that moral considerations always represent pro tanto normative reasons for agents to do things. Insofar as both thoughts are intuitive, this is an attractive view — at least, more attractive than its standard counterpart which categorically rejects both.

Now of course, this alternative error theory too harbours unintuitive commitments. Apart from the obvious one, which is the denial that moral considerations invariably supply agents with sufficient normative reasons for action, it also rejects the thought that there are at least some sufficient normative reasons for action. By contrast, this is a thought accepted by its standard counterpart — and to this extent, the latter may be thought of as having an edge over the alternative version. Whenever someone denies a thought so dear to ordinary thought as this, the burden of proof is on her to support her claim. But how does one go about arguing for such a radical claim?

*8.2.2 Support for P1**

The alternative theory hinges upon the ground-breaking P1*. Here, I envision ways of supporting it. It seems to me that support for P1* is unlikely to come in substantive form. It is of course true that, as it stands, the premise is substantive: it tells that there *are no* sufficient normative reasons. My point is that it is unlikely that defenders of this premise will find further substantive things to say in its favour. In the case of standard error theory, the all-important P3 came declares *substantively* that there are only instrumental reasons, and finds support — at least in partly — in the form of independent substantive considerations. An important such considerations (expressed in chapter 4) is the claim that the true theory of reasons is practical instrumentalism. Another, which is in the service of the aforementioned, is the claim that instrumental reasons are metaphysically grounded, whereas non-instrumental reasons are not.

But how does one argue on substantive grounds that there are no sufficient normative reasons to begin with? It cannot, of course, do so indirectly — by showing that the only kinds of reasons there are are *such-and-such*, and that as a result, reasons that do not fulfil that description cannot be said to exist. Rather, one may do so by showing directly that there are no sufficient normative reasons at large (or of any kind): that any kind of sufficient normative reasons that can be said to exist do not in fact exist. At any rate, I know of no plausible substantive argument to the effect that there are no sufficient normative reasons. But I think it may seem at least initially plausible to argue for P1* in ways that are genuinely conceptual. In what follows, I draw on work from Copp (1997) and (more recently) from Baker (2018) in order to construct arguments in favour of the substantive claim that there are no sufficient normative reasons. The arguments are conceptual and they purport to infer substantive conclusions. They both aim to object to the very intelligibility or coherence of the notion of a sufficient normative reason. It must be noted that both Copp and Baker attack the notion of an *all-things-considered* ought, but I think the arguments are also straightforwardly about the notion of a sufficient normative reason. The connection is

easily discernible: the former features in a judgement about what an agent ought to do given all relevant considerations that are at play, and the latter favours the course of action that is most justified given all relevant considerations (pro tanto reasons) that are at play.

Circularity

Take an example of what may be called *a normative conflict (*or *a sticky situation,* cf. McPherson forthcoming). You recognise the existence of a number of pro tanto normative reasons that you seem to have, and which call for incompatible actions. You know you have a pro tanto reason to do as morality says, a pro tanto reason to do as etiquette would have you do, and a pro tanto reason to act according to your best interest. Recognising the existence of these reasons, you ask: What do I *really* ought to do? In asking this question, you do not seem to be not interested in a conclusion about an action that is merely supported by this or that pro tanto reason (cf. Copp 1997: 100). Rather, you seem to be interested in an action that is *most* reason supported: not merely supported by instrumental, moral or other considerations, but rather supported by sufficient normative considerations.

Suppose I tell you: given all these pro tanto reasons, what you have most reason to do is to Φ. At this point, you may wish to scrutinise my judgement. You are, however, not interested in the issue of what you have sufficient normative reason to do — what you find interesting is the very meaning of that judgement. You tell me: you say that I have sufficient normative reason to Φ, a judgement that has supposedly taken all other considerations that are at play (the other pro tanto reasons) in account. I say: yes. You go on: but what affirms the truth of this judgement? Must it not be a *further* judgement? Relatedly: what affirms the status of that consideration — which you picked out as my sufficient normative reason — as the normatively most important one? Must it not be a *further* consideration, that is, a *further* pro tanto reason which happens to be the most important one?

But, then, for every further consideration that does the relevant affirming (or for every further judgement that tells me what I have most reason to do), there must be a *further* one that does the same (cf. Copp 1997: 101). The vicious circle is obvious. For every consideration, R, that we come up to explain the normative importance of a sufficient normative reason, there must be a further consideration, R_2, which in turn explains the normative importance of R, and yet another, R_3, which in turn explains the normative importance of R_2, and so on *ad infinitum*. You now seem to be suggesting that the judgement that I initially gave you is nonsense, insofar as it presupposes, and hence its intelligibility hinges on, a notion that has untenable conceptual implications.

Unanalysability

A second argument objects to the notion of a sufficient normative reason for action on the grounds that it eludes proper analysis. The argument is inspired by Baker (2018) who objects to the notion of an *all-things-considered ought* on similar grounds.

Suppose you are interested in the concept of a sufficient normative reason — you want to know more about it. By way of analysis, you mention the concept WITCH

and all the interesting things we can say about it: 'It is the concept of a person who possesses magical powers', 'The concept of a person who has made a pact with the devil', etc. But when it comes to providing a similar analysis of SUFFICIENT NORMATIVE REASON, you say, there are no interesting things to say besides some obvious truisms: 'It is the weightiest consideration', 'It is the consideration that determines the justified course of action', 'It is the consideration that overrules other competing ones', etc . You want an analysis of SUFFICIENT NORMATIVE REASON which will be given in *non-circular* terms.

Moreover, you are quick to mention that you want an analysis that is given in *normative terms*. Elaborating on this, you tell me that such an analysis cannot be given in either psychological terms or any other terms that 'change the subject' (cf. Baker 2018: 236). An example of a subject-changing analysis would define SUFFICIENT NORMATIVE REASON as some consideration pertaining to what strong desire you have; e.g. 'It is the consideration that is entailed by your strongest desire'. Recalling one of your favourite authors, you insist that the weight of reasons is a *normative*, rather than, say, a psychological — *matter* (Schroeder 2007: 100).

Those so called *sufficient normative reasons*, you go on, are supposed to have some special property in virtue of which they determine the justified course of action — and importantly, they do so in the face of further competing considerations that are at play. The characterisation of that special property, you now think, is typically metaphorical or otherwise hopelessly vague and ambiguous (cf. Baker 2018: 236). You now seem to be suggesting that the notion of a sufficient normative reason is nonsense, or at least in some other way defective, insofar as it eludes proper analysis.

8.3 Assessment of the Alternative Error Theory

We now have a complete sketch of an alternative error theory. This includes the basic commitments of the view, some important implication, along with some form of support for the crucial premise that it hinges upon. In this thesis, the alternative view is doomed to receive less attention than the standard version, but there are at least three questions that it is pressing to address here. First, do we have good reason to accept it? Second, if we do, what is the significance for this thesis? For example, in the face of the alternative error theory, and given that it is plausible, should the moralist concede defeat to the sceptic of moral reasons?

8.3.1 Is the Alternative Theory Plausible?

As already mentioned, the plausibility of the alternative error theory hinges on the plausibility of P1*, which in turn, hinges on arguments that may be offered in its support. If we do not have good reason to accept those arguments, we have no good reason to accept the premise and the theory at large. Now, I sketched two arguments in the previous section, and I said that there may be *initially* plausible. But I do not think the arguments are particularly convincing. The first argument, which complains that the notion of a sufficient normative reason implies an infinite regress and hence involves untenable commitments, makes a false presupposition — namely, that a consideration representing a sufficient normative reason is in need of further, independent affirmation. That is, why think that there must be something above and beyond a sufficient normative reason that provides it with, or contributes to, its

normative significance? In thinking that there must be, critics neglect the plausible idea that the normative significance (or 'supremacy') of a sufficient normative reason may be primitive (cf. Baker 2018: 6). Sufficient normative reasons may be well regarded as fundamental normative elements the significance of which is not in need of independent affirmation. Following Baker, 'if normative non-naturalism is correct, then whichever normative properties are most fundamental — the property of being a reason, say — admit of no further explanation' (ibid).

Regarding the second argument, things are quite similar. Recall that it complains that the relevant notion eludes proper analysis — one that others, say, the concept of a witch admit of. Moreover, the notion of a sufficient normative reason seems to imply that a sufficient normative reason has a special property in virtue of which it stands out from the rest of competing considerations and determines the justified course of action. But when analysing the notion, there seems to be no such property for us to specify. Therefore, the notion must be in some way defective.

In response, the first thing to note here is that the argument may show the notion is circular, but it certainly does not show that it is incoherent. At worst, it shows that it is in need of further analysis. Compare: the theory of big bang may be incomplete in virtue of the fact that it does not tell us what exactly came before the great explosion, but for this reason it cannot be said to be incoherent. It may be incomplete — but still the best story we have. Analogously, the notion of a sufficient normative reason may be incomplete, but it is the best one we have to account for the intuition that on some occasions one course of action is more justified than another, and for that matter, most justified than the rest. This is important as it is the latter type of criticism, one about incoherence, that seems to license the conclusion that there are no sufficient normative reasons. Compare: if you want to make me conclusively think that a truly omnipotent being has never existed, telling me that the historian's account on it is incomplete, or even nonexistent, need not do. What you need to point to is something stronger: that for it to have existed, it must have been possible to create something so heavy that even itself cannot lift. In other words, you need to point to an incoherence involved in the very idea of the thing the existence of which you are denying, at least if you seek to rule out its existence on conceptual grounds alone.

Apart from the inadequacies of the two arguments offered in its support, the alternative theory seems to face an independent challenge. In short, it may be incoherent *itself*. Recall that the alternative view has it that there are no sufficient normative reasons at large, though there can well be pro tanto normative reasons. A direct implication of this is that it may well be true that moral norms invariably supply agents with pro tanto normative reasons for action. For if such reasons do exist, and there is an intuitive link between what we are morally required to do, on the one hand, and what we have reason (of some sort or other) to do, why should the alternative error theorist who is only committed to a radical denial of sufficient normative reasons deny this? Allowing for this possibility, we have noted, it is an important part of what makes the alternative theory an attractive view.

But allowing for the possibility of pro tanto normative reasons, while at the same time categorically denying the possibility of sufficient normative reasons, may lead to the theory swallowing itself. The worry is that the notion of a pro tanto normative reason does not make sense in the absence of a notion of sufficient

normative reason: it is rooted in the latter — it indispensably implies it. In short: what does it mean for a consideration to be a *defeasible* reason for action, or to this or that extent, if it does not mean that it is *potentially* a sufficient reason for action? If it is indeed part of the notion of a pro tanto reason that it has at least the potential to be sufficient, then it makes at least some reference to the notion of a sufficient normative reason. And if the latter is incoherent, as an argument in favour of the alternative view would have it, then that argument is incoherent much like the notion that it deems so! The upshot here is not that the alternative theory itself is incoherent, only that conceptual arguments such as the one offered in its support may be. In other words, it is not incoherent per se to deny that sufficient normative reasons exist but pro tanto reasons do. Rather, what might be incoherent is to attempt to support this view via arguments to the effect that the notion of sufficient normative reason is incoherent.

Following the above worry, the alternative error theorist might wish to cease to be neutral on the question of pro tanto reasons. That is, she may also grant that such reasons do not exist. But now her theory loses its credentials: it is no more an attractive theory, which despite denying the intuition that moral norms invariably supply agents with sufficient normative reasons, nevertheless allows that they do invariably supply agents with *some* reasons. Moreover, if the alternative error theory cannot find support in any substantive considerations, and hence it hinges crucially on conceptual arguments to the effect that the notion of a sufficient normative reason is incoherent, then the theory is doomed to remain unsupported.

8.3.2 Should the Alternative Theory be Troubling to the Moralist?

The alternative error theory sketched is deserving of being considered in this context — given that it purports to lend support to the same conclusion that this thesis has been opposed to: that there are no sufficient normative reasons supplied by moral norms. My primary aim in this chapter, however, is not to conclusively argue that the alternative error theory is implausible — though I do have doubts as to whether the all-important P1* can be supported. Rather, my real interest lies in sketching an alternative theory, which may share the same conclusion as the standard version, but which need not trouble the moralist of this section who resists the view that moral norms do not invariably supply agents with reasons. In what follows, I explain why I think this.

At least, moral considerations do not 'lose' to competing ones

The natural first thing to say here is that the alternative error theory is essentially an error theory about all *sufficient normative reasons* (henceforth, just *reasons*) — rather than an error theory about *sufficient normative reasons invariably supplied by moral norms* (henceforth, just *moral reasons*). In short: it is a *global* error theory — as opposed to a *moral* error theory.[113] The alternative theory maintains that there are no moral reasons, *because there are no reasons*. The standard theory maintains that there are no moral reasons, *because there are only reasons for agents to perform actions that favour the satisfaction of their ends*. Insofar as both theories

[113] I say 'global' here to mean about all sufficient normative reasons. Again, 'global' here can be misleading as it might suggest the denial of all types of normative reasons: pro tanto and sufficient. For such a global error theory, see (Köhler & Ridge 2013).

support the underlined statement, they seem equally troubling for a defender of moral reasons. Yet there seems to be a comforting — albeit somewhat elusive — thought in thinking that there are no moral reasons *because there are no reasons to begin with*, as opposed to *because there are only reasons of a certain kind*.

Let us try to locate the elusive thought. Perhaps we can do so by way of analogy. Compare: there seems to be something comforting in knowing that I did not win the race, *because someone else won it*, as opposed to *because the race never took place*. Granted: in both scenarios, I never won the race. But the fact that the race never took place implies that I never *lost* to anyone. Similarly, I think, the claim that there are no reasons to begin with implies that moral considerations never 'lose' to other competing ones in the event of a normative conflict. Consider Andy, a dedicated immoralist who has the end of wreaking misery to others. Andy is currently exiting a burning building and has the opportunity to pull the fire alarm to save the lives of others. There seem to be two competing considerations here: the fact that Andy has the end of wreaking misery to others, and the fact that by doing so he would save the lives of others. The latter is a moral consideration (as it plays a central role in determining what morality requires of Andy), the former is a psychological one about Andy. If the standard error theory is true, and all reasons are instrumental to the agent's ends, the moral consideration that is at play does not supply Andy with a reason to pull the fire alarm, but the competing psychological one does. If the alternative error theory is true, and there are no reasons to begin with, the moral consideration that is at play does not supply Andy with a reason to pull, but neither does the competing psychological one. In the event of the standard theory, moral considerations invariably 'lose' to psychological (or instrumental) ones. In the event of the alternative theory, moral considerations never lose to any competing ones. Is this not somewhat comforting?

We can make the above suggested contrast more intelligible by observing two contrasting implications that follow from each theory. On the standard theorist's view, dedicated immoralists have no (sufficient normative) reason to act morally, though they may well have (sufficient normative) reason to act selfishly (or even immorally): in accordance with their own ends (which may even be immoral). By contrast, on the alternative theorist's view, dedicated immoralists have no (sufficient normative) reason to act morally, nor do they have any such reason to act selfishly. It is thus *not* an implication of the alternative theorist's view that selfish (potentially immoral) behaviour is justified in any normatively robust sense. The moralist, who in the face of the alternative theory being true may be forced to abandon her commitment that moral norms invariably supply agents with reasons, can now switch her allegience to the following claim: dedicated immoralists have no (sufficient normative) reason to act either selfishly or immorally because there are no such reasons to begin with. And though this does not mean that dedicated immoralists have (sufficient normative) reason to act morally, it at least does not mean that they have such reason to act immorally. This is a lesser pill to swallow than the claim, which the moralist would have to accept had the standard error theory been true, that dedicated immoralists have (sufficient normative reason) to act immorally.

Relatedly, the moralist may regard the premise that there are at least some reasons as *a starting point* in the particular discussion that she engages in. She may say: when I claim that there are no moral reasons, I *presuppose* that there are some

reasons to begin with. I do so, she may continue, because my inquiry is driven by a background concern: to resist the thought that the dedicated immoralists' ends supply her with reasons that are weightier than those supplied by moral considerations. Insofar as someone denies the premise that plays the role of the presupposition in the moralist's inquiry, she does not share her background concern. Thus the 'stranger' arguably engages in a similar, though subtly distinct, discussion.

The above is, of course, nor to say that the stranger's claim makes no sense if she does not accept the presupposition, nor tha anyone who denies the latter is not worthy of consideration. Far from it. In particular, our moralist must acknowledge that in the same way that one can intelligibly claim that there are no benevolent witches *because there are no witches to begin with*, one can claim that there are no (sufficient normative) reasons invariably supplied by moral norms *because there are no such reasons in the first place.*

8.4 Aftermath

This thesis has isolated a particular thought which belongs to the vast majority (if not all) of those engaging in ordinary moral discourse, namely that moral norms invariably supply agents with sufficient normative reasons, and took on the task of defending it against the *standard* error theorist's attack. A number of arguments have been considered, all initially promising but — as I have hopefully suggested — ultimately flawed. A reasonable next inquiry would be looking for further arguments in support of the standard error theorist's view. As I have explained in this chapter, the standard view is the most devastating view for the defender of moral reasons, as it not only denies the existence of moral reasons, but it can also be used as an instrument for justifying selfish (even immoral) behaviour. The latter may be read as a practical concern, as the dedicated immoralist's ability to convince herself that her actions are truly justified sounds like trouble, but it is also a theoretical defeat of sorts. That is, if the moralist cannot save the thought that there are moral reasons, it would be much worse if she also had to surrender what I here called a background concern: resisting the thought that the dedicated immoralist's ends supply her with reasons that are weightier than those supplied by moral considerations.

The best defence is a good offence. The background concern is true, and must be preserved at all cost. The moralist must identify alternative ways that the detrimental standard error theory may be supported, and seek to spot their inadequacies. For example, perhaps practical instrumentalism can be said to be superior to non-instrumentalism with respect to metaphysical grounding (see 4.7) — if so, in virtue of what methodological principle? Can we apply *that* methodological principle against other cases, and see how well it intuitively fares with respect to them?

Another question worth investigating is the plausibility of the alternative error theory. I have suggested that support for it is unlikely to come in any substantive form, and that the conceptual route is that theorist's only hope. Can a plausible, substantive theory of reasons be developed which would rule out the existence of sufficient normative reasons as such? Moreover, I cast doubt on the plausibility of the conceptual arguments considered here — arguments which have been advanced by respectable thinkers. Perhaps they are right and I am wrong. First, is a sufficient normative reason in need of affirmation, or can its normative significance be said to be primitive?

Whatever the answer, is it consistent with our way of thinking about other domains? Second, does the notion of a pro tanto reason presuppose that of a sufficient reasons, such that the former makes no sense without the possibility of the latter?

Bibliography

Audi, Robert. "Prospects for a Naturalization of Practical Reason: Humean Instrumentalism and the Normative Authority of Desire." *International Journal of Philosophical Studies* 10, no. 3 (2002): 235 – 63.

Baber, H. E. "The Ethics of Dwarf-Tossing." *International Journal of Applied Philosophy* 4, no. 4 (1989): 1-5.

Baker, Derek Clayton. "Skepticism About Ought Simpliciter." *Oxford Studies in Metaethics* 13 (2018).

Bedke, Matthew S. "Practical Reasons, Practical Rationality, Practical Wisdom." *Ethical Theory and Moral Practice* 11, no. 1 (2008): 85-111.

Blackburn, Simon. "Supervenience Revisited." In *Exercises in Analysis: Essays by Students of Casimir Lewy*, edited by Ian Hacking, 59--74: Cambridge University Press, 1984.

———. *Ruling Passions: A Theory of Practical Reasoning.* Oxford University Press UK, 1998.

Brink, David O. "Moral Realism and the Sceptical Arguments from Disagreement and Queerness." *Australasian Journal of Philosophy* 62, no. 2 (1984): 111 – 25.

Broome, John. "Does Rationality Give Us Reasons?". *Philosophical Issues* 15, no. 1 (2005): 321–37.

———. "Is Rationality Normative?". *Disputatio* 2, no. 23 (2007): 161-78.

Burgess, John P. "*Against Ethics*." *Ethical Theory and Moral Practice* 10, no. 5 (2007): 427-39.

Carroll, Lewis. "What the Tortoise Said to Achilles." *Mind* 4, no. 14 (1895): 278-80.

Copp, David. "The Ring of Gyges: Overridingness and the Unity of Reason*: David Copp." *Social Philosophy and Policy* 14, no. 1 (1997): 86-106.

Cullity, Garrett, and Berys Gaut. "Introduction." In *Ethics and Practical Reason*, edited by Garrett Cullity Berys Gaut, 1-27: Oxford University Press, 1997.

Dennett, Daniel C. "*Darwin's Dangerous Idea*." *Behavior and Philosophy* 24, no. 2 (1996): 169-74.

Dreier, James. "Humean Doubts About the Practical Justification of Morality." In *Ethics and Practical Reason*, edited by Garrett Cullity and Gaut Berys, 81--100: Oxford Clarendon Press, 1997.

Dworkin, Ronald. "Objectivity and Truth: You'd Better Believe It." *Philosophy and Public Affairs* 25, no. 2 (1996): 87-139.

Enoch, David. "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action." *Philosophical Review* 115, no. 2 (2006): 169-98.

———. "How Is Moral Disagreement a Problem for Realism?". *The Journal of Ethics* 13, no. 1 (2009): 15-50.

———. "Shmagency Revisited." In *New Waves in Metaethics*, edited by Michael Brady: Palgrave-Macmillan, 2011.

Falk, W. D. "Goading and Guiding." *Mind* 62, no. 246 (1953): 145-71.

Finlay, Stephen. "*The Error in the Error Theory*." *Australasian Journal of Philosophy* 86, no. 3 (2008): 347-69.

———. "The Obscurity of Internal Reasons." *Philosophers' Imprint* 9 (2009): 1-22.

———. "Errors Upon Errors: A Reply to Joyce." *Australasian Journal of Philosophy* 89, no. 3 (2011): 535 - 47.
Foot, Philippa. "Morality as a System of Hypothetical Imperatives." *Philosophical Review* 81, no. 3 (1972): 305-16.
Garner, Richard T. "*On the Genuine Queerness of Moral Properties and Facts*." *Australasian Journal of Philosophy* 68, no. 2 (1990): 137 – 46.
———. *Beyond Morality.* Vol. 45: Temple University Press, 1994.
Gibbard, Allan. "Reasons Thin and Thick." *Journal of Philosophy* 100, no. 6 (2003): 288-304.
Goldman, Alan H. *Reasons from Within: Desires and Values.* Oxford University Press, 2009.
Harman, Gilbert. "*Moral Relativism Defended*." *Philosophical Review* 84, no. 1 (1975): 3-22.
Holroyd, Jules. "Responsibility for Implicit Bias." *Journal of Social Philosophy* 43, no. 3 (2012): 274-306.
Hooker, Brad, and Bart Streumer. "Procedural and Substantive Practical Rationality." In *The Oxford Handbook of Rationality*, edited by Piers Rawling and Alfred R. Mele, 57--74: Oxford: Oxford University Press, 2003.
Hubin, Donald C. "What's Special About Humeanism." *Noûs* 33, no. 1 (1999): 30-45.
Hume, David. *A Treatise of Human Nature.* Oxford University Press, 1738.
———. *An Enquiry Concerning Human Understanding.* Oxford University Press, UK, 1748/2007.
Joyce, Richard. "Error Theory."
———. "*The Fugitive Thought*." *Journal of Value Inquiry* 34, no. 4 (2000): 463-78.
———. *The Myth of Morality.* Cambridge, UK ; New York: Cambridge University Press, 2001.
———. *The Evolution of Morality.* Life and Mind. Cambridge, Mass.: MIT Press, 2006.
———. "*The Error in 'the Error in the Error Theory'*." *Australasian Journal of Philosophy* 89, no. 3 (2011): 519-34.
Katsafanas, Paul. *Agency and the Foundations of Ethics: Nietzschean Constitutivism.* Oxford University Press UK, 2013.
Kelly, Daniel, and Erica Roedder. "Racial Cognition and the Ethics of Implicit Bias." *Philosophy Compass* 3, no. 3 (2008): 522–40.
Kolodny, Niko. "Why Be Rational?". *Mind* 114, no. 455 (2005): 509-63.
Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity.* Oxford University Press, 2009.
Langford, C. H., and Paul Arthur Schilpp. "The Notion of Analysis in Moore's Philosophy." *Journal of Symbolic Logic* 8, no. 4 (1943): 149-51.
Lillehammer, Hallvard. "The Doctrine of Internal Reasons." *Journal of Value Inquiry* 34, no. 4 (2000): 507-16.
———. "The Argument from Queerness." In *The International Encyclopedia of Ethics*, edited by Hugh LaFollette, 2013.
MacIntyre, Alasdair C. *After Virtue: A Study in Moral Theory.* Vol. 92: University of Notre Dame Press, 1981.
Mackie, John L. *Ethics: Inventing Right and Wrong.* Penguin Books, 1977.
Margolis, Eric, and Stephen Laurence. *Concepts: Core Readings.* MIT Press, 1999.

McPherson, Tristram. "Authoritatively Normative Concepts." In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau: Oxford University Press.
———. "Ethical Non-Naturalism and the Metaphysics of Supervenience." In *Oxford Studies in Metaethics Vol 7*, 205, 2012.
Millgram, Elijah. "Williams' Argument against External Reasons." *Noûs* 30, no. 2 (1996): 197-220.
Nietzsche, Friedrich Wilhelm. *The Antichrist.* Prometheus Books, 1911.
———. *On the Genealogy of Morality.* Vol. 9: Cambridge University Press, 1998.
Olson, Jonas. "Reasons and the New Non-Naturalism." In *Spheres of Reason*, edited by Simon Robertson: Oxford University Press, 2009.
———. *Moral Error Theory: History, Critique, Defence.* Oxford University Press, 2014.
Pettit, Philip, and Michael Smith. "Backgrounding Desire." *Philosophical Review* 99, no. 4 (1990): 565-92.
Plato. *Republic.* Oxford University Press UK, 380 BC/2008.
Quine, Willard V. O. "Two Dogmas of Empiricism." *Philosophical Review* 60, no. 1 (1951): 20–43.
Railton, Peter. "On the Hypothetical and Non-Hypothetical in Reasoning About Belief and Action." In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Nigel Gaut, 53--79: Oxford University Press, 1997.
Robertson, Simon. *Spheres of Reason: New Essays in the Philosophy of Normativity.* Oxford University Press, 2009.
Scanlon, Thomas. *What We Owe to Each Other.* Vol. 66: Belknap Press of Harvard University Press, 1998.
Schroeder, Mark. *Slaves of the Passions.* Oxford University Press, 2007.
Sebastian Köhler, Michael Ridge. "Revolutionary Expressivism." *Ratio* 26, no. 4 (2013): 428-49.
Shackel, Nicholas. "Still Waiting for a Plausible Humean Theory of Reasons." *Philosophical Studies* 167, no. 3 (2014): 607-33.
Shafer-Landau, Russ. *Moral Realism: A Defence.* Oxford University Press, 2003.
———. "A Defence of Categorical Reasons." *Proceedings of the Aristotelian Society* 109, no. 1pt2 (2009): 189-206.
Shemmer, Yonatan. "Desires as Reasons." *Philosophy and Phenomenological Research* 75, no. 2 (2007): 326–48.
Silverstein, Matthew. "Teleology and Normativity." *Oxford Studies in Metaethics* 11 (2016): 214-40.
Sinclair, Neil. "Reasons, Inescapability and Persuasion." *Philosophical Studies* 173, no. 10 (2016): 2823-44.
———. "Reasons Internalism and the Function of Normative Reasons." *Dialectica* 71, no. 2 (2017): 209-29.
Singer, Peter. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1, no. 3 (1972): 229-43.
———. *Practical Ethics.* Cambridge University Press, 1979.
Sinnott-Armstrong, Walter. *Moral Skepticisms.* Oxford University Press, 2006.
Smith, Michael. *The Moral Problem.* Blackwell, 1994.
Sorensen, Roy A. *A Brief History of the Paradox: Philosophy and the Labyrinths of the Mind.* Oxford University Press, 2003.

Thomson, Judith Jarvis. "A Defense of Abortion." *Philosophy and Public Affairs* 1, no. 1 (1971): 47-66.
Velleman, David. *Practical Reflection.* Vol. 94: Princeton University Press, 1989.
———. *The Possibility of Practical Reason.* Vol. 106: Oxford University Press, 2000.
———. "Replies to Discussion on the Possibility of Practical Reason." *Philosophical Studies* 121, no. 3 (2004): 277-98.
Wiggins, David. "On Being in the Same Place at the Same Time." *Philosophical Review* 77, no. 1 (1968): 90-95.
Williams, Bernard. *Moral Luck: Philosophical Papers, 1973-1980.* Cambridge University Press, 1981.
———. *Ethics and the Limits of Philosophy.* Routledge, 1985.
Zangwill, Nick. "Moral Mind-Independence." *Australasian Journal of Philosophy* 72, no. 2 (1994): 205-19.