What to do when the world doesn't play along:
Life after moral error theory

JESSICA ISSEROW

A thesis submitted for the degree of
Doctor of Philosophy of
The Australian National University
August 2017

Statement

This thesis is solely the work of its author. No part of it has previously been submitted for any degree, or is currently being submitted for any other degree. To the best of my knowledge, any help received in preparing this thesis, and all sources used, have been duly acknowledged.

JESSICA ISSEROW
15th August, 2017

Acknowledgements

I had the pleasure and good fortune of writing my PhD at the philosophy department at the Australian National University, a graduate student's paradise. I was also fortunate to have had an all-star supervisory panel, whose diversity helped me to see issues from different philosophical perspectives. I first want to thank the chair of my panel, Benjamin Fraser, for being incredibly supportive over the years, and for providing insightful comments, raising important challenges and making valuable suggestions. I am also immensely grateful to Daniel Nolan, who was incredibly generous with his time, and whose feedback was always constructive and edifying. I would like to thank Kim Sterelny for his help, encouragement, and for making my time at the ANU so philosophically rewarding. I am grateful to Nic Southwood for providing straight and honest feedback, acknowledging when my work was up to scratch and having a laugh with me whenever it wasn't.

I am also thankful to those at the ANU who made my time there so enjoyable and philosophically stimulating. My work on this thesis has benefitted greatly from discussions with Clare Due, Melissa Ebbers, Edward Elliott David Kalkman, Colin Klein, Christopher Lean, Leon Leontyev, Eric Llamas, Alexander Sandgren, and Lachlan Umbers. Special thanks is owed to Edward Elliott, who was the preview audience for all of the presentations I have given on this work. I also wish to thank audiences at the ANU, at annual meetings of the Australasian Association of Philosophy, and many conferences at Kioloa.

I spent most of the final year of my PhD at the University of Leeds, and I will be forever grateful to the philosophers there for their help and kindness. (And, of course, for the formidable karaoke sessions as well.) I am especially grateful to Daniel Elstein, Ulrike Heuer, Gerald Lang, Pekka Värynyen, and Jack Woods for meeting with me to discuss my research, as well as Richard Rowland, who met with me while visiting there.

I should also register my appreciation for the funding that I received. This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

On a more personal note, I would like to thank (all four of) my parents, whose love and support has been unwavering throughout the years. They have always encouraged me to pursue my dreams, and for that I am eternally grateful. Finally, I would like to thank Edward Elliott for his love, for believing in me, and for being the best friend that I've ever had.

Abstract

This work addresses the 'what next?' question for moral discourse, which concerns the best choice of action given the truth of the moral error theory. The moral error theory comprises two claims: (i) that moral discourse is used assertorically, and (ii) that moral assertions systematically fail to state truths. The upshot of the moral error theory is that nothing is really right or wrong—indeed, that the very idea of things being right or wrong is fundamentally mistaken. And yet, I argue, there are strong arguments in favour of moral error theory. With such far-reaching implications, we'd do well to have some guidance regarding what we ought to do upon coming to believe that the moral error theory is true.

In the first part of this work, I evaluate the answers to the 'what next?' question that have been proposed in the current literature. These include a systematic revision of our moral concepts (revisionism), preserving moral language in the spirit of a useful fiction (fictionalism), ridding ourselves of moral discourse entirely (abolitionism), and making do with our current erroneous moral discourse (conservationism). I argue that none of the first three proposals offer us an entirely satisfactory answer to the 'what next?' question. Conservationism is the most promising solution still on the table.

However, conservationism is yet to be fully developed. In the second part of this work, I develop and motivate my own version of conservationism, and show that it is the most attractive response to the 'what next?' question; one that is capable of securing the many desirable practical goods that our moral practices provide.

This page intentionally left blank

Contents

INTRODUCTION

Error theories and the 'what next? question

In his *On the Plurality of Worlds*, David Lewis recommends that we adopt a "simple maxim of honesty" when constructing our philosophical theories: "never put forward a…theory that you yourself cannot believe in your least philosophical and most commonsensical moments" (1986, p.135). It is safe to say that not everyone has taken Lewis's suggestion to heart. Philosophers have told us that contrary to what we believe, there are no beliefs (Churchland 1981)—or any philosophers for that matter (Unger 1979)! They have also denied the existence of colours (Boghossian & Velleman 1989), numbers (Field 1980), sexual perversion (Priest 1997), emotions (Griffiths 1997), and even moral facts (Mackie 1977, Joyce 2001, Olson 2014).

Assuming, as these philosophers do, that the sentences that figure in our talk of numbers, colours, and the like purport to be stating truths, this means that we ought to be *error theorists* about these discourses. Roughly, to be an error theorist about a discourse is to hold that certain sentences of that discourse which seem to be clearly aiming at truth are incapable of achieving it; that is, the world is *just not the right way* for basic assertions made within that discourse to come out true. The error theorist about colour discourse, for example, takes sentences like 'roses are red' and 'violets are blue' to be stating purported truths. But she also holds that such sentences cannot be true because, strictly speaking, *there are no colours*.

Of course, error theories are not merely the province of philosophers. As Richard Joyce and Simon Kirchin observe,

> There is nothing terribly complicated, esoteric, or unfamiliar in the idea of taking the error theoretic stance towards a problematic subject matter...It is the attitude that sensible persons take towards phlogiston, astrology, the Loch Ness monster, and the existence of reliable causal relations between severed rabbits' feet and episodes of good luck. (2010, pp.x-xi)

Nonetheless, one is apt to encounter many surprising error theories within philosophical circles;[1] for philosophers often deny the existence of objects, properties, or relations that are quite familiar to us. Unlike error theories regarding astrology and lucky rabbits' feet, error theories targeted at colours, beliefs, and moral facts are bound to strike us as controversial—at least on first appearances. Our talk of colours doesn't seem remotely on a par with talk of astrology. Using the stars to predict one's fortune seems wrongheaded in a way that taking apples to be red is not.

As such, philosophers have had much to say about whether we *should* opt for error theories in familiar discursive domains. Frank Jackson, for example, advises that we "work on the general presumption that the folk are not badly confused", and takes this to count against an error theory about colour discourse (1998a, p.103). Peter van Inwagen cautions that subscribing to an error theory carries the threat of a more global kind of scepticism. Anyone "…who denies what practically *everyone* believes is…", he warns, "…adopting a position according to which the human capacity for knowing the truth about things is radically defective" (1990, p.103, emphasis in original).

Yet comparatively little has been said about just what ought to be done in the event that *we do* subscribe to an error theory of some form or other. Should an error theory of some kind turn out to be plausible, then what ought we to do with the erroneous discourse? This is the 'what next?' question (WNQ) that follows our belief in an error theory; it is the question regarding what we ought to do with a discourse if we take an error theory regarding that discourse to be true.[2]

The focus of this thesis will be the WNQ that accompanies *a moral error theory* in particular. Just as error theorists about colour deny that anything is really red or blue, moral error theorists deny that anything is really right or wrong (Mackie 1977, Joyce

[1] This is not to suggest that one is not apt to encounter many surprising error theories *outside* philosophical circles. Error theories about King Arthur, phlogiston, and anthropogenic global warming, for instance, have all been raised and contested, but not primarily by philosophers. Thanks to Daniel Nolan for pointing this out.

[2] WNQs are traditionally formulated as questions regarding what we *ought* to do with an erroneous discourse. But for reasons to be discussed later on (§2.1), it is controversial whether the WNQ that shall form our primary focus—the WNQ for moral discourse—can be a normative question. I will argue that it can.

2001, Olson 2014). According to them, moral discourse is guilty of a systematic error, and so, no moral claims can be true.[3] For the philosophical purposes of this thesis, the moral error theory is assumed rather than defended. I take as my starting point the assumption that the moral error theory *is* true (and moreover, that *we believe* it is true,) and provide an answer to the WNQ that follows.[4] (I will specify who 'we' are in §2.2.) As I will suggest, a proper investigation of this question can deliver some important lessons regarding how we should proceed when attempting to answer analogous WNQs in other discursive domains.

In this work, I will evaluate the solutions that have been proposed in response to the WNQ for moral discourse in particular. These include:

> **Abolitionism:** ridding ourselves of moral discourse entirely.
> **Revisionism:** revising our moral concepts and/or how we use moral language.
> **Fictionalism:** preserving moral language in the form of a useful fiction.
> **Conservationism:** making do with our current, erroneous moral discourse.[5]

None of the first three proposals, I shall argue, offer us a fitting solution to the WNQ for moral discourse; some are likely to be infeasible, others unstable, and if enacted, all are likely to result in the loss of the many desirable practical goods that our erroneous moral discourse provides. To my mind, conservationism is the most promising option, but it is yet to be properly developed. In the final part of the work, I will develop my own brand of conservationism, which is well-motivated, feasible, and (I argue) has the best chance of securing the benefits of our error-ridden moral discourse.

My purpose in this introduction will be to familiarise the reader with error theories and the sorts of answers that are proposed in response to the WNQs that often accompany them. I do so to provide the reader with a sense of the problem that the thesis addresses. A more detailed discussion of moral error theory is the task of the

[3] More specifically, *a proper subset of* these claims cannot be true. I will get to this clarification shortly.

[4] Although I am concerned with a situation in which this conjunction holds—one in which (i) the moral error theory is true, and (ii) we believe that it is true—I will often frame the WNQ only in terms of ones of these conjuncts for ease of expression.

[5] Conservationism sometimes goes under the name 'inconsistent nihilism'. See Pigden (1991), and Nolan, Restall, and West (2005, p.314).

first chapter. Clarifying the WNQ for moral discourse and motivating the project of addressing it is a job for the second.

I begin here by offering a characterisation of error theories more generally (I). After discussing more familiar error theories (e.g., witches, dragons), I explain the sorts of considerations that motivate error theories within philosophy (II). Following that, I outline the kinds of answers that are usually offered in response to WNQs (III). These proposals shall be considered when we later move on to evaluate different answers to the WNQ for moral discourse. I conclude with a blueprint of the remainder of the thesis (IV).

§I. Error Theories

Error theories are typically directed at a target discourse (cf. Miller 2010). A discourse, as I shall understand it, is a domain of talk and thought that is structured around a particular set of concepts and associated beliefs. Witch discourse, for example, is structured around concepts such as <witch>, <evil>, and <magic>.[6] For the folk of the 16th century, this discourse also involved a characteristic set of beliefs; the belief that witches consorted with the devil, the belief that they had magical powers, and the belief that they used such powers to further nefarious ends.

When a particular discourse centrally involves reference to properties, objects, or relations that (we believe) fail to exist, then we will likely be error theorists about that discourse. (There is a distinction to be made between error theorists who think that the relevant properties, objects, or relations do not *in fact* exist and those who think that they *couldn't possibly* exist—a matter to which I will return in IIb.) Presumably, most of us are error theorists about dragon discourse and witch discourse. For neither enormous, fire-breathing lizards that fly nor women with magical powers who consort with the devil have a place in our critically considered ontology. Accordingly, dragon discourse and witch discourse—both of which centrally make reference to such creatures—are *error-ridden* discourses.

[6] In what follows, I shall use <x> to refer to the concept of x, and 'x' to refer to the term.

Although it is tempting to characterise an error theorist as one who holds that *none* of the sentences of a particular discourse are true, it is more accurate to understand her to be targeting only *a relevant subset* of the sentences of that discourse. An error theorist about dragon discourse, for instance, would likely allow that some of its negative ('there are no dragons'), second-order ('dragon discourse is erroneous'), trivial ('all dragons are dragons') sentences could be true; for the assertion of such sentences does not commit the speaker to the existence of dragons. She could also allow that some of the non-atomic sentences of dragon discourse are true. The sentence 'either there are dragons or the Eiffel Tower is in Paris' for example, is true in virtue of the truth of its second conjunct.

Thus, we can, to begin with, take an error theorist about a discourse *D* to be one who denies that *a proper subset of* the sentences of *D*—hereafter, the '*D*-sentences' of *D* —are true.[7] These *D*-sentences include the positive, atomic, first-order, and non-trivial sentences of *D*, and exclude at least some of its negative, non-atomic, second-order, and trivial sentences. Among the *D*-sentences of dragon discourse, for example, would be 'there are many dragons in England', and 'a dragon stole my treasure'. These are the kinds of sentences that the error theorist is targeting when she claims that none of the *D*-sentences of her target discourse are true.[8]

Moreover, we should note that to be an error theorist about a discourse *D* is not *only* to claim that none of its *D*-sentences are true. It is more accurate to say of the error theorist that she takes the *D*-sentences of her target discourse to be somehow *aiming at truth* but systematically failing to secure it. A crowd cheering 'boo' and 'hoorah!' at a football game, for example, would be failing to state truths with their cheers. But that wouldn't license an error theory about football supporter discourse; for the chants of football fans aren't in the market for truth.

In order to advance an error theory about a particular discursive domain then, it is important that the sentences of the target discourse be *truth-apt*; they must be the

[7] Whether the error theorist holds that all of the *D*-sentences of *D* are *false* is another matter. She might prefer to characterise them as neither true nor false (Joyce 2001, pp.6-9).

[8] There are some technical complications here. (See Sinnott-Armstrong 2006, pp.33-7 for discussion.) One such complication concerns the validity of particular inferences made within the discourse. Moral error theorists, for instance, may want to deny the validity of the inference from something's not being morally wrong to its being morally permissible. (See Pigden 2007, Olson 2011a.)

sorts of things that can be true or false. Error theorists typically take the sentences of their target discourse to be *declarative* sentences that express propositions. And they take the utterances of such sentences to be *assertions*—speech-acts that put forward propositions to be considered as true. When I say 'there is a dragon on my doorstep', the error theorist about dragon discourse takes me to be asserting something about the world rather than say, make-believing that there is a dragon on my doorstep. But she holds that a relevant subset of these assertions (those that involve uttering *D*-sentences—hereafter, '*D*-assertions') fail to state truths.

Following (and slightly paraphrasing) Joyce (2001, p.9) then, we can say that to be an error theorist about a discourse *D* is to claim that:

1. The discourse is used assertorically, and
2. The *D*-assertions of the discourse systematically fail to state truths

Thus, to be an error theorist about dragon discourse would be to first understand cartographical markings that warn, 'here be dragons', complaints to the effect that 'there is a dragon wreaking havoc upon our village', and many other claims of dragon discourse to be *D*-assertions that purport to describe something about the world. Secondly, it would be to hold that none of these *D*-assertions state truths because there simply aren't any dragons. The 'error' of dragon discourse thus consists in a kind of reference-failure; dragon discourse is error-ridden because something that the discourse requires in order to successfully refer does not exist.

I have spoken of referring terms like 'dragon'. But error theories certainly aren't restricted to referring expressions. Many are tied to other parts of language. One may be an error theorist about adverbs like 'possibly' and 'necessarily'.[9] Or one may target particular sorts of predicates; 'red' and 'blue', for example (Boghossian & Velleman 1989).[10] The alleged error of colour discourse may be said to consist in its predicates never being literally satisfied (as opposed to its referring terms failing to refer). Whatever the nature of the expressions in question though, we can take a discourse to be error ridden just in case something that that discourse requires in order for (i) its re-

[9] Quine (1953) arguably comes close.

[10] One might even be an error theorist about the referring terms of a discourse without being an error theorist about the discourse in general. Nolan (2014) explores this possibility for moral discourse.

ferring expressions to successfully refer and/or (ii) its positive predicates to be literally satisfied does not exist.

The real philosophical meat lies in specifying just what this 'something' is. To this end, error theorists typically distinguish between the different kinds of commitments that a discourse carries. The error theorists' central claim is that one or more of the *core*—or, alternatively, the *non-negotiable* or *necessary*—commitments of the target discourse are not satisfied in (at least) the actual world. They are careful to note, however, that discourses carry a number of commitments and not all are non-negotiable. One commitment of dragon discourse, for example, is that dragons have a fondness for treasure. But this commitment seems negotiable. Presumably, we would not continue to be error theorists about dragon discourse if we were to discover an enormous, magical, fire-breathing lizard in the Australian outback, even if that creature had little or no interest in gold.

Yet discourses also seem to have *non-negotiable* commitments (Joyce 2001, pp.3-4). The superstitious folk of the 16th century not only believed that there were witches. They also thought that witches had a number of important properties—properties without which a person simply wouldn't count as a witch. These folk didn't merely believe, for example, that witches destroyed farmers' crops; it was also important to them that a witch be someone who use *magical powers* to further such pernicious ends. If someone were to stir up mayhem in a village by pouring too much pesticide onto its maize, we would still be error theorists about witch discourse. To explain this diagnosis, we can appeal to the fact that this pesticide-pourer fails to satisfy what seems to be a necessary commitment of witch discourse; she is unable to boast any magical powers. Were we to use the term 'witch' to refer to the pesticide-pourer, then we would not be competently partaking in witch discourse.

These claims regarding non-negotiable commitments set particular standards that must be satisfied in order to vindicate the discourse under investigation. To say that a fondness for treasure is a *negotiable* commitment of dragon discourse is to say that in order to vindicate dragon discourse, there need not be magical, fire-breathing lizards with a fondness for treasure. (A magical, fire-breathing lizard that shuns wealth will suffice.) Similarly, to say that the existence of creatures with magical powers is a *non-negotiable* commitment of witch discourse is to say that in order to vindicate witch discourse, there must be creatures who, *inter alia*, have magical powers. If there are no such creatures, then witch discourse is an error-ridden discourse.

We can understand the error theorist's non-negotiability claim as a *semantic* or *conceptual* claim; it concerns the meaning of the terms that figure centrally in the target discourse. When error theorists claim that having magical powers is a non-negotiable commitment of witch discourse, they mean to say that ridding witch discourse of this commitment amounts to changing the subject—it amounts to talking about *schwitches* rather than witches.[11]

Spelling out the non-negotiable nature of these commitments in more familiar philosophical terminology is far from straightforward. Sometimes the non-negotiability claim is cashed out in terms of conceptual entailment (Joyce 2008, pp.65-6; Kalf 2013, p.925). For example, we might (simplifying somewhat) take the applicability conditions for <witch> to be as follows: in order to use <witch> correctly, one must apply it to something that is both female and has magical powers. If we were to take either femaleness or having magical powers out of <witch>, then whatever (if anything) we end up referring to simply isn't a witch. This, we might say, is owing to conceptual entailment; the sentence '*x* is a witch' entails '*x* has magical powers' in virtue of the concept <witch> that is expressed by the term 'witch' in the sentence '*x* is a witch'.

An alternative way to explain the nature of a non-negotiable commitment is to appeal to the notion of a presupposition (Joyce 2001, pp.8-9; Kalf 2013, p.926). A statement S_1 presupposes another statement S_2 if denying S_2 renders the kind of talk one is engaged in when uttering statements like S_1 somehow inappropriate. A frightened villager's stating, 'The witch is going to destroy my crops next!' (S_1), for example, seems to presuppose 'there are witches' (S_2). This is because if we were to deny that there are any witches (that is, deny S_2), we would render the villager's talk of witches (that is, S_1) somehow inappropriate.

[11] The schm- prefix is sometimes used to denote a shift in focus away from something real and significant and onto something much less interesting and important (as in Dennett 2006). This is how I intend for the schm- locution here; few (if any) people are interested in schwitches. But the prefix may also be used in a less disparaging way to simply denote a change of subject.

An error theory traditionally proceeds in two stages—one conceptual, one ontological (Joyce 2001, p.5). The conceptual stage involves identifying (at least some of) the non-negotiable conceptual commitments of the target discourse. As we have seen, these commitments provide us with a kind of job description (or partial job description); they tell us (at least in part) what something in the world would have to be like in order for the discourse to be vindicated. During this initial stage, it is common to formulate a statement such as the following:

For every x, x is an F only if Px[12]

So, for the term 'witch' as it was used by the folk of the 16th century, we may have:

For every x, x is a witch only if x has magical powers

Before she can establish her error theory, the error theorist must establish that the discourse fails at the second, *ontological stage*. The ontological stage requires, roughly, that we scan the world for what is needed to vindicate the term under investigation:

Does there exist an x, such that Px?

Or, in the witch case:

Does there exist an x, such that x has magical powers?

If we answer the ontological question in the negative, then we countenance an error theory about the relevant discourse—here, an error theory about witch discourse. We can therefore take error theories to arise in those instances in which the world fails to 'play along' with us; when we speak of and readily believe that there are particular objects, properties, or relations, but no such objects, properties, or relations are to be found.

A small clarification is needed before proceeding. I take the strategy outlined above to be a very popular way of establishing an error theory—especially within philosophy. But I do not wish to rule out other means of doing so. The necessary condition(s) identified by an error theorist need not amount to a conceptual or analytic truth. Someone could it seems, be an error theorist about luminiferous ether

[12] Some formulate the error theorist's conceptual claim as 'for any x, Fx if and only if Px and Qx and Rx' (e.g. Joyce 2001, p.5). This would be a complete job description. But it is not essential for the error theorist to appeal to a set of necessary and sufficient conditions for being an *x* in order to establish her error theory. Her conceptual claim need only appeal to a necessary condition (or a number of jointly necessary conditions) for being an *x*. (See Hussain 2004, p.156.)

even if they couldn't point to a semantically necessary condition for being luminiferous ether that they thought wasn't satisfied.[13] This having been noted, I will mostly restrict my focus in what follows to error theories which proceed by way of the two-step strategy outlined above; these will be the most common in the cases under investigation.

§II. Error Theories in philosophy

The reader will presumably agree that we should be error theorists about dragon discourse—along with talk of witches, astrology, and lucky rabbits' feet. But they are perhaps less likely to be persuaded by many of the error theories put forward by philosophers, which target our talk of beliefs, colours, and moral facts (amongst other things). However, philosophers are sensible folk. They do not deny the existence of these posits of commonsense without good reason. Here, I will explain how error theories in philosophy typically proceed, detailing which moves are commonly involved in their conceptual and ontological stages.

§IIa. *The conceptual stage*

As we have seen, the task of the conceptual stage is to provide us with a job description; to tell us what kind of property, object, or relation we are to look for in order to determine whether a discourse is error-ridden. In formulating that job description, an error theorist usually appeals to the necessary or 'non-negotiable' conceptual commitments of the target discourse. Yet we may very well wonder what *justifies* an error theorist's choice of particular conceptual commitments over others. And we might further wonder just *whose* commitments these are.

Beginning with the latter issue, most error theorists take the necessary commitments of their target discourse to be the (at least tacit) commitments of the linguistic community that employs it. Error theorists about our talk of colours, morality, and beliefs, for instance, are usually concerned with the relevant terms as they figure in

[13] Thanks to Daniel Nolan for bringing this to my attention.

ordinary usage.[14] When they deny that there are any colours, what they are usually denying is that there are any colours in the sense of 'colours' that is used by non-philosophers as well as philosophers. To borrow Jackson's (1998a) terminology, the error theorist's conceptual claim is usually intended to capture the underlying conceptual commitments of 'the folk theory' of colour. It is not typically intended as a claim about some esoteric or unfamiliar understanding of colour.

Admittedly, identifying the conceptual commitments of the folk is far from straightforward. And not all error theorists proceed in the same way. But given that their conceptual claim is often presented as an analysis (or partial analysis) of how a particular term is used and (at least tacitly) understood within a linguistic community, it is common for them to partake in some form of *conceptual analysis*.[15] Conceptual analysis typically involves applying the 'method of possible cases'; we present a number of (partially specified) possible scenarios, and ask whether the term under investigation intuitively applies in those scenarios (Jackson 1998a, pp.31–32). Our answers allow us to surmise the applicability conditions of the term, which in turn helps us to formulate an analysis of it.

Often, conceptual analysis also involves appealing to a set of 'platitudes' or commonsense statements to which any competent user of the term would consent (Smith 1994, Jackson 1998a). Some consider empirical research important here—for instance, studies that test the conceptual intuitions of the linguistic community (Knobe & Nichols 2007). Others deny that empirical studies have any meaningful role to play (Kauppinen 2007). A more moderate position concedes that empirical research will sometimes be relevant for the purposes of conceptual analysis, but denies that it is typically required (Jackson 1998a).[16]

[14] Mackie, for example, takes the conceptual claim of his moral error theory to be appealing to "the traditional moral concepts of the *ordinary man*" (1977, p.35, emphasis mine). Boghossian and Velleman's error theory about colour discourse is similarly concerned with what *ordinary people* are saying when they call "…something red, in an everyday context" (1989, p.100).

[15] The seminal defence of the method of conceptual analysis is found in Jackson's *From Metaphysics to Ethics* (1998a), though he himself does not use it to establish an error theory of any kind. For other defences, see Jackson (1998b), Bealer (1998), and Nolan (2009). For critical discussion, see Laurence and Margolis (2003).

[16] Conceptual analysis seems to be more welcome in some philosophical quarters than others. It is commonly associated with the so-called 'Canberra Planning' methodology for solving location prob-

The output of these methods is (ideally) an analysis that provides the error theorist with the necessary commitments of her target discourse. These commitments are taken to form part of the meaning of the relevant terms, and to reflect how they are used within the relevant linguistic community. However, it need not follow that ordinary persons are able to articulate such commitments explicitly. A conceptually competent user of a discourse will have particular inferential and judgmental dispositions in virtue of which she is best interpreted as taking particular conceptual commitments for granted. But *that she has them* need not be something that she knows (Smith 1994, p.38). A helpful analogy here is grammar (Jackson, Stich & Mason 2009). Although competent speakers of a language have a mastery of its grammar, they need not be able to articulate grammatical rules explicitly. Just as the grammarian's task is to make explicit the rules that are implicit in ordinary speakers' grammatical know-how, it is the conceptual analyst's task to make explicit the structure that our concepts (often tacitly) have.

Most error theorists' conceptual claims are therefore intended to reflect the underlying conceptual commitments of ordinary speakers who make use of the relevant discourse. And they often justify their choice of 'non-negotiable' commitments by conducting conceptual analysis, a process that characteristically involves consulting (hopefully common) intuitions and folk platitudes. It should be noted, however, that not *all* error theorists are interested in the terms of their target discourse as they figure in ordinary usage. Many error theories are aimed at certain sectors of scientific discourse, and so, their proponents are principally concerned with how the relevant terms are used within the scientific community. Edouard Machery's (2009) error theory about concept discourse, for example, targets *the psychologist's notion* of a concept as a body of knowledge that is stored in long-term memory and used by default in higher cognitive processes (e.g., categorisation and inductive reasoning).

lems for various phenomena (mental states, colours, and what-have-you); that is (roughly), systematising commonsense intuitions, building a job description of the phenomenon to be located, and hoping that something will satisfy it, or at least come close.

§IIb. *The ontological stage*

Unlike the conceptual stage, the ontological stage is comparatively straightforward. The question 'are there *x*s?' is relatively to the point. It is at least arguably less divisive than 'what is necessary in order to be an *x*?' Yet it isn't at all clear from what has been said so far how exactly the ontological stage proceeds.

It can proceed in a number of different ways. We might simply scan the world to see whether anything in it is capable of satisfying the job description provided by our conceptual claim. Alternatively, (and more plausibly), we might consult our best scientific theories: do they suggest that something in the world is capable of satisfying the relevant job description—or, do they allow us to make an inference to the best explanation that there are such things? Should we turn up empty-handed, then an error theory looms.

Of course, error theories need not be established by *a posteriori* means. We can plausibly arrive at error theories about talk of square circles without putting on our lab coats. Error theories can also be established using *a priori* methods; we could, from the armchair, simply demonstrate that the terms under investigation (e.g., square circles) are incoherent (Joyce & Kirchin 2010, p.xvi).

Moreover, and relatedly, error theories can differ in the *modal strength* of their ontological claims. The error theorist's ontological claim may be a contingent one: she may simply deny that objects, properties, or relations of kind *K* exist in our world. An error theory about witch discourse plausibly belongs in this category. There are no witches in our world. But there are surely some lurking about in other sectors of possibility space. In some possible worlds, witch discourse succeeds in referring.

Other error theorists may be more ambitious, and argue that properties, objects of relations of kind *K* could not exist in *any* world—that their existence is impossible. Presumably, we are necessary error theorists about talk of male vixens, and intersecting Euclidian parallel lines; such things are not to be found in any possible world.

§IIc. *Controversy*

Earlier, I claimed that the conclusions of error theories may often strike us as surprising. By now, I hope to have shown that these conclusions can be reached by employing transparent and familiar philosophical methodologies—there's no trick up the error theorist's sleeve. Still, these methodologies are not free from controversy. Quite a few challenges have been raised against the arguments for error theories.

To begin with, some philosophers deny that the qualities singled out in the error theorist's conceptual claim should be understood as *jointly necessary* commitments of the relevant discourse—that some property, object or relation must satisfy *all* of them if we are to avoid an error theory. Sometimes, the problem here is a local one; the adversary disagrees with the particular conceptual claim that the error theorist has put forward—that is, with the content of her analysis. Some philosophers, for example, reject the moral error theorist's claim that moral discourse is conceptually committed to the existence of particular sorts of reasons (e.g., Finlay 2008, Shafer-Landau 2005).[17]

More commonly though, the problem reflects a deep, methodological disagreement with the error theorist; quite a number of philosophers are sceptical about claims regarding the non-negotiable commitments of a discourse. This scepticism can be cashed out in different ways. Some hold that *every* conceptual commitment of a discourse is negotiable in principle (following Quine (1951)). More moderately, others argue that very few commitments of a discourse (far fewer than error theorists suppose) are non-negotiable. As such, we need only require that some object, property, or relation in the world exhibit *most of* the qualities listed in the conceptual claim in order to avoid an error theory. Error theorists simply ask too much in insisting that something must satisfy *all* of them. Put differently, there need not be anything in the world that meets the error theorist's job description *perfectly* in order for the discourse to be vindicated; often, an imperfect deserver can be given the job, and this will suffice to avoid an error theory (Lewis 1994, pp.415-6; Jackson 1998a, p.35; Braddon-Mitchell 2003, 2006). Yet another source of scepticism comes from those who deny that the discourse in question was assertoric to begin with. Many expressivists about moral discourse, for instance, will deny that moral claims were ever in the market for truth and falsity.

Like anything else in philosophy, there are some who are likely to see the error theoretic project as misguided from the outset. Some theorists hold that philosophi-

[17] As we shall see (in chapter 1), the reasons in question are *categorical* reasons, the force and legitimacy of which holds independently of an agent's unique ends or goals.

cal reflection can offer us no good arguments against the existence of certain kinds of objects, properties or relations—for that, we need to appeal to scientific considerations (e.g., Horwich 1998, p.88). Others deny that the meaning of a term should be understood by appeal to some analysis of it, as the error theorist's conceptual claim seems to suggest (Kripke 1972, Putnam 1975a, Burge 1979). Hostility towards error theories can also be rooted in Moorean considerations; we might think that error theories amount to a denial of commonsense, and that commonsense claims or "Moorean truisms" are "…more certainly truths than any evidence that is brought against them" (Armstrong 2006, p.160.)

I will not attempt to address these positions and the arguments made by their advocates head-on. As I have indicated, my project assumes that a moral error theory succeeds on its own terms. (Though I will draw attention to the considerations in its favour in chapter 1.) That said, it's worth noting that the various lines of resistance canvassed above are not always as powerful as they may first appear. *Contra* Horwich, it is highly plausible that philosophical reflection *can* produce good arguments against the existence of certain kinds of objects, properties, or relations. One could, it seems, mount a decent defence of an error theory about monotheistic discourse on purely logical grounds; by claiming that it is logically impossible for anything to be omnipotent, for instance (Daly & Liggins 2010, p.214). Likewise, Armstrong is right to note that we should afford due respect to commonsense. But commonsense surely doesn't have *absolute* authority in philosophical enquiry (Lewis 1986, p.134). We can, after all, be radically mistaken about the nature of the surrounding world; what strikes us as obvious is sometimes false. Sometimes, the evidence brought to bear against commonsense beliefs can be more powerful than the case in their favour. We may even be in a position to offer a 'debunking explanation' of these commonsense beliefs, which accounts for the widespread error (a point to which I'll return in chapter 1).[18]

Though it is not my foremost intention in this work to defend moral error theory against the challenges that have been brought to bear against it, I believe that my arguments do have implications for its acceptability. As I will suggest (in chapter 2),

[18] For further discussion, see Daly and Liggins (2010), who also defend error theories against arguments that appeal to norms of charity.

hostility towards error theories often derives not merely from doubts concerning their methodologies, or worries about their counterintuitive conclusions, but from a fear of their *practical consequences*. Some are concerned with the real-life ramifications of adopting a moral error theory. I will propose that we can make some headway in assuaging this concern by providing a satisfying answer to the WNQ for moral discourse.

§III. Answering the WNQ

The WNQ is the question regarding what we ought to do with an error-ridden discourse. It has the following, conditional structure:

> **The 'What Next?' Question**
> If an error theory in a particular discursive domain is true, and moreover, we believe it is true, then what ought we to do with the error-ridden discourse?

As I will explain (in chapter 2), we assess any candidate answer to the WNQ by looking at the extent to which it makes for a fitting or appropriate response to the case under consideration. This will depend, amongst other things, upon the extent to which that answer promotes particular interests, or enables us to achieve certain ends.

Given that the relevant interests and ends are likely to vary depending upon the nature of the erroneous discourse, we should expect WNQs in different discursive domains to invite different solutions. Sometimes, continuing to make use of an error-ridden discourse might be the best option available to us. We might, for example, be convinced by Buddhist arguments against the existence of an enduring self, but continue with our personal identity talk in any case because we simply could not make do without it. However, in many cases, the recommendation that we persist with an erroneous discourse won't make for sound advice. Once they discovered that there were no witches, it was surely a *good* thing for the superstitious folk of the 16th century to cease with their witch discourse, together with the public burning ceremonies that talk of witches encouraged. (It was an especially good thing for those who were erroneously accused of witchcraft.)

Broadly speaking, there are four kinds of answers to the WNQ that follows an error theory: revisionism, abolitionism, fictionalism, and conservationism.[19] As we will see, these answers have been proposed in response to the WNQ for moral discourse. I shall discuss these proposals in more detail in chapter 2. We can content ourselves with a quick preview for now.

Abolitionists recommend that we cease using the relevant error-ridden discourse. The position can be motivated by an appeal to epistemic considerations; perhaps discourse *D* and the ontology to which it is committed must stand or fall together. If there are no *x*'s, then perhaps we ought to stop talking about them, lest we continue to believe and assert falsehoods. Abolitionist proposals can also be rooted in practical concerns. Their proponents often argue that employing discourse *D* incurs a great number of real-life costs. As such, we would do better not to preserve *D* in any form—revised, fictionalised, or otherwise. Many abolitionists are motivated by both sorts of considerations; in the event that a discourse can boast neither truth nor usefulness, they think we have doubly good reason to get rid of it. It has been recommended, for example, that we cease to use the term 'Boche' on ethical grounds (because it licenses unethical inferences) as well as epistemic ones (because its constitutive inference rules are unreliable) (Dummett 1973, p.454). Error theorists who take moral discourse to be on-balance harmful have also motivated abolitionism by appealing to both practical and epistemic considerations (Hinckfuss 1987; Garner 1994, 2007; Greene 2002; Burgess 2007; Moeller 2009; Marks 2013a; Ingram 2015).

A second strategy is *revisionism*, which consists in somehow modifying the error-ridden discourse. One means of doing so is to rid the discourse of its problematic conceptual commitments. For example, (and simplifying), if there exists no *x* such that P*x* and R*x* and Q*x*, but there does exist an *x* such that P*x* and R*x*, then we might revise *x*-discourse such that it is no longer committed to the first conjunction, but only to the second. Another kind of revisionist move is to change how the discourse *is used*. We might, for example, refashion our talk of *x*s such that it consists in giving expression to non-cognitive attitudes rather than beliefs. Both moves have been explored in response to moral error theory (the former by Lutz (2014), the lat-

[19] There is another option—'propagandism'—that I shall mention and set to the side in chapter 2.

ter by Köhler and Ridge (2013) and Svoboda (2017)).[20] Revisionist proposals align with our epistemic preferences. Unlike the *D*-sentences of discourse *D,* some of the *D**-sentences of discourse *D** will be true, or won't be in the market for truth. Either way, we won't be expressing false beliefs with our *D**-assertions. Revisionism often aligns with our practical interests as well, insofar as the revised, substitute discourse confers similar advantages to the original, erroneous discourse.[21]

A third proposed answer to the WNQ is *fictionalism*. Fictionalists think that it is (on-balance) in our practical interests to continue to employ an erroneous discourse in some form; even if the discourse cannot boast truth, it may nonetheless be far too useful to dispense with altogether. Fictionalists don't, however, recommend that we preserve an error-ridden discourse in *its current form*. (That would be to hold onto false beliefs, after all.) Instead, they advise us to preserve the discourse in the form of a useful fiction. This typically involves substituting fictive attitudes for our full-blooded (and false) *D*-beliefs. On some fictionalist proposals, doing so allows our *D*-assertions to be assertions that state truths within a fiction rather than literal truths. The fictionalist manoeuvre is a very popular one; it has been explored in response to error theories about numbers (Field 1980, Balaguer 2009, Leng 2010), colours (Boghossian & Velleman 1989), and possible worlds (Rosen 1990, Brogaard 2006), as well as moral error theory (Joyce 2001, 2005; Nolan Restall, and West 2005).[22]

Finally, there is the *conservationist* option. Like fictionalists, conservationists take the relevant error-ridden discourse to be incredibly useful. But they part ways from

[20] I should note that Köhler & Ridge's proposal is in fact intended as a response to an error theory about *all* normative claims—not only moral ones.

[21] Of course, given the error theorist's non-negotiability claim, revisionist proposals effectively change the subject. So revisionism might be better thought of as *replacement-ism*. But this doesn't prevent us from assessing the proposal. We can still debate the merits and demerits of using schmorality rather than morality.

[22] Some fictionalists advise us to adopt non-cognitive attitudes of make-belief towards moral propositions (e.g., Joyce 2001). Given this, it might seem that there is not much light between certain varieties of fictionalism and certain varieties of revisionism; for both work by removing the aspiration to truth from a discourse. Indeed, Joyce characterises his fictionalist proposal as recommending that we *become* non-cognitivists with respect to moral discourse (2001, pp.200-1). My main reason for putting these options into different categories concerns the distinct *kinds* of non-cognitivist attitudes that the fictionalist recommends. As we shall see (chapter 5), there are some interesting properties associated with attitudes of make-belief in particular.

the fictionalist in recommending that we hold onto *D* in its current, *erroneous* form. This involves continuing to make *D*-assertions and preserving our false *D*-beliefs. Interestingly, conservationists advise us to believe the error theory about *D* as well—though they usually suggest that we only attend to this belief in a specified range of contexts (See for example, Olson 2014.) Advocates of such proposals often have some story to tell regarding how such doxastic back-and-forth is possible, and why our interests are best served by preserving the erroneous discourse in its current form rather than revising or fictionalising it.

§IV. Game Plan

The thesis begins with a chapter devoted to explaining the moral error theory and the arguments in support of it. The first task of the chapter will be to identify the distinctive meta-ethical assumptions that underwrite the position. The second will be to offer a condensed history of moral error theory. I then move on to explore Mackie's (1977) seminal articulation of the view, and pry apart his many different arguments for it. Following that, I distinguish some important developments of Mackie's moral error theory. One (that of Olson 2011a, 2014) is primarily rooted in concerns having to do with the non-naturalness and/or explanatory impotence of moral properties. Another (that of Joyce 2001) derives from concerns regarding the special kind of reasons that morality purports to supply. I shall argue that the latter variety of moral error theory is more plausible, and its success is what shall be assumed for the remainder of the work.

The job for Chapter 2 will be to clarify and motivate the project of providing an answer to the WNQ for moral discourse. I begin by explaining what kind of question the WNQ for moral discourse is; whether it is (or can be) normative in character, and to whom an answer to it is addressed. I then motivate the task of answering the WNQ for moral discourse, and outline the answers to this question that are available in the current literature. I conclude by pointing towards the philosophical pay-offs of the project.

The remainder of the thesis explores different answers to the WNQ for moral discourse. Chapter 3 considers the abolitionist option. After having pinned down the variety of moral abolitionism of interest, I examine the case in favour of the abolitionist's proposal, doing some work to construct the strongest versions of her arguments. I will argue that these do not support the abolitionist's central claim that we would be better off without morality, and will outline some promising means by which we could control for the costs that she brings to our attention. If I am right, then the abolitionist's arguments do not so much support abolitionism as they support reform. I will further propose that abolitionism is likely to be an infeasible option going forward; ridding ourselves of morality may very well be something that we cannot do.

Chapter 4 assesses revisionism. Following a discussion of concepts and conceptual change, I distinguish our revisionist from a certain kind of success theorist: the reformist. There are important differences between these two projects. But as we shall see, there are also some important connections. In the critical discussion, I develop a number of arguments against revisionism, all of which are intended to show that her 'schmorality' is likely to be a rather poor stand-in for morality. Some of these problems are local problems for the revisionist's proposed schmoralities, which seem unlikely to be very useful to us. Other problems are more global, and suggest that any candidate schmorality is likely to fall short of giving us what we want. I shall argue that morality—unlike science—is not plausibly a domain in which concepts can be substantially modified and continue to be put to good use. Scientific concepts are far more amenable to (fruitful) modification than moral concepts, and this is owing to the distinctive functions of scientific discourse.

Chapter 5 explores the prospects of moral fictionalism. Following an overview of fictionalism more generally, I consider two varieties of moral fictionalism. I then move on to assess a number of challenges that have been directed against each variety. In my view, these challenges should only suggest to us that moral fictionalism needs to be further refined—not that it ought to be abandoned. The remainder of the chapter outlines a more pressing problem for the moral fictionalist. I will argue that the fictional attitudes that she intends to substitute for our error-ridden moral beliefs are neither a stable enough nor a strong enough basis for securing the many practical benefits of our error-ridden moral practices.

In chapters 6 and 7, I argue that conservationism is the most fitting response to the WNQ for moral discourse. The job of chapter 6 is theory-building. Here, I will begin by considering the most developed variety of conservationism—that of Olson (2014)—and drawing attention to where there is room for improvement. I then take these lessons on board, and develop my own brand of conservationism, which preserves the benefits of Olson's approach while filling in some important gaps. Among my primary tasks will be (i) to motivate taking the conservationist's attitudes towards both first-order moral propositions and the moral error theory to be beliefs, (ii) to make sense of the idea that a moral error theorist could attend to her beliefs in moral propositions in some contexts and to her belief that the moral error theory is true in others, and (iii) to offer a suitable justification for overriding the presumption against intentionally cultivating false beliefs. In chapter 7, I switch to the defensive, responding to some challenges and potential concerns. Among these are problems with taking attitudes that seem insensitive to evidence to be beliefs, the general disadvantages of cultivating false beliefs, objections to pragmatic (as opposed to epistemic) reasons for belief, and doubts pertaining to the feasibility of the proposal. I conclude the thesis with a summary of the discussion.

CHAPTER 1

(Almost) Everything You've Ever Wanted to Know About Moral Error Theory

For the purposes of this thesis, I will be assuming that a moral error theory is true. So it's best to begin by getting clear on what it is exactly that I will be assuming. The primary goals of this chapter will be to explore the arguments in favour of moral error theory, and to specify which variety of the position will form my background assumption for the remainder of this work.

Moral error theory is a position in *meta-ethics*, a field of philosophy that is concerned, *inter alia*, with the nature of moral properties, language, and judgments. My first task in this chapter will be to locate the moral error theorist in meta-ethical space—to explain the distinctive meta-ethical assumptions that underwrite her position (§1.1). Following that, I offer a condensed history of moral error theory (§1.2). (Discussing Mackie alone would make for a rather thin historical background.) I will also compare and contrast moral error theories within the Mackiean tradition to those outside of it. The latter task will help us to better distinguish the moral error theories that shall form our focus from other varieties.

With that background out of the way, I explore Mackie's (1977) seminal articulation of moral error theory in his *Ethics*: *Inventing Right and Wrong* (§1.3).[23] As we shall see, Mackie develops a number of arguments in support of the position. For the most part, I will restrict my critical focus to what I take to be the most important of these: the metaphysical argument from queerness.

There have been a number of developments of Mackie's moral error theory. One such development—that of Olson (2011a, 2014)—is primarily rooted in concerns having to do with the non-naturalness and/or explanatory impotence of moral prop-

[23] All unattributed citations henceforth to Mackie in this thesis will be to his *Ethics* (1977).

erties. In my view, however, a more philosophically plausible development of Mackie's error theory is one rooted in concerns having to do with the special kind of *reasons* that morality purports to supply.[24] This is the variety of moral error theory articulated and defended by Richard Joyce (2001).[25] The task for §1.4 will be to draw out the finer lineaments of Joyce's moral error theory, whose success shall be assumed for the remainder of the work.

§1.1 MORAL ERROR THEORY IN META-ETHICS

Before we examine the arguments for moral error theory, it will be useful to first get clear on some of the philosophical assumptions that underwrite it. As one might expect, error theory is not an especially popular position in meta-ethics, and not all of the moral error theorist's foundational assumptions are shared by other philosophers. I will therefore begin by locating the moral error theorist in meta-ethical space, distinguishing her from non-cognitivists and success theorists in particular (§1.1.1). I will also explain the different positions that meta-ethicists (error theorists included) can occupy regarding the truth-conditions of moral claims (§1.1.2). Introducing these distinctions will also be helpful in supplying the reader with some important conceptual and terminological background for the remainder of the work.

§1.1.1 *Cognitivism & Non-cognitivism, Success & Error*

As I have noted (§I), error theorists take the speech-acts of their target discourse to be assertions that aim at truth but systematically fail to secure it.[26] This makes our moral error theorist a *cognitivist.* Cognitivism is a thesis about the kinds of mental attitudes that are involved when we take something to have a particular moral quality,

[24] To be sure, Olson appeals to the problematic nature of moral reasons as well. But as we shall see, concerns about metaphysical queerness play a far more central role in his case for moral error theory than they do in Joyce's arguments.

[25] All unattributed citations henceforth to Joyce in this thesis will be to his *The Myth of Morality* (2001).

[26] As I will clarify shortly (§1.2.2), the 'systematically' qualification is important here; it is needed to distinguish the moral error theorists who shall form our focus not only from other kinds of moral error theorists, but also from certain kinds of success theorists.

and the kind of language that is at play when we declare that something has that quality—when we judge or declare that some action is morally wrong, for example.

According to the cognitivist, the attitudes involved in our judging something to have a particular moral quality (e.g., taking it to be wrong) are *beliefs*. Since beliefs are the sorts of attitudes that are capable of being true or false, moral judgments are truth-apt on the cognitivist picture. And since assertions are the sorts of speech-acts that give expression to beliefs, moral utterances take the form of *assertions*.

Given the above package of claims, it is natural for cognitivists to take moral judgments to be *descriptive* or *representational*. When we believe or assert that φ-ing is wrong, the cognitivist takes us to be aiming to describe or represent certain features of the world. Since describing something involves attributing certain properties to it, cognitivists take someone who utters an indicative sentence of the form '*x* is F' (where 'F' is a moral predicate) to be asserting that *x* has the property of being F. To assert 'torturing kittens for fun is wrong', for example, is to attribute the property of wrongness to the act of torturing kittens for fun. The cognitivist takes these predicative moral sentences to express propositions, and so, she takes the truth or falsity of our moral assertions to depend upon the truth or falsity of the propositions that they express.

The moral error theorist is far from being a lone cognitivist. (Cognitivism is the majority position in meta-ethics (see Bourget & Chalmers 2014 p.476).) But she is the only cognitivist who takes moral claims to be systematically false.[27] All other cognitivists are *success theorists*; they think that at least some moral claims are *true*. Success theory is thus the complement of error theory with respect to cognitivism.

[27] As I suggested in the introduction, there are some complications here. For one thing, a moral error theorist may want to allow that some negative ('it is not the case that lying is wrong') non-atomic ('either lying is wrong or the Eiffel Tower is in Paris'), trivial ('wrongness is wrongness') moral propositions could be true. That is to say, she might want to restrict her claim to (what I earlier called) the '*D*-sentences' of moral discourse. She might also want to deny the validity or appropriateness of certain inference rules (e.g., those that allow one to infer from something's not being wrong that it is permissible). For discussion, see Joyce (pp.6-7), Sinnott-Armstrong (2006, ch.3), Pigden (2007), Joyce & Kirchin (2010, p.xii), and Olson (2011a). I will put these complications to the side in what follows. For ease of expression, I shall simply speak of our moral beliefs, judgments, or claims being systematically false, or none of them being true.

Like other cognitivists, moral error theorists part ways from *non-cognitivists*, who think the attitudes involved in our taking something to have a particular moral quality are non-cognitive; for example, emotions (Ayer 1952), prescriptions or imperatives (Hare 1952), acceptances of norms (Gibbard 1990), or fictional attitudes (Kalderon 2005). Non-cognitivists take moral utterances to be expressions of these attitudes. Unlike utterances that give expression to beliefs, these utterances are not in the market for truth and falsity.

This is, admittedly, *a simplification* of non-cognitivism. Many contemporary non-cognitivists deny that we must choose between the thesis that moral judgments involve only non-cognitive attitudes and the thesis that they involve only beliefs. In Michael Ridge's (2006) words, one can be *ecumenical*, and take moral judgments to implicate both sorts of attitudes. But a simplified characterisation of non-cognitivism will suffice for my purposes. In the interest of getting clear on those distinctions that are most relevant for understanding the moral error theory, we can understand non-cognitivism as I have described it above.

§1.1.2 *The truth-conditions of moral judgments*

Although cognitivists agree that moral judgments are truth-apt, they disagree upon which features of the world *determine* their truth and falsity. They disagree, in other words, upon the truth-conditions of moral claims. Geoff Sayre-McCord (1986, p.10) offers a helpful taxonomy of the different stances that one might take on the matter:

> **Subjectivism**
> The truth of moral judgments depends upon facts about some individual.
>
> **Inter-subjectivism**
> The truth of moral judgments depends upon facts about some group of individuals.
>
> **Objectivism**
> The truth of moral judgments is independent of facts about any individual or group of individuals.

Note that the facts about individuals or groups of individuals need not only include facts about how those individuals or groups *actually are*. Lewis's (1989) variety of subjectivism, for instance, takes moral truths to depend upon what each individual would desire to desire under idealised reflection. Similarly, some intersubjectivists understand moral truths to depend upon the standards upon which idealised agents would converge (e.g., Scanlon 1998).

Objectivists take the truth-conditions of moral judgments to be mind-*independent*. That is to say, they take moral truths to hold independently of us—of our whims, our desires, and our preferences.[28] On this view, moral truths are not made true "…by virtue of their ratification from within any given actual or hypothetical perspective" (Shafer-Landau 2003, p.15). In meta-ethical circles, objectivists who are success theorists are commonly known as *moral realists*. (This distinction between objectivists and non-objectivists will later help us to distinguish different types of moral error theorists.)

Our moral error theorist clearly parts ways from the success theorist in denying that any of our moral judgments are true. But she can agree with (at least some of) them about *what it would take* for these judgments to secure truth—a moral error theorist can certainly take a stance on the *truth-conditions* of moral claims. For example, (and as I will later explain), Mackie seems to concede to the moral realist that in order for our moral judgments to be true, there would have to be particular kinds of mind-independent properties. Both parties are therefore objectivists. But Mackie denies that any of these moral judgments *are* true, and so, he is no realist.

§1.1.3 *Moral error theory in meta-ethics: a summary*

We have now located the moral error theorist in meta-ethical space. Unlike non-cognitivists, the moral error theorist thinks that the attitudes involved in our taking something to have a particular moral quality are beliefs, and that our moral utterances are assertions. And although the moral error theorist parts ways from success theorists in denying that any of our moral judgments are true, she can nonetheless agree with (at least some of) them regarding what it would take for those judgments to be true.

[28] Moral realists need not claim that moral facts do not depend upon our attitudes or feelings *in any respect*. Many will want to claim that the wrongness of jabbing a person with a sharp stick can depend upon the fact that it caused them to feel pain. (See Shafer-Landau 2003, p.15.)

Before proceeding, it's worth noting that moral error theory does not always sit comfortably with meta-ethicists, many of whom are concerned to *vindicate* moral discourse; that is, to vindicate the highly intuitive idea that some things *really are* morally right or wrong. One concern for the error theorist pertains to the practical consequences of denying that this is so. As Joyce points out, there are "…worries about what might *happen* if a moral error theory were to become widely accepted as true" (p.231, emphasis in original). One task of this thesis will be to quell such concerns. Contrary to what we might fear, life after moral error theory would not be all that bad—or so I shall argue.

It's also worth noting that widespread resistance to moral error theory among moral philosophers shouldn't necessarily raise doubts about the cogency of the position. As Joyce and Kirchin speculate,

> The real explanation for the dearth of real-life moral skeptics plying their wares in the philosophical marketplace may be nothing more insidious than a natural process of self-filtration: Those who are drawn to moral philosophy sufficiently to publish works on the topic are more likely than not to be antecedently hostile towards moral skepticism. By analogy, consider theology. One need not believe in God in order to be a capable theologian, but how many atheistic theologians does one really expect to find in the profession? (2010, p.ix)

Following Joyce and Kirchin, there may be a strong selection bias at work here. One should not be so quick to interpret a near-consensus among (what is likely) a highly-biased sample of philosophers as evidence against the plausibility of moral error theory.

§1.2 MORAL ERROR THEORY: A GENERAL OVERVIEW

My primary focus in this thesis will be moral error theories in the Mackiean tradition: those which build upon Mackie's statement of the position in his *Ethics*. However, Mackie was certainly not the first to suspect that not all was well with morality. Nor do his arguments represent the only way in which one might go about establishing a moral error theory. Here, I briefly consider some possible precursors to Mackie (§1.2.1), and explore other roads to moral error theory (§1.2.2). I mention the latter largely for comparative purposes; they are not roads that we shall travel.

§1.2.1 *A selective history*

Philosophers seem to have flirted with the idea of moral error theory long before Mackie. Indeed, some take David Hume to have been an early supporter of the position (e.g., Olson 2011b, pp.27-30; 2014, ch.1). It is debatable whether moral error theorists can confidently claim Hume as one of their own. But many do take him to have offered a promising *explanation* for our widespread error—how it is that we may have come to falsely believe that some things really are right and wrong (e.g., Joyce 2006, pp.123-6; Olson 2011b pp.33-4). The Humean hypothesis, roughly, is that we experience moral properties as objective features of the world as a result of 'projecting' our sentiments outward. When we feel for someone who suffers, for example, we project that feeling onto the world and perceive their suffering as something that *demands* our sympathy. And so, we form the relevant moral judgment; it seems to us to be a fact of the world that it is *morally wron*g to cause suffering (see Joyce 2006, pp.125-6).

This Humean hypothesis is commonly thought to be congenial to moral error theory. Error theorists are famously met with a Moorean challenge; given the counterintuitive nature of their claim that nothing is really right or wrong, their position (the thought goes) must surely be false—something must have gone awry somewhere. (See Huemer 2005, pp.115-7; Armstrong 2006, p.160.) But if the moral error theorist can offer a plausible explanation as to *how* our false moral beliefs have "become established and [are] so resistant to criticism", then she may have the resources to *explain away* the intuitions that underlie the Moorean challenge (Mackie, p.42; see also Joyce, ch.6; Olson 2011a).

Karl Marx (1848/1977) and Friedrich Nietzche (1887/1994) may also be counted among the moral error theorist's predecessors. Indeed, some (e.g., Pigden 2007) interpret Nietzsche as a moral error theorist *cum* fictionalist. But it remains debatable whether his and Marx's respective projects do amount to second-order critiques of morality. Perhaps these works are better thought of as first-order criticisms of particular moral belief-systems (Olson 2014, p.16).

Another figure worthy of mention here is Gilbert Harman (1977), who develops (without ultimately endorsing) an argument from explanatory irrelevance.[29] Harman claims that moral facts do not seem to feature in the best explanations of our moral judgments. In order to explain these judgments, it seems that we need only appeal to the relevant natural facts that prompt them (e.g., some children setting fire to a cat), together with facts about our moral sensibilities. Since moral facts themselves appear to be explanatorily irrelevant, we appear to have no good evidence for their existence. Harman's argument from explanatory irrelevance has been reserviced by some moral error theorists (e.g., Olson 2014, ch.7; Joyce 2006, ch.6).

I have been rather brief here, largely for considerations of space. What the above discussion should suggest to us, though, is that there has been sympathy for moral error theory and related positions for quite some time. Historically, many philosophers have taken seriously the idea that there are no moral facts. So should we.

§1.2.2 *The many roads to error*

Moral error theorists are something of a mixed bag, so it's best to avoid lumping them all together. There are other possible routes to moral error theory aside from those that we see in the Mackiean tradition. One such route could proceed on religious grounds; an error theorist could take moral discourse to presuppose the existence of a divine being, and then regard a statement of atheism as a statement of moral error theory.

Alternatively, an error theorist might be tempted by the thought that some of our moral claims operate upon false metaphysical assumptions about agents and their place in the world. Our attributions of moral responsibility, for example, might falsely presuppose that we have genuine free will (Pereboom 2001). One might even endorse an error theory that is restricted to the referring terms of morality (Nolan 2014). A distinctive mark of these latter two possibilities is that they only affect *certain kinds* of moral judgments, leaving others unscathed. In this respect, they crucially dif-

[29] Harman is no error theorist; for he thinks that there *are* (relational) moral facts (1977, p.132). His aim, as I understand it, is not to establish that evidence for the existence of moral facts is unavailable to us—only that it is needed.

fer from the moral error theories that shall form our focus, which attribute a *systematic* error to moral discourse.

It's worth pausing to note this feature of the moral error theories that shall be our primary concern; for it helps us to distinguish them not only from other varieties of moral error theory, but also from positions that can be adopted by success theorists. There is a sense in which the committed utilitarian regards the moral judgments of her deontologist colleagues as systematically mistaken; she thinks that many of their claims about what is right or wrong are false. But the utilitarian still thinks that some things *are* right and wrong. She may even be concerned that her deontologist friends are being led morally astray. Our moral error theorist, by contrast, denies that there is *any such thing* as moral rightness or wrongness. The error that she identifies is systematic in a far more deep and pervasive sense; it prevents the property of moral rightness or wrongness from attaching to *anything*.[30] (See Hussain 2004, pp.158-60.)

Our moral error theorist is therefore more ambitious than those who restrict their target to certain sectors of moral language. But she is less ambitious than others. Normative error theorists (e.g., Streumer 2011) deny that *any* normative claims—and, *a fortiori,* any moral claims—are true. So they adopt a moral error theory by implication (Joyce & Kirchin 2010, p.xiii). The moral error theorist with whom we shall be concerned is different: she "…thinks that there is something especially problematic about morality, and does not harbor the same doubts about normativity in general" (Joyce & Kirchin 2010, p.xiii; see also Joyce 2014, p.844). (Whether or not a moral error theorist *can* coherently indict moral discourse while leaving the rest of normative discourse intact is admittedly a contentious issue—one that I shall address in §2.1.)

The above discussion distinguishes moral error theorists in the Mackiean tradition from their close cousins. But it's also worth distinguishing these error theorists from moral *sceptic*s. The sceptic doesn't necessarily think that our moral claims are system-

[30] I will mostly characterise moral error theory by appealing to so-called *thin* moral terms such as 'good', 'bad', 'right', and 'wrong'. But it's worth noting that the moral error theory may very well apply to *thick* moral terms (e.g., 'just', 'cruel') as well. See Joyce (p.176).

atically false; she merely claims that we lack moral knowledge.[31] Though there are a number of different routes to moral scepticism (Sinnott-Armstrong (2006) offers a helpful overview), evolutionary debunking arguments have become increasingly popular in recent years. Debunkers claim, roughly, that we can supply an evolutionary explanation of our moral faculties which appeals only to their adaptive utility and nowhere to their truth. Given this, they argue that we should be far less confident that our moral beliefs are indeed true (Joyce 2006, Street 2006).

Many moral error theorists take these debunking genealogies to be very friendly to their proposal (e.g., Olson 2014, ch.7; Joyce, ch.6). Such genealogies purport to supply us with an explanation of our moral judgments and practices that nowhere presupposes the existence of any moral facts. Loosely following Harman, some error theorists have appealed to this explanation in support of their contention that moral facts are explanatorily irrelevant and should be eliminated from our ontology (e.g., Olson 2014, p.147; cf. Joyce, p.168).

§1.3 Moral facts: they're not here, they're queer

Moral error theory arguably first gained traction in contemporary philosophy when Mackie articulated the position in his *Ethics* (1977).[32] However, Mackie presented quite a number of arguments for moral error theory in his seminal work on the topic, and some will be more important for our purposes than others. I will briefly outline these arguments, commenting along the way on their relative importance in establishing a moral error theory (§1.3.1). The majority of the discussion will be devoted to the metaphysical argument from queerness, which, to my mind, is by far the most important among these (§1.3.2).

Following that, I explore Olson's (2011a, 2014) development of Mackie's error theory (§1.3.3). Like Mackie, Olson emphasises the problematic kind of *queerness* that

[31] Though not every moral sceptic need be a moral error theorist, we would expect all moral error theorists to be moral sceptics. If there are no moral truths, then, presumably, there is no moral knowledge either. Some moral error theorists (e.g., Joyce 2006), pursue the sceptical project (largely) independently.

[32] Mackie also discussed moral error theory in his 'A Refutation of Morals' (1946). But philosophical discussion of the position was only really ignited following the publication of his *Ethics* (1977).

moral properties would have. On closer inspection, however, these queerness-based arguments do not seem to me to be especially plausible (§1.3.4). I will therefore ultimately distance myself from moral error theories which proceed by raising queerness concerns, aligning myself instead with Joyce's development of Mackie's error theory (to be covered in §1.4), which strikes me as more promising.

§1.3.1 *Mackie's many arguments for moral error theory*

Let's begin with Mackie's well-known *argument from disagreement*. The backbone of this argument is the observation that people's moral values differ (or seem to differ) in significant ways—especially among persons who are temporally and spatially distant from one another. On that much, philosophers tend to agree.[33] What they disagree upon is the precise ambition of the argument that follows, together with its role in establishing a moral error theory.

It is common to understand Mackie to be using his initial observation to put forward an inductive argument. (See for example, Brink 1984, p.115; Olson 2014, p.73.) "The actual variations in…moral codes", Mackie tells us, "are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values" (p.37). Thus, Mackie seems to think that there are two competing explanations available for moral disagreement. Either (i) moral truths are determined by individuals' attitudes and commitments, and moral disagreement issues from different attitudes and commitments, or (ii) there are objective moral truths, and moral disagreement persists because some have failed to obtain knowledge of them. Mackie argues that the former explanation is to be preferred: moral disagreement is better explained by the hypothesis that moral truths reflect different lifestyle choices than by the hypothesis that it issues from differential access to moral truth. (One might think that the conclusion of this argument is moral relativism rather than moral error

[33] That is to say, they agree that Mackie is claiming to make such an observation. Not everyone agrees that moral disagreements *do* in fact issue from differences in moral values. One might offer a "defusing explanation", which identifies the source of these disagreements in bias, irrationality, or disagreements over non-moral facts. (See Doris and Plakias (2008) for discussion.) I will discuss some of these defusing explanations in §3.2.2.

theory. But as we shall see, Mackie takes morality to be conceptually non-relative; in our earlier terms, he is an *objectivist*. Given this, he does not view moral relativism as a vindication of moral discourse)

Though the above interpretation is rather common, it is certainly not the only one. Others understand Mackie to be putting forward the conceptual claim that ordinary moral discourse is underwritten by the expectation that we will converge upon our moral judgments, following suitable argument and reflection (Lillehammer 2014, p.97; cf. Streumer 2011). His pointing towards widespread moral disagreement may therefore be intended to show that this is an expectation made in vain.

Mackie's other arguments for moral error theory include *the epistemic argument from queerness*s, and *the argument from queer supervenience*. The first of these is intended to establish that given the metaphysically queer character of moral properties, we would require an equally queer epistemic faculty for detecting them—a faculty that, Mackie argues, we have no good reason to attribute to anyone (p.41). The second seeks to show that the supervenience relation between non-moral properties and queer moral properties would similarly be queer. It is because we operate under the assumption that moral properties supervene upon non-moral properties—that there can be no difference in the moral features of a person, action, or state of affairs without a non-moral difference—that we can say of an action that it is wrong because it is an act of gratuitous killing. Yet what in the world *explains* this supervenience relation—"just what *in the world* is signified by this because?" Mackie asks (p.41). He doesn't think that any plausible answer is forthcoming.

As should be clear, both of these latter arguments are heavily premised upon *the metaphysical argument from queerness*; it is only once Mackie takes himself to have established that moral properties themselves are queer that he is in a position to argue that this implicates a queer supervenience relation, together with a queer epistemic faculty for detecting moral properties. Simply put, the metaphysical argument from queerness is "load bearing" (Joyce & Kirchin 2010, p.xvii). This speaks to the importance of the metaphysical argument from queerness, which, to my mind, is the most important component of Mackie's case for moral error theory. Not only does Mackie himself understand this argument to be "more important" and "certainly more generally applicable" (p.38), but it also forms the foundation of many of his other arguments.

The argument from disagreement, by contrast, has questionable ambitions, and its role in establishing moral error theory remains unclear. In particular, it remains unclear whether the argument from disagreement is a mere handmaiden to the argument from queerness (an observation that boosts Mackie's overall case), a complementary point (one that derives from the very same considerations), or an independent, stand-alone argument. For these reasons, I will confine myself to Mackie's metaphysical argument from queerness in what follows.

§1.3.2 *Mackie's metaphysical argument from queerness*

Mackie's metaphysical argument from queerness can be interpreted in different ways, and it has inspired different readings and developments of moral error theory. My goal in the remainder of §1.3 will be to explore what I shall call *the queerness reading* and its descendants. This reading takes Mackie at his word, understanding his moral error theory to be rooted in the idea that moral properties would be objectionably queer. The queerness reading differs in important ways from Richard Joyce's (2001) development of Mackie's error theory, to which I attend later on (§1.4).

With those qualifications in mind, we can now turn to Mackie's metaphysical argument from queerness. This argument is premised upon the claim that moral properties, if they were to exist, would have to be *objectively prescriptive* (exactly what Mackie means by 'objectively prescriptive' is something that I will discuss in some detail below).[34] Mackie takes this to be a non-negotiable conceptual commitment of moral discourse. It is, he maintains, "part of what our ordinary statements mean", and underwrites "the traditional moral concepts of the ordinary man" (p.35).[35] His conceptual claim is therefore as follows:

> **Mackie's conceptual claim**
> Moral discourse is conceptually committed to the existence of objectively prescriptive properties.

[34] Mackie seems to be making a broader claim—one that does not only target moral properties, but is directed at moral "entities or qualities or relations" more generally (p.38). I speak of properties here for ease of expression.

[35] I follow others (e.g., Smith 1994, p.64; Svavarsdóttir 2001, p.145) in thinking that Mackie takes non-negotiability to be a matter of conceptual entailment.

Mackie denies the existence of such peculiar properties, which he famously dubs metaphysically "queer". This gives us:

> **Mackie's ontological claim**
> There are no objectively prescriptive properties.

Let's begin by homing in on the conceptual claim. We can start with an initial sketch of what it would take for properties to be objective or prescriptive. The notion of prescriptivity captures the important sense in which moral properties seem to be "action-guiding" (p.23). When some person, action, or state of affairs has some moral property, this has action-guiding force; it has important implications for *what we do*.

What of objectivity? Mackie understands objectivity in terms of being "part of the fabric of the world" (p.15), and "prior to and independent of our choices" (p.30, p.43). We might therefore interpret the 'objective' component of 'objective prescriptivity' as the idea that moral discourse is conceptually committed to the *mind-independence* of moral properties. If something had the property of being good, bad, right, or wrong, then this would not be made true by facts about the perspectives of any (actual or hypothetical) agent or group of agents.

Taken alone, neither objectivity nor prescriptivity is problematic. Scientific claims are presumably objective in the sense that Mackie intends. The number of protons in a hydrogen molecule is plausibly "prior to and independent of our choices". Yet there doesn't seem to be anything metaphysically queer about hydrogen molecules. And the rules of backgammon are surely 'prescriptive' in that they have implications for what we do when playing backgammon. But none of us think that there's anything metaphysically suspect about board games. According to Mackie, the problem arises when these two features are put together—in something's being *objectively prescriptive*.

But what is it for something to be objectively prescriptive? There is some disagreement on this issue. Although most agree that Mackie understands objectivity as I have characterised it above, it is less clear what the notion of prescriptivity involves. And this is likely owing to a lack of clarity on Mackie's behalf—specifically, his consideration of two very different sorts of ontologies that he thought would be able to accommodate objectively prescriptive properties.

One ontological framework that would be able to admit such metaphysically queer properties, Mackie thought, was a Platonic world of forms. In this world, simply coming to know that something participates in the form of goodness (that is, that something is good) necessarily motivates any agent to pursue it (roughly, to do what is morally required of them). The second framework that, Mackie suggests, could accommodate objectively prescriptive properties is Samuel Clarke's envisioned world in which "… a situation would have a demand for such-and-such an action somehow built into it" (p.40). These are strikingly different examples, and they suggest two very different ways in which moral properties could be action-guiding; one concerns moral motivation, the other, moral demands (Garner 1990).

On the motivational reading, the objective prescriptivity of moral properties concerns their purporting to be both mind-independent and *intrinsically motivating*. Moral properties, being motivating in and of themselves, would necessarily motivate any agent (with moral knowledge) to act as they morally ought quite independently of what they are like. I am going to put this motivational reading the side in what follows. My first reason for doing so is philosophical; this reading gives rise to an implausible variety of moral error theory. It is easy to disagree with Mackie from the start if he thinks that moral discourse is conceptually committed to moral knowledge necessarily inducing moral motivation.[36] (Moral knowledge is not usually taken to be any sort of guarantor of virtuous conduct.) And disagree many have on that basis. (See for example, Brink 1984.) As we shall see, these problems do not carry over to the competing (and to my mind, superior) reading, which takes Mackie's talk of prescriptivity to allude to the intrinsically *directive* nature of moral properties—to the fact that they would *demand* certain actions of us. We can presumably have demands thrust upon us even if we are not always motivated to comply with them (Garner 1990, p.144; see also Joyce, p.30).

There are also strong exegetical grounds to prefer the directive reading. Mackie passes over considerations having to do with moral motivation rather quickly. The directive reading, by contrast, is backed up with philosophical argument—namely, Mackie's discussion of categorical imperatives (to which I turn shortly). There is

[36] This is not to deny that moral discourse may be conceptually committed to weaker claims pertaining to the connection between moral judgment in motivation. (More on this in §1.4.3.) For in-depth discussions of Mackie's commitments on this score, see Sinnott-Armstrong (2010) and Dreier (2010).

nothing in this discussion which suggests that the issue with moral properties is their claim to intrinsic motivational power. Nowhere does Mackie link categorical imperatives with motivational efficacy. That Mackie is primarily concerned with morality's intrinsic demandingness is therefore far more congruent with his discussion of categorical imperatives than the reading which would have him pushing for controversial claims about moral motivation being core conceptual commitments of moral discourse.

Let us therefore devote our attention to what I have called a *directive* reading of prescriptivity. This reading is suggested by Mackie's appeal to the sort of world that Clarke envisages. It is further supported by his discussion of imperatives. Mackie tells us that in denying the existence of objectively prescriptive properties, he means to deny that any "categorically imperative element is objectively valid" (p.29). Clearly, some work needs to be done to unpack this idea. I will begin by attending to the distinction between categorical and hypothetical imperatives.

Borrowing heavily from Kant, Mackie takes the distinguishing mark of hypothetical imperatives to be that their validity crucially rests upon the recommended action's being a means to the satisfaction of some desire that the addressee has. For example, we might take 'you ought to open the window' to be a *hypothetical imperative* because there is something of a tacit suffix in place along the lines of '…provided that you desire to cool down'. If you didn't desire to cool down (or have some other desire that would be served by closing the window), this imperative would not be valid. That is to say, the imperative would not legitimately apply to you; it would in some sense "evaporate" (Joyce, pp.31-5). The validity of a *categorical imperative*, by contrast, does not depend upon the desires of the addressee. A categorical imperative can legitimately apply to an agent independently of whether she has a desire that would be served by her compliance.

Mackie argues that *moral imperatives* are categorical; whether or not an agent morally ought to φ does not depend upon whether φ-ing would satisfy any of her desires (p.29). On the face of it, that claim seems eminently plausible. Presumably, we don't take the validity of the imperative 'you morally ought not to torture kittens', to depend upon what the addressee's desires are. Moral imperatives seem to legitimately apply to us quite independently of what our desires happen to be.

However, moral imperatives don't seem unique in this respect. As Phillipa Foot (1972) points out, the imperatives of various institutional systems of rules (e.g., those of etiquette) also seem to legitimately apply to us independently of our desires. We do not withdraw the imperative 'you ought not to chew with your mouth open', for example, when the agent has no desire that would be served by closing her mouth while scoffing down her food. What then, distinguishes these (presumably) ontologically respectable categorical imperatives from those that Mackie relegates to the queer camp?

Mackie's answer is that the categorical imperatives generated by these systems of rules are rendered legitimate by *institutions*. We can explain such standards by appealing to institutional facts such as publicly recognised rules of conduct. Mackie is keen to emphasise that these institutional facts are "…constituted by human thought, behaviour, feelings, and attitudes" (p.81). The distinguishing feature of institutional rules, then, is that they are explicitly *mind-dependent*.

Not so for moral imperatives. According to Mackie, moral imperatives not only purport to be "action-directing absolutely, not contingently…upon the agent's desires" (p.29) (as do other, metaphysically kosher categorical imperatives). They also purport to *transcend* institutional frameworks—to be legitimised by "requirements which are simply there, in the nature of things" (p.59). Moral imperatives are therefore distinct in purporting to impose *objective* requirements upon us.

Thus, Mackie thinks that objectively prescriptive properties entail the existence of imperatives that are *objectively valid*—valid in virtue of facts having nothing at all to do with any agent(s). (p.29). If something were morally wrong, then there would be a demand upon us not to do it—a demand that legitimately applied to us independently of any desires that we may have, and independently of any institutions to which we might subscribe. Moral properties purport to be *intrinsically directive*; to be action-guiding in and of themselves.

Mackie therefore takes moral properties to be queer because they entail the existence of demands that are issued by the world itself. He thinks that we have very good reason to deny that any such properties exist. Exactly what Mackie's reasons are, however, is open to interpretation. Many take his central claim to be that such metaphysically queer properties cannot hope to find a place in our natural world (e.g., Kirchin 2010, p.175). Given their objective prescriptivity, moral properties would be unlike any natural property; they would, it seems, have to be *non-natural* (Copp 2006, pp.9-10). A global naturalist, Mackie refuses to countenance the existence of anything so spooky.

(It isn't altogether clear what these theorists think Mackie takes a natural property to be. (Though this can be forgiven, given that the notion is notoriously difficult to pin down.) Following a suggestion from Hampton (1998, p.22), we might reasonably interpret Mackie as holding that a natural property is any property that is that is countenanced by the best available scientific theory of reality, and to be saying that moral properties could not be countenanced by such a theory.)

Alternatively, perhaps Mackie is putting forward an argument from explanatory irrelevance.[37] He might think that since our best scientific theories "neither recognize nor require for explanation any" objectively prescriptive properties, "we are not licensed to believe that they exist" (Hampton 1998, pp.21–2). On this latter interpretation, Mackie's basic idea is that we should not believe in moral properties because they don't figure in our best account of the world. They are ontologically profligate.

§1.3.3 *Olson's development*

As we have seen, Mackie is moved to endorse a moral error theory by the thought that moral properties would be metaphysically queer. On the interpretation that I favour, the queerness in question concerns the mind-independent and intrinsically directive character that moral properties would need to have, were they to exist. Jonas Olson (2011a, 2014) has recently developed this queerness-based argument for moral error theory, arguing that moral facts entail the existence of queer, *irreducibly normative favouring relations*.

Before we can understand what 'irreducibly normative favouring relations' are, we need to get clear on the concept of a *normative reason*. To this end, it will be helpful to

[37] Though I cast this as an 'alternative' interpretation, one may think there is very little light between the two. But as Lillehammer suggests, there does at the very least seem to be an intuitive difference. The claim that "the very idea of objective value offends against both our best theory of the universe, and our best theory of how the universe is known" seems stronger than the claim that "there is no need to postulate objective moral values in order to explain why we make the moral claims we do" (2011, p.59).

first distinguish normative reasons from motivating reasons (Smith 1994, pp.94-98). Normative reasons are facts (or true propositions) that *justify* our acting in particular ways.[38] (I intend for 'acting' to be read broadly to refer to different kinds of responses that an agent can give; e.g., believing certain propositions, or feeling particular sorts of emotions.) We speak of normative reasons when we say that there is no reason to panic, or that people generally have good reasons to exercise regularly. *Motivating reasons*, by contrast, are facts that *explain why* we acted, or will act, as we do.[39] Motivating reasons are what we speak of when we say that a criminal fled overseas in order to evade the authorities, or that someone chose to visit Australia because there are fabulous beaches there.

Secondly, we should specify not only what a normative reason is not, but what it *is*. A normative reason for φ-ing is, minimally, a consideration that counts in favour of φ-ing (Scanlon 1998, p.17). One popular way of cashing this out is to take normative reasons to be facts that stand in particular kinds of relations to an agent and a type of response that she can give.[40] For instance, the fact that there is delicious coffee in Melbourne might be a reason for a coffee-lover to choose Melbourne as a future holiday destination.

Finally, and in order to properly understand Olson's development of moral error theory, we need to distinguish (i) the fact that is a reason and (ii) the property of being a reason. In the above example, the fact that there is delicious coffee in Melbourne is the fact that *is* the reason. So far, so good—nothing mysterious. Where the mystery comes in, Olson thinks, is in trying to tie down *the property of being a reason*. In the scenario above, we have (a) a person who loves coffee, (b) the fact that there is delicious coffee in Melbourne, and (c) the option of choosing Melbourne as a future

[38] Whether or not facts differ from true propositions, and whether, if they do, reasons are to be identified with one or the other, is an issue that I put to the side here. (See Alvarez (2010, pp.151–8) for discussion.) I will sometimes refer to these facts or true propositions as 'considerations'.

[39] Some think that motivating reasons are properly thought of as mental states—for example, belief-desire pairs (e.g., Davidson 1963, p.687). But there are good theoretical reasons to understand them to be facts. (See Dancy 2000, Alvarez 2010, McDowell 2013.)

[40] To my knowledge, this triadic understanding is the most popular way to make sense of normative reasons. But it is not mandatory. Some may hold that there could be normative reasons that aren't connected in any way to any agent. (See Schroeder (2007a) for discussion.) Others—most notably, contrastivists—may think of reasons as four-place relations. (See Sinnott-Armstrong 2008.)

holiday destination. But then there seems to emerge from (a), (b), and (c) this property of being a reason—this property of 'counting-in-favour-of'. (The fact that there is delicious coffee in Melbourne 'counts in favour of" the coffee lover choosing Melbourne as a holiday destination.) Olson (2011a, 2014) finds this counting-in-favour-of relation mysterious. If our reasons language cannot do away with it—if, that is, our talk of reasons cannot be spelled out in wholly naturalistic, non-normative terms—then he thinks we should deny that any reasons exist.

Fortunately, eschewing the counting-in-favour-of relation is sometimes possible. Olson thinks that this can be achieved for *hypothetical reasons*, which can be reduced to empirical facts about agents' desires and (actual or believed) means of bringing about their satisfaction (2011a, p.78). A hypothetical reasons-claim such as '*a* has a reason to visit Melbourne', for example, can be reduced to the claim that visiting Melbourne will or is likely to satisfy some of *a*'s desires—a desire to sample the world's best coffee, say.

We can also do away with this "queer" counting-in-favour-of relation when we speak of our reasons to act in accordance with institutional systems of rules. Olson regards these reasons-claims as metaphysically innocent. The claim that a chess player has a reason not to move her rook diagonally, for example, can be reduced to facts about her engagement in the relevant rule-governed activity (2011a, p.65).

Importantly, Olson doesn't merely think that in these cases, the fact that is the reason is reducible. (Presumably, most of us agree with *that*—chess pieces aren't ontologically fundamental.) Olson thinks that *the property of being a reason* is reducible, where reduction is achieved by eschewing the understanding of the property of being a reason in terms of a counting-in-favour-of relation. In order for any claim to the effect that *a* has a reason to φ to be true, that claim must be reducible to empirical facts. And he thinks that both hypothetical and institutional reasons-claims satisfy this constraint.

However, Olson argues that the same is not true of *moral* reasons-claims. We take the fact that torturing kittens is morally wrong to provide a reason to anyone to refrain from doing so, quite independently of their desires or their participation in any rule-governed activities. Moral facts therefore entail facts about *categorical reasons*. But since categorical reasons depend upon neither an agent's desires nor her engagement in any rule-governed activities, moral reasons-claims leave the property of being a reason unreduced; we are left with an irreducibly normative reasons-relation. Olson's conceptual claim, then, is as follows (2014, pp.123-4):

> **Olson's conceptual claim**
> Moral facts entail that there are facts that favour certain actions, where the favouring relation is irreducibly normative.

Olson therefore takes moral discourse to be committed to the existence of irreducible normativity. And this, he thinks, is a problem; for irreducibly normativity is "queer" and "metaphysically mysterious" (2014, pp.135-6). Since moral facts entail facts about queer counting-in-favour-of relations, we should take moral facts to be queer as well.

Unfortunately, Olson doesn't say much to illuminate what he means by "queer". He does characterise a queer fact as one that is "ontologically suspicious" (2014, p.84). But that doesn't exactly help. My impression is that Olson takes moral facts to be ontologically profligate. (See Morton & Sampson (2014) for a similar diagnosis.) This is suggested by his discussion of the slipperiness of the term 'queerness':

> Neutrinos, aardvarks, and impressionist paintings may strike us as *prima facie* queer, but when we reflect on how they fit into the natural order of things it is unlikely that we will continue to view them as queer. On reflection, we realize that they are actually parts of the best explanations of some of our observations and beliefs. (2014, p.87)

These remarks suggest that Olson understands queerness in terms of *explanatory irrelevance*; queer facts or relations are those facts or relations that do not figure in our best account of the world.

This interpretation gains further plausibility once we reflect upon the role that evolutionary accounts of our moral beliefs and practices play in Olson's development of moral error theory. Olson explains the presupposition of irreducible normativity in moral language and practice by appealing to the popular idea that it was important for helping societies to survive and reproduce. (I.e., it was important for people to think that they had reasons to act as they morally ought, independently of their ends.) This evolutionary story, he argues, need nowhere presume that *there is* any irreducible normativity. Since we can (plausibly) explain all that needs to be explained without appealing to any such queer facts or relations, Olson thinks we should deny that any such facts or relations exist (2014, p.147).

It therefore seems to be considerations of explanatory irrelevance that move Olson to endorse the following ontological claim:

> **Olson's ontological claim**
> There are no irreducibly normative favouring relations.

Given that (i) there are no irreducibly normative favouring relations, and (ii) moral facts entail that there are such queer relations, Olson denies that there are any moral facts; there are no facts about what is morally right or wrong, nor facts about what we morally ought or ought not to do.

§1.3.4 *The queerness of the queerness charge*

For the purposes of this thesis, I will be assuming what I take to be the most promising variety of moral error theory. To this end, I now want to distance myself from those canvassed above, which proceed by way of raising worries about metaphysical queerness. These arguments can be rather elusive; it is not always clear what exactly queerness consists in. On first appearances, talk of queerness seems to signal a distinct and interesting case against moral properties. On closer inspection, however, arguments from queerness seem to reduce to more familiar strands of philosophical argument, none of which are terribly convincing.[41]

Mackie sometimes equates queerness with 'non-naturalness'. He proposes, for example, that the term 'morally good' "…is used as if it were the name of a supposed non-natural quality, where the description "non-natural" leaves room for the peculiar evaluative, prescriptive, intrinsically action-guiding aspects of this supposed quality" (p.32). So it is reasonable to take Mackie to be objecting to moral properties on account of the fact their non-naturalness. Given this, the queerness charge would seem to rest upon a general hostility to properties that are not amenable to empirical investigation.

Yet it is not clear that any such hostility is warranted. Most philosophers are comfortable talking about beliefs as relations between a speaker and a proposition. But propositions aren't obviously amenable to empirical investigation (though this will

[41] For paper-length treatments of this problem, see Shepski (2008), and Morton & Sampson (2014).

likely depend upon one's metaphysics of propositions). Likewise, many are happy to countenance the existence of *abstracta* such as numbers, given that they are (arguably) indispensable to our best scientific theories (Quine 1951, 1976; Putnam 1979). Indeed, we seem to be committed to the existence of numbers *whether or not* they are susceptible to empirical investigation. Given "…that we rest our empirical investigations *on* such mathematical facts" it may very well be "…a mistake to reject all facts not susceptible of empirical investigation" (Shepski 2008, p.380).

Of course, it may be thought that in these latter cases, the relevant phenomena earn a place in our ontology in virtue of earning a place in (our best) explanations for various phenomena. It is debatable whether *moral* properties can earn their ontological keep. Perhaps this is Mackie and Olson's real worry; the queerness of moral properties concerns their ontological profligacy—there is no need to posit them in order to do any explanatory work.

It's not obvious, however, that *explanatory* indispensability is the only gateway into an ontology; moral properties may be indispensable in other respects. (See Enoch 2011.[42]) It's also debatable whether a failure to do explanatory work licenses *disbelief* in the relevant entities, as opposed to suspension of belief (Sinnott-Armstrong 2006, pp.44-5). These are, admittedly, contentious issues. But they do suggest that moral error theorists who levy the charge of ontological profligacy have their work cut out for them. Ultimately, the strength of such arguments is likely to depend upon the relative power of the arguments in favour of moral non-naturalism—or indeed, the existence of non-natural properties more generally.

In summary, then, I think that there are a number of problems with queerness-based arguments for moral error theory. Once these arguments are demystified, they seem to be nothing over and above more familiar arguments against non-natural properties. But none of these are especially convincing. I think we can do better. In what follows, I outline what I take to be the most plausible route to moral error theory.

[42] I should note that it is in fact irreducibly normative truths that Enoch takes to be indispensable (to deliberation rather than explanation)—not moral truths in particular. But he does think that the latter are afforded entry into our ontology as well "on the wings of the indispensability argument" (2011, p.91).

§1.4 THE TROUBLE WITH CATEGORICAL REASONS

I turn now to Richard Joyce's development of Mackie's moral error theory. Importantly, Joyce doesn't take the central issue with moral discourse to be the non-naturalness of moral properties. His foremost concern is that moral requirements lack the special kind of *normativity* that they purport to have. (As we shall see, these two claims are in fact dissociable.) Before we can understand Joyce's developments, though, we need to understand the particular account of normative reasons and rationality to which he subscribes. My first task will be to clarify Joyce's views on this matter (§1.4.1). I will then explain the conceptual (§1.4.2) and ontological (§1.4.3) components of his moral error theory.

§1.4.1 *Reasons and rationality*

Joyce's moral error theory is premised upon the understanding of normative reasons that he adopts. Joyce subscribes to *reasons internalism*, which we can formulate schematically as follows:

> **Reasons internalism**
> An agent *a* has a normative reason *r* to φ only if *r* is meaningfully connected to certain kinds of motivational fact(s) *m* about *a*.[43]

Reasons internalism comes in different varieties, and each differ in their finer details (Finlay & Schroeder 2015). One such detail concerns what sort of motivational fact *m* must be. Given the popular Humean understanding of desires as intrinsically motivational mental states, it is common to take *m* to be a fact about an agent's desires—for example, the fact that she has certain desires that would be served by her φ-ing. Internalists also disagree about whether *m* must be a motivational fact about *a* as she *actually* is, or something that would hold true of her under particular circumstances—for example, under conditions of rationality (Smith 1994), in a state of reflective equilibrium (Brandt 1979), or in awareness of all relevant contingencies (Darwall 1983).

[43] In Stephen Darwall's (1992) terminology, reasons internalism is an 'existence' form of internalism; it concerns what it takes *for there to be* a reason. It is to be distinguished from 'judgment internalism', which concerns what it takes *for an agent to be judging* that she has a reason.

Joyce takes reasons internalism to be highly plausible (pp.109-12), appealing to the popular idea that normative reasons must be capable of *explaining* an agent's actions. (See Davidson 1963, Williams 1981, Schroeder 2007b.) Assuming, as Joyce does, that intentional action requires the presence of a desire (assuming, that is, a Humean Theory of Motivation), this explanatory requirement speaks in favour of internalism. Given that the internalist ties normative reasons to an agent's motivations, an agent's normative reasons are capable of explaining her actions—they are reasons *for which* she could act.

Joyce distinguishes two kinds of normative reasons. An agent has an *objective reason* to φ just in case φ-ing will satisfy her ends, and a *subjective reason* to φ just in case she is justified in believing that she has an objective reason to φ (p.53).[44] (Joyce uses 'ends' as a catch-all phrase for desires and interests. I will adopt this convention in what follows.) Popeye, for example, may have a subjective reason to eat a bowl of spinach when he justifiably believes that doing so would satisfy his desire for food and cause him no harm. However, Popeye would have an objective reason not to eat the spinach had it been poisoned by Bruno.

Joyce also draws a connection between rationality and reasons, taking an agent to be *practically rational* to the extent that she is guided by her subjective reasons (p.54). On this view, an agent is practically rational to the extent that she takes what she justifiably believes to be the necessary means to her ends. Practical rationality therefore "yields only hypothetical imperatives" (p.51); what we are rationally required to do depends upon the ends that we have.[45] It's worth clarifying that it doesn't follow from this that all practical reasons are 'instrumental', in one common sense of 'in-

[44] Following R.J. Wallace (2003), some may complain that this "multiplies reasons beyond necessity". What Joyce calls 'subjective reasons' are perhaps "…not really reasons at all, but rather beliefs of agents about what they have reason to do." I am inclined to regard this as a largely terminological quibble, and am happy for 'subjective reasons' to denote what agents justifiably believe their objective reasons to be.

[45] I should note that Joyce rejects a rudimentary, Humean form of instrumentalism, according to which these must be an agent's actual ends, or her strongest desire at the moment of decision (p.69). He goes on to refine his account of subjective reasons, suggesting that a subject "…S has a subjective reason to φ if and only if she is justified in believing that S+ (S granted full information and idealized powers of reflection) would advise S to φ" (p.100). This qualification may, as Joyce suggests, make practical instrumentalism more plausible. But nothing that I have to say here hangs upon it.

strumental' (Beardman 2007, pp.260-1). If I have the end of φ-ing and ψ-ing is a necessary means for φ -ing, then I have an instrumental reason to ψ. But this is not to say that I have an instrumental reason *to φ*. One need not think, for example, that I have an instrumental reason to φ because I already have a reason to pursue my ends. (I will return to this issue in §2.1.2.) The instrumentalist can hold that reasons transmit from *intrinsic* ends or desires to means (where an intrinsic or non-instrumental desire is one that an agent does not merely have as a means of satisfying some other desire). Most sophisticated instrumentalists do (e.g., Hubin 2001).

The connection that Joyce draws between subjective reasons and rationality is rooted in the thought that rationality attaches not to the content of a belief, desire, or choice, but to "the manner at which it is arrived" (pp.55-6)—rationality is a form of negligence rather than ignorance (see also Korsgaard 1996a). It is therefore preferable to take an agent's acting in accordance with what she *justifiably believes* that she has reason to do determine her rationality. On this view, what it is rational for an agent to do and what she has an objective reason to do can come apart. Popeye's decision to eat the poisoned spinach, for example, may be characterised as *incorrect* but *perfectly rational* (see Wedgwood 2003).

On the flipside, an agent's *failure* to act in accordance with her subjective reasons is thought to render her rationally criticisable. Given that rational criticism is presumably only warranted if an agent fails to comply with her *best* reasons (one is not obviously rationally accountable for failing to act on any just any old *pro tanto* reason), Joyce understands an agent's normative reasons to be what she has *most* or *decisive reason* to do (pp.50-1).[46] Practical rationality is therefore understood as a normative framework that tells us what our *all-things-considered* reasons for acting are.

As I understand Joyce, he takes rationality rather than reasons to be the foundational notion here. The litmus test for a normative or "real" reason is that it would be irrational to ignore it (p.41, p.51, pp.80-1). This is because an adequate theory of

[46] As I understand Joyce, he doesn't take this to be true of *all* normative reasons—only those that are relevant to determining an agent's rationality. He claims to be using 'normative reason' in "…a restricted sense, to mean something that is justified according to practical rationality" (fn.21). So this view need not conflict with the common understanding of normative reasons as *pro tanto* reasons, the balance of which decides what an agent ought to do, all-things-considered (Broome 2004, Dancy 2004).

normative reasons must "…make out reasons to be precisely those things that forestall a 'So what?' response" (p.81). Joyce regards this condition as particularly important, for he thinks that practical rationality is immune to legitimate questioning:

> Can we imagine someone questioning practical rationality: "Yes, I recognize that there is a practical reason for me to φ, but what is that to me?...This, it seems to me, is incoherent ...Even to ask the question "Why should I be interested in practical rationality?" is to ask for a reason. Thus even to question practical rationality is to evince allegiance to it… (pp.49-50).

Given that practical rationality itself is 'un-so-what-able', Joyce conjectures that an agent's normative reasons, responsiveness to which determines her rationality, should similarly be un-so-what-able. And this result, he seems to think, is easy to secure if we respect the internalist constraint that normative reasons must be grounded in an agent's motivations. Since a normative reason is a consideration to which an agent cannot rationally respond 'so what?', an agent's normative reasons must be considerations that are potentially engaging for her—considerations that could potentially motivate her, or could provide her with an adequate justification for an action by her own lights (p.108).[47]

Loosely following Joyce, I will refer to this theory of reasons and rationality as "practical instrumentalism". There are two features of this view that are worth noting. First, the practical instrumentalist constructs a link—or as it is sometimes put, a "nexus" (Smith 2007)—between reasons and rationality. She is therefore to be distinguished from those who take rationality to be silent at the level of reasons (e.g., Broome 2005). Second, the practical instrumentalist does not take rationality to (merely) be a faculty of the mind. Rationality can involve *doing* something—not merely believing or intending something when one believes or intends something else. She therefore differs from those who take rationality to be a system of requirements that govern only relations among one's psychological attitudes (e.g., Broome 2005, Kolodny 2005, Southwood 2008). In what follows, I will sometimes refer to this as a

[47] For a complementary point, see Hubin (1999, p.39), who argues that on such views, the charge of irrationality is "unshruggable". An irrational agent "…cannot avoid the motivational force of our judgment about him by pleading that he does not care about the ends in question".

distinction between 'act theories of rationality' and 'psychological theories of rationality' (loosely following Broome 2005, p.325).

§1.4.2 *Joyce's conceptual claim*

Paraphrasing slightly, Joyce's argument for moral error theory has the following structure (p.77):

J1. If *a* morally ought to φ, then *a* morally ought to φ regardless of what her ends are.[48]

J2. If *a* morally ought to φ, then *a* has a reason for φ-ing.

J3. Therefore, if *a* morally ought to φ, then *a* can have a reason for φ-ing regardless of what her ends are.

J4. But there is no sense to be made of such reasons.

J5. Therefore, *a* is never under a moral obligation.

The last two premises are Joyce's ontological claim—I return to them later. The first three constitute his conceptual claim, and require some unpacking. **J1** is inspired by the cornerstone of Mackie's error theory. Joyce thinks that categorical imperatives are a necessary feature of a system of moral requirements (p.62). The guiding thought here is that morality is *categorically applicable*; someone can be morally required to φ even if φ-ing would not satisfy any of her ends (p.56). **J1** is therefore a claim about the *content* of moral requirements; they are not restricted to actions that would serve an agent's interests, or satisfy her desires.[49]

However, Joyce recognises that the requirements of institutional systems of rules are also categorically applicable. Yet he does not take issue with them. What, then, distinguishes moral requirements from the requirements of (say) etiquette? **J2** suggests an answer. What sets moral requirements apart, Joyce claims, is their purporting to have a distinct kind of practical authority—an authority that consists in the *provision of reasons* (p.38). Though the requirements of morality and etiquette are both categorically applicable, only the former are *categorically reason-giving*; moral requirements purport to provide all agents with reasons to comply, and this is so independently of

[48] I use the amended version of J1 from Joyce (2016), rather than the original formulation in Joyce (2001).

[49] I follow Shafer-Landau (2005) in understanding J1 to be a claim about the content of moral requirements, and J2 to be a claim about their normative import.

whether compliance would further their ends. **J2** is therefore a claim about the *normative import* of moral requirements—they purport to carry reason-giving force (pp.39-41).[50]

The third premise (**J3**) is Joyce's conceptual claim and follows from the two premises that precede it. **J3** informs us that moral requirements purport to supply agents with *categorical reasons* for action—reasons that legitimately apply to them independently of their ends. In Joyce's view, then, categorical reasons are a core conceptual commitment of moral discourse. Any system of values that leaves them out "…simply does not count as a "morality" at all" (p.177). Thus, we have:

> **Joyce's conceptual claim**
>
> Moral discourse is conceptually committed to the existence of categorical reasons.

Joyce therefore holds that if moral discourse is to be vindicated, there must be categorical reasons. This means that there would need to be some actions for which a valid categorical imperative could be directed at any agent, where a valid categorical imperative is one for which a "real reason" could be given to any agent (in particular circumstances) to comply (p.44). In order to vindicate moral discourse, then, there would have to be a class of actions (that are properly called moral actions) that any agent would have a real reason to do when she finds herself in the relevant circumstances.

As I noted in the introduction, my goal in this thesis is not to offer a sustained defence of moral error theory. However, I also noted there that claims regarding the non-negotiable conceptual commitments of a discourse can be controversial. So it's worth briefly rehearsing some of the considerations in favour of this one.

To begin with, Joyce's conceptual claim would seem to accord with our intuitive judgments. He appeals to the example of Plato's Lydian shepherd, Gyges, who acquires a ring of invisibility and uses it in morally heinous ways to further his ends. Gyges has no desires or interests that would be served by restraint. Nonetheless, it

[50] Another way to understand this idea is to invoke a well-known distinction between two kinds of normativity (Parfit 2011, pp.267-8). Both morality and etiquette are normative in a "rule-involving" sense—for both are constituted by requirements. But only moral requirements purport to be normative in a *robust* sense as well—to necessarily provide *reasons* to act as morality requires. On this point, see also Foot (1972, p.308), Railton (1986, p.165), and Dreier (1997, p.84). Joyce prefers to characterise the distinction as a distinction between "institutional reasons" and "real reasons" (pp.34-41).

seems highly intuitive to say "…both that Gyges ought not kill innocent people and that he has a reason not to kill innocent people" (p.41). It is, in other words, highly intuitive that Gyges has a normative reason to cease with his behaviour, quite independently of what his ends happen to be.

Though the appeal to intuition here is powerful, it may not be quite enough to motivate the claim that the existence of categorical reasons is a *non-negotiable* conceptual commitment of moral discourse. What distinguishes the negotiable from the non-negotiable? Joyce's answer is that we need to examine how moral concepts are characteristically *used* (2006, pp.201-2; 2012, p.95). His suggestion, roughly, is that if a concept could not be used in the characteristic ways that moral concepts typically are, then it would not properly be called a *moral* concept. And in Joyce's estimation, any concept of a moral requirement or obligation that did *not* entail categorical reasons could not perform the characteristic uses of moral concepts.

One such characteristic use of moral concepts concerns *evaluations* of agents and actions. We take an evaluation of an action as right or wrong to have important implications for how agents *ought to behave*. Joyce (2006) considers Jack, who wants to kill John and has no desire that would be served by sparing his life. Plausibly, we should want to say that Jack's killing John would be morally wrong. But if we did not take the wrongness of Jack's actions to entail categorical reasons for him to refrain, then it seems that we should rather be forced to say "From the moral point of view, killing John was unacceptable" and to add "but Jack had…no real reason to refrain" (Joyce 2006, p.204). As Andres Luco (2016, p.2516) observes, the latter concession seems to destabilise "an important use of moral concepts…It throws into doubt the conviction that knavish individuals ought to act in accordance with moral evaluations".

Moral concepts are also characteristically used for the purposes of *criticism*. We typically criticise people for wrongful behaviour, and take ourselves to be justified in doing so. But it's hard to see how we *could* be so justified if we did not take the person criticised to have any real reason to act otherwise. In order to make sense of moral criticism, we must, it seems, suppose that wrongdoers have *reasons* not to act as they do. Eric Vogelstein has recently afforded this idea a paper-length treatment, arguing that "…morally wrong action licenses a critical attitude towards the wrongdoer—an attitude that, is justified only if the wrongdoer has failed to comply with her reasons" (2013, p.1084).

I take the considerations above to lend quite a bit of plausibility to Joyce's claim that moral discourse is conceptually committed to the existence of categorical reasons. We do, it seems, ordinarily suppose that agents can have moral reasons to refrain from wrongful behaviour even when doing so would frustrate their ends. And characteristic uses of moral language—for the purposes of evaluation or criticism, for example—likewise seem to carry with them the presumption that wrongdoers have genuine reasons to fulfil their moral obligations, quite independently of what their ends happen to be.

§1.4.3 *Joyce's ontological claim*

Joyce takes moral discourse to be conceptually committed to the existence of categorical reasons. **J4**, however, tells us that "there is no sense to be made of such reasons". As I understand Joyce, the relevant claim is not simply an epistemic one about our abilities to make sense of certain phenomena. (Physicists, for example, may have no problem countenancing the existence of particular kinds of subatomic particles, even if they often struggle unsuccessfully to make sense of them.) As we shall now see, his real contention is that the notion of a normative reason is itself fundamentally at odds with the notion of a categorical reason. Given this, there simply *cannot be* any categorical reasons. Joyce's ontological claim is therefore as follows:

> **Joyce's ontological claim**
> There are no categorical reasons

Recall that a valid categorical imperative is one from which we can extrapolate normative reasons. What Joyce's ontological claim denies is that moral imperatives are valid imperatives—that all agents necessarily have normative reasons to act as morality requires independently of their ends. Shafer-Landau (2005, p.113) offers a helpful summary of the reasoning at play here:

1. If there are categorical reasons, then, for any reasonable agent S, S might have reason to φ, but fail to be engaged by φ.
2. Reasons cannot fail in this way.
3. Therefore, there are no categorical reasons.

To demonstrate, recall the case of Gyges. We want to say that Gyges has a reason to stop killing innocent people. But Gyges lacks any desire or interest that speaks in favour of such restraint. The fact that killing innocent persons is morally wrong simply has no sway upon him. Thus, if agents like Gyges have categorical reasons to refrain in such cases, then agents can have normative reasons that fail to properly engage them. But given practical instrumentalism, normative reasons cannot fail in this way; for normative reasons are, by definition, the sorts of considerations that are capable of *engaging* the relevant agent. Thus, Joyce concludes that there are no categorical reasons—*a fortiori,* there are no moral reasons.

Before concluding, I should note that there is a common strategy for attempting to evade a moral error theory—one which consists in arguing that moral requirements are in fact requirements of practical rationality. A well-known proponent of this strategy is Michael Smith (1994).[51] Smith agrees with Joyce that the reason-giving force of moral requirements does not depend upon any agent's *actual* ends (i.e., that conceptually speaking, morality is categorically reason-giving). However, he claims that their reason-giving force can be vindicated by considering every agent's *counterfactual* ends—specifically, the ends that every agent would have if they were fully rational.

Smith takes our normative reasons to be determined by what our fully rational selves would desire for us to do in any particular circumstance *c*. He argues that our ideal selves would *converge* upon these desires, giving all agents the same normative reasons (in any circumstance *c*). Since moral reasons are a proper subset of these normative reasons, the account, if successful, vindicates the idea that all agents have normative reasons to act in accordance with moral requirements independently of what their ends happen to be.

Joyce seems to concede that moral discourse could be vindicated if our ideally rational selves would converge upon their desires.[52] But he denies that they would. This is because even on Smith's analysis, there appears to be an important kind of *path-*

[51] Another is Christine Korsgaard (1996b).

[52] This is suggested by his broader claim that "the only hope" for evading a moral error theory is "...to defend the thesis that practical rationality delivers categorical imperatives, and then to forge a connection between the imperatives of practical rationality and those of morality" (p.51).

dependency involved in what determines each agent's normative reasons; what *a*'s ideally rational self wants her to do will always depend in some way upon what *a actually* desires. This makes it exceedingly unlikely that our ideal selves would converge upon their desires. Given that people vary immensely in their goals, beliefs, and ideologies, there is "every reason to think that" these differences will "transfer to their idealised versions" (Joyce, p.76; see also Sobel 1999, Bukoski 2016). Since we cannot expect all agents to converge upon their normative (and hence, moral) reasons, we cannot maintain that all agents have normative reasons to act in accordance with moral requirements.[53]

Joyce's exchange with Smith brings out two interesting features of his moral error theory. First, it suggests that this error theory, if true, is *necessarily* true. In order to vindicate moral discourse, moral requirements must be necessarily reason-giving; all possible agents must have normative reasons to comply with them. Since moral requirements are not necessarily reason-giving, our moral claims are, it seems, *necessarily* false. (This verdict is shared by Coons (2011, p.87); though see Brown (2013, p.631).)

Second, Joyce is largely in agreement with Smith about the conceptual commitments of moral discourse; both agree that there is a tight conceptual connection between what an agent is morally required to do and what she has normative reason to do. Yet Smith also defends a tight conceptual connection between moral requirements and *rational motivation*, suggesting that if an agent judges that she morally ought to φ, then she will, insofar as she is rational, be motivated to φ (1994, p.61). This is, of course, a variety of motivational internalism—albeit one that is considerably less strong than the Platonic idea (considered in §1.3.2) that moral judgment (or knowledge) induces *overriding* motivation. Smith's variety of motivational internalism isn't wholly implausible given his conceptual claim that moral requirements are requirements of practical rationality; it requires only the (reasonable) bridging principle

[53] Both Smith and Joyce regard convergence as important. Smith takes normative reasons to be conceptually *non-relative* (i.e., he thinks it is a conceptual truth that all agents have the same reasons in any circumstance *c*). A divergence in agents' normative reasons would therefore amount to an error theory about normative reasons on his view. Joyce takes *moral* reasons in particular to be conceptually non-relative (pp.95-99). Since divergence in agent's normative—and hence, moral—reasons leads to moral relativism (not all agents would have the same moral reasons in any circumstance *c*), moral discourse is not vindicated.

that rational agents are motivated to do what they have all-things-considered reason to do. Given this, the moral error theorist's conceptual claim is sometimes described as a claim about the connection between moral judgment and rational *motivation*.[54] (See for example, Braddon-Mitchell 2006, West 2010.)

To my mind, Joyce's reasons-based argument for moral error theory has a far greater measure of plausibility than the one rooted in queerness concerns. We could of course rephrase Joyce's argument in terms of queerness as well if we so wished. (Joyce himself (pp.30-1) briefly suggests that the alleged queerness attaches to the notion of "moral bindingness"; to the idea that someone is somehow normatively bound by moral demands independently of whether she cares about them.) What is important to appreciate is that the reasons-based argument seeks to establish only that moral requirements lack the distinct kind of normativity (or 'reason-givingness') that they purport to have. There is no need to take issue with anything's being non-natural, or ontologically profligate—the charge isn't that moral properties would be creepy, kooky, mysterious, or spooky.

Indeed, someone who endorses Joyce's variety of moral error theory need not think that the existence of any non-natural properties is necessary in order rescue moral discourse from error. What is needed in order to vindicate moral discourse is for all agents to have normative reasons to act in accordance with moral requirements—a condition that could be satisfied if idealised agents were to converge upon their desires. Following Smith (2012), there doesn't seem to be anything spooky or non-natural about the notion of convergence among idealised agents.

Joyce's error theory also seems to do away with the notion of *mind-independence* to which Mackie subscribed. Joyce seems happy to allow that moral claims have *intersubjective* truth-conditions; such claims would be true, he thinks, if all possible agents would converge upon their normative reasons.[55]

[54] Though Joyce himself is likely to resist this characterisation, since he denies that motivational internalism is a core conceptual commitment of moral discourse (pp.17-29), and explicitly rejects Smith's variety of the view (pp.22-3, pp.64-67).

[55] That convergence-based error theories differ in important ways from those rooted in concerns having to do with non-naturalness has been noted by Lillehammer (2004, pp.97-100; though cf. Hussain 2004).

For the purposes of this thesis, I shall be assuming the success of the Joyce's variety of moral error theory. When I speak of 'moral error theory' in what follows, then, this should be taken to refer to the reasons-based variety in particular, unless otherwise indicated.

§1.5 Moral Error theory: summing up

This thesis is not about the moral error theory *per se,* but about what comes *after* it. But before we can explore what comes after, we need to know what comes before; we need to understand what the moral error theory is. The primary goals of this chapter were to identify the background meta-ethical assumptions that underwrite the moral error theory, to explain the arguments available for it, and to specify which variety of the position will form my background assumption in this thesis. These tasks were useful for the purposes of getting clear on what the moral error is, and providing the reader with some of the necessary conceptual and terminological background for the remainder of the work.

Chapter Two

What is the 'what next?' question for moral discourse?

In the last chapter, I explained what the moral error theory is, and the kinds of considerations that might lead us to believe that it is true. For the remainder of this thesis, however, I want us to assume that the moral error theory *is* true and, moreover, that we believe that it is true. Our concern will be the question that follows this assumption: what ought we to do next?

WNQs are certainly not unique to meta-ethics; they currently occupy the minds of philosophers of mathematics, philosophers of mind, and even those who have embraced an error theory about colour. But I do not take the sole justification for addressing our WNQ to be an argument from precedent. There is more to be said to motivate the project, and saying more is the task of this chapter.

I will begin by explaining what kind of question our WNQ is (§2.1). As I have characterised it, the question is a normative one—a question regarding what we *ought* to do if the moral error theory is true. However, it is not obvious that our WNQ *can be* a normative question. The arguments marshalled in support of moral error theory may support an error theory about *all* normative reasons—or so some philosophers have argued. If they are right, then we are to have no recourse to normative language. Others have claimed that the moral error theorist is committed to denying the existence of epistemic reasons, which (it is alleged) are also categorical in nature. If they are correct, then we cannot speak of reasons for belief. I will explore and defend what I take to be the best strategies for responding to these over-generalisation problems. Doing so can, I hope, earn me the right to speak of various kinds of (non-moral) reasons for the remainder of the work.

The tasks for §2.2 and §2.3 respectively will be to specify to whom our WNQ is addressed, and to motivate the project of providing an answer to it. I will then outline the different answers to the WNQ for moral discourse that have been proposed in the current literature (§2.4). (All but one will be explored in the remainder of the thesis.) The chapter concludes by pointing towards the philosophical pay-offs of the project (§2.5). These dividends are not confined to moral error theorists. The project should also be of interest to those contemplating error theories in their own sectors of philosophy, and indeed, to meta-ethicists more generally.

§2.1 CAN THE WNQ FOR MORAL DISCOURSE BE A NORMATIVE QUESTION?

Our first order of business will be to explore whether the WNQ for moral discourse can be a normative question. It is not uncommon for philosophers to characterise the question in this way; as a question regarding what we *should* do, or *ought* to do, or have *most reason* to do with moral discourse, following the moral error theory (e.g., Joyce, p.177; Nolan, Restall & West 2005, p.310; Lutz 2014, p.352; Svoboda 2017, p.3). In doing so, they would appear to presume that the moral error theorist's rejection of categorical reasons leaves many other, non-moral reasons intact.

Yet it is not obvious that the moral error theorist's arguments can be insulated in this way. Some think that she is committed to an error theory about epistemic reasons as well (Cuneo 2007, Rowland 2013). Others argue that she is committed to an error theory about *all* normative reasons (Hampton 1995, Korsgaard 1997, Shafer-Landau 2003, Raz 2005, Bedke 2010) These are, of course, over-generalisation worries; for an error theory about all normative reasons (or even just epistemic reasons) seems undesirable. And insofar as the moral error theory entails this more global thesis, so too, it seems, is moral error theory. These worries would also seem to be bad news for the project of answering our WNQ. If there are no reasons to do or to believe anything, then, *a fortiori,* there can be no reason to do anything with or believe anything about moral discourse going forward.

My goal in this section will be to earn the right to normative language for the remainder of the work. I shall argue that the moral error theorist is not committed to an error theory about epistemic reasons (§2.1.1), nor to a more thoroughgoing error theory about all normative reasons (§2.1.2).

§2.1.1 *The over-generalisation problem for epistemic reasons*

I shall tackle the over-generalisation problem for epistemic reasons and the more global over-generalisation problem separately, beginning with the former.[56] Some philosophers argue that the error theorist's attack upon moral reasons is likewise an attack upon reasons for belief. Epistemic reasons, they hold, are similarly categorical; whether or not an agent has a reason to believe some proposition *p* does not depend upon whatever ends she happens to have (Cuneo 2007). Richard Rowland, for example, writes that

> …our understanding of epistemic reasons and justification also entails that there are categorical reasons…it seems that there is reason for everyone to believe that dinosaurs once roamed the earth regardless of what they want to believe. (2013, p.4)

I think that Christopher Cowie's is the strongest response to this line of argument, and I will draw heavily upon his work in what follows, before adding some developments of my own.

As Cowie (2014a, pp.117-8) notes, intuition-eliciting cases like Rowland's do not make for a very dialectically effective response to the moral error theorist. She will simply deny that anyone has a categorical reason to believe that dinosaurs once roamed the earth. There is, of course, a claim in the vicinity that she will endorse; she will concede that the fossil record provides *evidential support* for the claim that dinosaurs once roamed the earth. But this is just to say that the evidence supplied by the fossil record raises the probability that dinosaurs once roamed the earth. It is not to say that the evidential support relation supplies anyone with a *categorical reason* to believe that dinosaurs once roamed the earth.

Jonas Olson enlists a similar strategy. As he observes, 'epistemic reason' is ambiguous between 'evidence' and 'reason for belief' (2014, p.158). The former notion is not itself normative, and, he suggests, it is enough to get us by; for evidence can still supply *hypothetical* reasons for belief—reasons that depend upon our having adopted

[56] There are good grounds for addressing the specific problem for epistemic reasons directly. The strategy of enlisting epistemic reasons as partners in guilt (or innocence) is well-known, and so, worthy of discussion in its own right. It will be also helpful to have in hand an account of epistemic reasons that is consistent with moral error theory.

the end of being evidence-responsive, for example.[57] (I return to the latter idea below.)

At this stage, it is open to the moral error theorist's opponent to maintain that evidential support relations *just are* categorical reasons for belief. But it's not obvious that this claim is plausible. To be sure, it might seem intuitive to say that the fact that some evidence *e* raises the probability of some hypothesis *h* is a reason to believe *h*. The crucial question, however, is whether this 'reason' is properly thought of as a *normative* reason.

Cowie is sceptical that it is, and I share his scepticism. On the common view of things, a normative reason for believing some proposition *p* is evidence that one *ought* to believe it (Kearns & Star 2009), or an explanation as to why one *ought* to respond to this evidence in a particular way (Broome 2013). And it just does not seem true that evidence for some proposition *p* is always evidence that one *ought* to believe that *p*. Cowie makes the point by appealing to banal truths:

> … Suppose that I possess, and am aware of possessing, evidence e that bears on some proposition p. But suppose that I have no interest in arriving at a true or evidentially supported belief about that proposition. And suppose that it would not serve any practical end for me to do so. If one nevertheless maintains that e is evidence that I ought to believe that p (and not merely that e is evidence for the truth of p), the burden is surely very much on them to explain why. (2014a, p.121)

So the error theorist will concede to her opponent that the fossil record provides evidential support for the proposition that dinosaurs roamed the earth. Likewise, she will concede that this evidential support relation is a reason to believe that dinosaurs roamed the earth *in one sense* of 'reason'. But she will deny that we are speaking here of a *normative* reason.

Just what *are* we speaking of, then? Cowie (2014a, pp.120-1) alludes to "institutional reasons", such as those supplied by the norms of etiquette. These norms may provide us with reasons to say, wear pants to dinner. But again, it does not necessari-

[57] For similar suggestions, see Lenman (2008) and Heathwood (2009).

ly follow that anyone has a *normative* reason wear pants to dinner.[58] Presumably, whether some individual has a normative reason to wear pants to dinner depends upon whether she has a normative reason *to participate in* the institution of etiquette in the first place. Cowie suggests that the same is true of epistemic reasons: if one is to have a normative reason to believe some proposition *p*, then one must "also require a reason to engage in the business of believing (the truth) with respect to that proposition" (2014a, p.121; see also Heathwood 2009, p.96; Olson 2014, pp.165-8).

I think this idea is important, and worth developing. The basic thought, I take it, is that there is a kind of epistemic activity or institution that is in some sense *normatively optional* for us. We may or may not have normative reasons to be in the business of responsible believing—to shape our beliefs in accordance with our evidence, and the like.

Thankfully, it seems that many of us do in fact have such reasons; for many of us do have ends that are served by responsible believing—the end of believing truths, and the end of not believing falsehoods, say. Assuming that many of us do have such epistemic goals, many of us will have epistemic reasons to respond appropriately to our evidence. (Insofar as responding appropriately to one's evidence is likely to be an effective means for arriving at truth, that is. I assume here that it will be—at least in typical circumstances.[59]) Moreover, even those who don't value truth *per se* may have good instrumental reasons to be responsible believers. After all, most of us care about furthering our ends, and we're generally better positioned to further those ends—whatever they may be—if we track truth effectively (Kornblith 1993, p.371; 2001, pp.158-9). This way of seeing things dovetails nicely with the error theorist's claim that all normative reasons are hypothetical—that is, contingent upon our (intrinsic) ends or desires.

[58] This assessment is not uncommon. It is, for example, plausible that undesirable social policies may supply legal reasons to harm others, without supplying any normative reasons to do so (Lillehammer 2002, pp.54-5; see also Joyce, pp.34-42).

[59] I do not assume that believing in accordance with one's evidence will *always* result in the acquisition of true beliefs, nor that believing truly will *always* further one's various goals—that would be too strong (see Stephens 2001). There are bound to be exceptions to such unqualified claims. But following Cowie, these complications should not "…obscure the basic point that forming and revising beliefs on the basis of one's evidence is of great practical utility" (2014b, p.4007; see also Wedgwood 2002).

The approach is, of course, a variety of epistemic instrumentalism; the view that epistemic rationality is a species of instrumental rationality, or rationality in the pursuit of one's goals. (I remain neutral here on the matter of whether these goals must be distinctively epistemic ones—the goals of believing truth and avoiding falsehoods, say.[60]) Loosely following Adam Leite (2007, p.456), we can characterise epistemic instrumentalism as follows:

> **Epistemic instrumentalism**
> A belief is epistemically rational when (and because) holding it is instrumentally rational given one's goals, and one has an epistemic reason to believe some proposition *p* when (and because) doing so would be instrumentally rational given those goals.

This view has an impressive fan base. (See, for example, Foley 1987; Kornblith 1993, 2002; Laudan 1990, 1991; Maffie 1990; Papineau 1999). But I do not pretend that it is uncontroversial. (For criticism, see Kelly 2003, Lockard 2013.) There will be some who are not particularly taken with this account of epistemic reasons. But my ambitions here are rather modest. I am only attempting to show that given epistemic instrumentalism, the error theorist has a reply to the over-generalisation worry for epistemic reasons, and there is a respectable notion of an epistemic reason to which she can appeal.[61] Adopting this understanding can, I hope, earn me the right to speak of epistemic reasons for the remainder of this thesis. (The reader is therefore free to take my conclusions in what follows to be conditional on the truth of epistemic instrumentalism—a matter that I haven't fully adjudicated here owing to considerations of space and priorities.)

[60] On this issue, see Lockard (2013), who distinguishes "intellectualist" varieties of epistemic instrumentalism, according to which epistemically rational beliefs are those which serve an agent's cognitive or epistemic goals, from "pragmatist" varieties, which do not require that the relevant goals be epistemic ones.

[61] It is no surprise that many moral error theorists (e.g., Joyce, pp.178-9; Olson 2014, pp.158-9) are (or very much seem to be) attracted to epistemic instrumentalism. The position is often hailed as a promising and respectable strategy for naturalising epistemic normativity, and shirking talk of categorical reasons. But it is not the only such strategy. One could instead interpret claims about epistemic rationality as expressions of preferences or tastes (see Field (2000)). Indeed, Stephen Ingram (2017) has recently recommended that moral error theorists be epistemic expressivists. I put this possibility to the side here.

There is, however, still a problem in need of address. We have assumed that individuals typically have good *hypothetical* reasons to be responsible believers (e.g. to respond appropriately to their evidence) because responsible believing is a means of realising certain (intrinsic) ends that they have. But perhaps these hypothetical reasons simply cannot exist without categorical ones. It is to this latter complaint that I now turn.

§2.1.2 *The over-generalisation problem for all normative reasons*

The moral error theory, recall, is premised upon an instrumentalist conception of normative reasons and rationality: "practical instrumentalism". On this view, an agent is practically rational to the extent that she takes what she justifiably believes to be the necessary means to her (intrinsic) ends. The practical instrumentalist also holds that there are only hypothetical reasons—specifically, reasons to adopt particular means when one has particular ends (Joyce, p.51).

Yet quite a few have voiced the suspicion that there must be categorical reasons lurking in the background of these hypothetical ones. Here is the problem. Practical instrumentalism tells us something about the *transmission* of reasons: they are transmitted from ends to means. But if reasons are to be transmitted, then there must be reasons to transmit. So—and here's the rub—it seems that the instrumentalist requires a *categorical, non-instrumental principle* as a foundation for hypothetical reasons; she must posit a categorical reason to have certain ends (e.g., the end of taking the necessary means to one's ends) if there is to be anything to transmit in the first place (Hampton 1995, pp.70-71; Korsgaard 1997, p.223; Raz 2005, p.23).[62]

Call this *the grounding challenge*; it is the challenge of saying just what (if anything) grounds or explains the normativity of hypothetical reasons, if not categorical reasons. If the challenge cannot be met—if, that is, hypothetical reasons must presup-

[62] There is a slightly different articulation of the charge that the moral error theorist is committed to an error theory about all normative reasons, according to which hypothetical reasons-relations are just as *metaphysically queer* as categorical reasons-relations (Shafer-Landau 2003, Bedke 2010). However, this challenge seems more pertinent to Olson's variety of moral error theory (which appeals to queerness concerns) than to that of Joyce (which does not). Since I am assuming the latter, I will not address this variant of the objection.

pose the existence of categorical ones—then (insofar as she continues to deny that categorical reasons exist) the moral error theorist seems committed to an error theory about practical reasons as well.

One tempting response to the grounding challenge is to dismiss any request for an explanation here as incoherent.[63] Joyce seems to flirt with this possibility when he suggests that asking for a reason to be rational is utter nonsense (p.49).[64] One is effectively demanding a practical reason to be guided by practical reasons ('what reason do I have to do what I have reason to do?') and so, one is presuming that *there are* practical reasons.

Though tempting, this line of response is unlikely to be dialectically effective. Many philosophers agree that asking for a reason to take the means to one's ends seems incoherent (e.g., Dreier 1997, p.95; Railton 2003, p.317). But they draw a rather different conclusion. The nonsense involved in challenging the claim that we have a reason to take the means to our ends is not that of *asking for a justification* for it. It is the nonsense of trying to use the claim that we have a reason to take the means to our ends to do the justificatory work! The nonsense evaporates as soon as we posit a categorical reason to pursue our ends. (See Dreier 1997, p.96; Railton 2003, pp.318-9.)

Here is what I take to be a more promising strategy. The instrumentalist's adversary assumes that she must offer *a particular kind of explanation* of an agent's reasons: she assumes that the instrumentalist must explain why someone has a reason to perform some action by showing that performing that action is a means of doing *something else that she has a reason to do* (Schroeder 2007b, ch.3). Suppose, for example, that Luke desires coffee and desires to go to a nearby cafe because there is coffee there.

[63] Some have also objected to the grounding challenge on account of its failure to distinguish between the normativity of rationality and the normativity of reasons (e.g., Broome 1999, Pauer-Studer 2009). However, I do not think that this move is open to our practical instrumentalist, who forges a connection between the two.

[64] This is not to say that *generally speaking,* the question 'why be rational?' is utter nonsense. It is only to say that it would seem to be nonsense given practical instrumentalism. If we understand rationality in terms of conforming with a set of requirements governing one's psychological attitudes (i.e., if we adopt what I earlier called 'the psychological understanding'), then the question certainly *is* coherent; for it seems that we can sensibly ask why we should act in accordance with such requirements (as many have—see Broome 2005, 2007; Kolodny 2005; Southwood 2008).

The adversary assumes that the practical instrumentalist must explain Luke's reason to go to the café as follows:

> *Luke desires coffee. What explains why he has a reason to go to the nearby café is that doing so is a means for Luke to do something that there is already a reason for him to do: to pursue his desires.*

As should be clear, this explanation runs headfirst into the grounding challenge. The instrumentalist's adversary will insist that any reason to pursue one's desires must be a categorical one.

But this is not the only explanation that the instrumentalist can offer. The practical instrumentalist is in the business of offering an analysis of the concept of having a reason. She is proposing that for all agents *a*, and all actions φ, *a* has a reason to φ iff *a* stands in the reasons-relation to φ. It is true that whether or not *this reasons relation holds* is not contingent upon an agent's ends. But as Joyce points out, this

> …does not imply that he has a reason to perform an action whether he likes it or not. It means that whether he likes it or not, if he stands in a certain relation to that action then he has a reason for performing it. (p.119)

As I understand Joyce, the basic suggestion here is that the practical instrumentalist can offer *a constitutive explanation* as to why some individual has a reason to perform a particular action. She can, in other words, explain why some agent has a reason (or ought) to do something by appealing to *what it is* to have a reason (or for it to be the case that one ought) to do something (cf. Hubin 2001).

Here, then, is our answer to the grounding challenge: we can explain why someone has a hypothetical reason to perform some action by appealing to *what it is* to have a reason. According to the practical instrumentalist, for some agent *a* to have a reason *r* to perform some action φ *just is* for her to have a particular (intrinsic) end *e* that would be served by her φ-ing.[65] In order to explain *a*'s reason to φ in some circumstance *c*, we can simply claim that *a* is currently in conditions in which her having a reason to φ obtains. There is no need to claim that she has any further reason—let

[65] Some practical instrumentalists (e.g. Joyce 2001) may want to refine this claim by speaking of what an idealised version of *a* (*a*+) would desire that she do. But I will simplify matters here for ease of illustration; for even an unqualified form of instrumentalism seems capable of evading the grounding challenge.

alone a categorical one—to further her ends. We can, for example, explain Luke's reason as follows:

> *Luke presently has the end of acquiring coffee, and there is coffee at a nearby café. What explains why Luke has a reason to go to a nearby café is that he is currently in conditions in which his having a reason to do so obtains; going to the nearby café is a means of satisfying his end of acquiring coffee.*

As Mark Schroeder (2007b, p.62) observes, such a constitutive explanation does not run into the same trouble as the initial explanation that the instrumentalist seemed forced to offer; for there is no need to appeal to any 'further reason' that Luke has to satisfy his desires. The practical instrumentalist can explain an agent's reasons simply by telling us that particular conditions obtain—the conditions in which some *r* is a reason for some agent *a* to perform some action φ.

I hasten to add that this is no philosophical magic trick; constitutive explanations are perfectly respectable. As Schroeder observes, it is perfectly respectable (and natural) to think that "…being three-sided makes a figure a triangle not because there is any further shape that all figures have, but because that is just what triangles are: three-sided figures" (2007b, p.62).

§2.1.3 *The WNQ: summing up*

I have argued that the moral error theorist is committed to neither an error theory about epistemic reasons, nor to a more global error theory about all normative reasons. It's worth noting, however, that even if these arguments were unsuccessful, not all would be lost for the project of answering the WNQ.

Even if we were led into wholesale normative nihilism, we could still make sense of asking what we *are* now to believe or to do. Addressing this latter question would not require us to form any normative beliefs (e.g., the belief that we ought to be moral fictionalists.) We would only need to form an intention or reach a decision—for example, to decide whether or not to be fictionalists.[66] Though a normative nihilist could not consistently ask what we *ought* to do with moral discourse going for-

[66] This is not to be confused with the question as to whether or not we *will be* fictionalists.

ward, she could still ask what *to do*—whether or not to be an abolitionist, say (see Southwood 2016, pp.62-3).[67]

One might object that we would have no reason *to care* about the answer to the question as to what we are to do going forward. Yet it seems that we can care about things without believing that we ought to care about them. Indeed, it is not at all implausible that many of us do sometimes care about things without believing that we ought to care about them—perhaps even while believing that we shouldn't. One might also worry that we would have no reason *to believe* any of the proposed solutions to this question. Yet it's not clear that this would prevent us from believing them (*pace* Streumer 2013). An inductive sceptic might think that there is no reason to believe that the sun will rise tomorrow. But it still seems that she could believe that the sun will rise tomorrow (Möller & Lillehammer 2015; see also Olson 2011d).

Accordingly, it is not obvious to me that normative nihilism *would* be utterly devastating for the project of answering the WNQ. In that case, we would simply devote our efforts to addressing the question as to what we are to do. However, my preference here will be to provisionally assume that the grounding challenge can be met. Given the preceding discussion, I hope that the reader will not think this an assumption made in bad faith.

Making this assumption also carries some distinct advantages. For one thing, we need not scrap all mention of reasons for the remainder of this work. (I suspect that doing so would be rather difficult). Moreover (and as I have noted), few if any of those who are in the business of answering the WNQ for moral discourse think that the arguments for moral error theory support an error theory about all normative reasons. In the interest of engaging directly with their arguments, it would seem best to follow their lead here, and to understand the WNQ for moral discourse as a normative question.

[67] I thank Nic Southwood for helpful discussions on this point.

§2.2 To whom is the WNQ for moral discourse addressed?

As I shall now characterise it (and, I hope, am now permitted to characterise it), the WNQ is the question regarding what we ought to do with moral discourse if we believe the moral error theory. The 'we' here suggests that I am concerned with a hypothetical situation in which we, as a linguistic community, have chosen to devise a collective solution to the WNQ. On this understanding of the problem, we're in it together.

But there is an alternative situation that could arise. We might unanimously believe the moral error theory, but choose to devise our answers to the WNQ *independently*. On this understanding of the problem, it's every moral every theorist for themselves. The WNQ instead asks us: what ought *some individual* to do if she believes the moral error theory?[68] This latter question is likely to invite quite different answers to the first. Presumably, our answer to the first question would be informed by certain ends that we share to some degree. But any answer to the second question would be heavily dependent upon the ends of the relevant moral error theorist.

In this work, I confine myself to the collective question. Three considerations speak in favour of doing so. The first concerns the philosophical interest of the project. In order to answer the question, 'what ought some individual to do if she believes the moral error theory?', we would need to identify some individual to whom our answer would be addressed. Yet whom would we choose? The most natural choice for me would of course be myself. But I expect that my readership would be rather limited (or, perhaps I should say, *more* limited) were this entire thesis intended as practical advice for me—as of course would the wider philosophical interest of the project.

The second justification for framing the WNQ question as a collective question is a straightforward, dialectical one: the vast majority of philosophers who are in the business of addressing the WNQ for moral discourse interpret the question along

[68] We could also ask about a *unilateral* moral error theorist; what ought she to do, given that she alone is in the now? I put this possibility to the side for the purposes of my investigation, which assumes that *everyone* is in the know about the moral error theory.

these lines (e.g., Nolan, Restall, and West 2005, Joyce 2005, Svoboda 2017). In the interest of properly engaging with their arguments, and comparing their proposals to my own, it seems best to follow their lead here.[69]

This second consideration—that of engaging with others' arguments on the same terms—speaks in favour of framing the WNQ as a collective question. But I admit that it is not a very principled justification for doing so. 'Everyone else is doing it' might pass as a justification for wearing flared jeans. But it won't always pass as a justification in philosophy. My third justification for understanding the WNQ as a collective question is a more principled one: many of the benefits of engaging in moral practice (on which more in §2.3.3) plausibly require some linguistic co-operation on our part. Moral discourse can, for example, be a useful resource in practical disputes.[70] Following Nolan, Restall, and West (2005, pp.212-3), talk of rights and duties forms part of a "well-established framework" of "tacit understandings", and this framework can be helpful when we need to adjudicate competing interests, or to divvy up the relevant burdens in an acceptable way. Yet it is difficult to see how moral language could be useful in these contexts if we couldn't take it for granted that our moral terms have something *meaningfully in common*; if we couldn't take it for granted that we were using terms like 'rights' and 'duties' in a broadly similar way, or that we all took claims about rights and duties to have certain kinds of implications for our behaviour, say.

This is just to say that some degree of linguistic co-operation is plausibly needed if we are to preserve the benefits of moral discourse going forward. Two moral error theorists who revise the meaning of their moral terms independently might come to

[69] This is not to say that there is no room to question the common assumption that this is a question to be asked and answered collectively. An interesting and perhaps under-appreciated aspect of the issue is the extent to which this is true. In answering the collective question, it is also difficult to avoid making assumptions about what makes for good collective decision-making processes; for example, the assumption that there are many widely-shared preferences and/or desires, and the assumption that it is these which we ought to try to satisfy as best as possible for the majority of people. These are interesting issues, but properly engaging with them is likely to take us too far afield (the literature on voting paradoxes comes to mind).

[70] Of course, some think that we'd be far more successful at resolving practical disputes *without* moral language (Hinckfuss 1987, §4.2; Garner 2007, pp.202-3; Ingram 2015, pp.240-1). I mention this view shortly, and address the arguments for it in chapter 3.

mean very different things when they declare an action 'wrong'. And those who pursue the fictionalist option independently might come to adopt moral fictions with radically different contents. If we devise independent answers to the WNQ, then it seems that there will be very little that we can take for granted in the moral (or schmoral, or pretence-driven) conversations of the future.[71]

We should therefore want to ensure that morality (or its future stand-in) constitutes a *common currency* (assuming for the moment that moral discourse is to be preserved in some form). We should want the moral (or schmoral, or pretence-driven) interlocutors of the future to be able to take certain things for granted in their moral conversations. Devising a *collective* solution to the WNQ helps us to achieve this. If, for example, we decide to collectively preserve moral discourse in the form of a useful fiction, then the moral interlocutors of the future will take that fiction to be a common currency in their moral conversations.

Moreover, even those who *don't* think that moral discourse ought to be preserved in some form have good reason to seek a collective answer to the WNQ. Moral abolitionists argue that our moral practices are harmful on-balance, and recommend that we scrap them altogether. But even *they* think that the benefits of doing so depend upon collective action. A well-known abolitionist argument is that doing away with moral discourse would help us to resolve interpersonal conflicts (Hinckfuss 1987*,* §4.2; Garner 2007, pp.202-3; Ingram 2015, pp.240-1). But achieving this presumably depends upon *all* of us ceasing to speak in moral terms. Conflict resolution is unlikely to be any easier if some parties continue to stubbornly appeal to their moral beliefs.

Of course, there is a potential difficulty with this collective policy-seeking. Some may fear that the question regarding what *we* ought to do with moral discourse is one that cannot be answered; for people surely have different interests that they want to further, and different desires that they want to satisfy. Why then, ought we to think that *any* answer to the WNQ will further each and every person's ends?

In response to this concern, we can point towards the various ends that many people *do* share. Presumably, most of us want to avoid an untimely demise, to see to

[71] This is not to say that moral communication would necessarily break down—a good conversational partner, after all, will often make her presuppositions explicit. But it does seem that moral discourse is less likely to play the same sort of positive social role(s) if it ceases to be a common currency.

it that our loved ones are well, and to live in a stable and co-operative society. Of course, not *everyone* has these ends. (Sensible knaves will remain.) But in order for an answer to the WNQ to be useful, it need not be useful for *everyone.* It need only be useful for *most* of us, who share a broad variety of interests and concerns.

But can we be justified in casting some people and their deviant interests to the side when addressing the WNQ? I think that we can; for we want our answer to the WNQ to be *useful.* And it is unlikely to be very useful if its intended audience are those who do not care very much about their survival, wish only ill for those closest to them, and eagerly await the downfall of modern society. Our answer is likely to be more useful if is addressed to (what I expect and hope is) the majority of the linguistic community.

I suspect that something like this thought is what guides those in the business of addressing the WNQ for mathematical discourse: what ought we to do with mathematical discourse if we believe that an error theory about numbers is true? Philosophers of mathematics don't pay attention to those with deviant interests when they attempt to answer this question. One finds little mention in such discussions of those who despise arithmetic, have little use for it, and would be very happy indeed if we dispensed with talk of numbers altogether.

This, I take it, is presumably because philosophers of mathematics are concerned with the interests of *the linguistic community on the whole* when addressing the WNQ for mathematical discourse. And that concern seems legitimate, as does their assumption that most of us do benefit from speaking of numbers. If this is indeed the concern that drives philosophers of mathematics, then we are partners in innocence. Or perhaps we are partners in guilt. Either way, I will follow their lead here. My WNQ will be the question regarding what is likely to be useful for *most* of us to do with moral discourse, following the moral error theory.

§2.3 MOTIVATING THE PROJECT

In what follows, I do some work to motivate the project of answering the WNQ for moral discourse. To this end, I will establish the initial plausibility of four claims that give rise to a tension. (These claims will also be useful for the purposes of taxonomising different answers to the WNQ in §2.4.) I begin by suggesting that (1) if we believe that the moral error theory is true, then we no longer believe that anything is morally right or wrong (§2.2.1). This, I will suggest, is a devastating result because (2) our subjective concerns (and consequent motivations and behaviour) covary fairly closely with our moral beliefs (§2.2.2). Thus, a significant practical consequence of our believing the moral error theory would be the loss of particular motivations and behavioural dispositions. This, I argue, would very much be a loss; for (3) morality is useful, and moral beliefs are helpful to have around (§2.2.3). Our allegiance to moral error theory therefore incurs a steep practical cost. And (4) it is not obvious that we can respond to the problem by simply persisting with our false moral beliefs; for doing so would seem ill-advised (§2.2.4).[72] An answer to the WNQ is therefore needed if we are to resolve this tension between our epistemic values and our other, real-world needs.

§2.3.1 *The devastating result of the moral error theory*

My goal in this section will be to argue that the moral error theory seems to carry worrying implications for first-order moral theorising; for if we believe the moral error theory, then we no longer believe that anything is morally right or wrong. This seems to be a devastating result. And I suspect many will think that such a result follows from the error theory quite straightforwardly (perhaps even rather obviously). We didn't after all, continue to organise witch-burning ceremonies once we discovered that there were no witches, or to seek out dragons to slay after we came to appreciate that there were none around.

However, this result is not as straightforward and obvious as it may initially seem. Precisely what sort of relationship holds between first and second-order ethics is a contentious issue. There is some debate as to whether there can be any traffic between these two 'levels' of ethical inquiry.[73] Some think not (e.g. Rawls 1974-1975;

[72] Throughout this work, I will, for ease of exposition, often speak of 'having false moral beliefs' and 'believing moral propositions'. What I really intend by this is false beliefs in positive, first-order, atomic, non-trivial, moral propositions such as 'cheating is wrong', and 'tyranny is unjust'.

[73] There is also some debate as to whether there *are* two levels of inquiry here. Some (e.g., Blackburn 1984, 1998; Dworkin 1996, 2011) think that all statements or questions *about* ethics are really statements or questions *within* ethics. I do not have space here to properly engage with their argu-

Mackie, pp.16). But this is something of a minority position; the majority do allow for this possibility.[74] (See for example, Dreier 2002, Enoch 2011). And they would seem right to do so. As Stephen Darwall (2006, pp.25-34) observes, first-order moral claims can sometimes rest upon meta-ethical ones. Deontologists, for example, would seem to presuppose a meta-ethical account of moral obligation according to which the right can come apart from the beneficial, and the wrong from the harmful. Likewise, our meta-ethical commitments can influence our choice of first-order moral theories. It is no coincidence that meta-ethical naturalists are often attracted to consequentialism. These examples are hardly idiosyncratic. And they do, I believe, put quite a bit of pressure upon those who insist upon the strict independence of first and second-order moral theorising.

It therefore seems plausible that what we say and do in the meta-ethics classroom can have implications for what we say and do in the normative ethics classroom, and vice versa. For my purposes here, it will not be necessary to sketch the precise conditions under which this obtains.[75] It will suffice to motivate the idea that as far *as the moral error theory is concerned*, these conditions plausibly do obtain. Many find this verdict plausible (if not obvious). Russ Shafer-Landau, for instance, writes that

> If there are no truths within morality—only a truth about morality, namely, that its edicts are uniformly untrue—then the enterprise of normative ethics is philosophically bankrupt. Normative ethics is meant to identify the conditions under which actions are morally right, and motives morally good or admirable. If nothing is ever morally right or good, then normative ethics loses its point. (2005, p.107)

Crispin Wright echoes these sentiments when he remarks that

> … as soon as philosophy has taught us that the world is unsuited to confer truth on any of our claims about what is right, or wrong, or obligatory, etc.,

ments, and I do not have much criticism to add beyond what has already been said by Bloomfield (2009), Shafer-Landau (2010), and Enoch (2011, ch.5). (This is not to deny that these theorists are getting at something. I will suggest that there is an important kernel of truth in what they say in §6.4.) In any event, I hope that the reader will not take my presuming that meta-ethics exists to be an illicit assumption.

[74] Note that one can concede this possibility *while maintaining* that at least some meta-ethical claims are neutral at the level of first-order ethics.

[75] For systematic attempts to do so, see Dreier (2002) and Enoch (2011).

> the reasonable response ought surely to be to forgo the right to making any such claims... If it is of the essence of moral judgement to aim at the truth, and if philosophy teaches us that there is no moral truth to hit, how are we supposed to take ourselves seriously in thinking the way we do about any issue which we regard as of major moral importance? (1996, p.2)

Following these theorists, it seems that a significant purpose of ethical theorising is to discover moral truths—truths about what is right or wrong, say. So long as we think that there are such truths, it seems that there could be some value in attempting to uncover them. But once we have come to believe that there is *no moral truth to hit at*, it would surely be incongruous (to say the least) for us to simply continue on as before.

We might also think that moral discourse chiefly consists in *invoking* these truths in our moral conversations; we often appeal to moral considerations in our attempts to convince others to act accordingly. But appealing to moral considerations in everyday contexts likewise seems odd if we're moral error theorists—what is the point of telling someone that they morally ought to φ if it is never the case that anyone morally ought to do anything?

The truth of moral error theory therefore seems to carry worrying implications for first-order moral theorising. If we are moral error theorists, then, presumably, we do not believe that there are any moral truths to be discovered in ethical theorising, nor any truths about what we morally ought to do that can be invoked in everyday contexts. The following, then, is what I take to be the devastating result of our having come to believe the moral error theory:

> **(1) Devastating result**
> We no longer believe that anything is morally right or wrong.

In chapters 6 and 7, I will argue that matters are in fact more complicated here than they first appear. But explaining why will take a good deal more work. At this stage, my purpose is to draw attention to what many theorists seem justified in assuming; that our believing the moral error theory is likely to have consequences for our ability to engage (or at least, engage seriously) in first-order moral theorising.

§2.3.2 *Doom and gloom?*

I have suggested that it plausibly follows from our believing the moral error theory that we no longer believe that anything is morally right or wrong. What remains to be shown, however, is that this devastating result really is *devastating*. To this end, we must ask what is likely *to happen* if we believe that nothing is really right or wrong. If total anarchy would result, then we will urgently need to search for a remedy. But if barely anything would change, then addressing the WNQ for moral discourse would seem to be of comparatively little practical—and perhaps even philosophical—importance.

The anarchy hypothesis isn't wholly unpopular; many have expressed despair at the idea that the moral error theory could be true. Joyce and Kirchin offer a nice rehearsal of the relevant concerns:

> Two thousand years ago, Aristocles of Messina asked "What evil deeds would he not dare, who held that nothing is really evil, or disgraceful, or just or unjust?" Paraphrasing Dostoyevsky, one might declare "If there is no moral truth, then everything is permitted." And Dr Johnson memorably said of the moral skeptic: "If he does really think that there is no distinction between virtue and vice, why, sir, when he leaves our houses let us count our spoons" (2010, p.xiv).

Yet such sentiments seem to express an unwarranted cynicism about human nature. If we believe the moral error theory, then it does seem to follow that we believe that nothing is really right or wrong. But are our moral beliefs all that stands between a Hobbesian war of all against all and civilised society as we know it? That seems implausible. After all, many of our subjective concerns align closely with our moral goals; most of us have strong non-moral desires to help others, to tell the truth, and to foster worthwhile friendships.[76]

[76] By non-moral desires, I mean desires with non-moral content. Compare: (i) a desire to be a good person, and (ii) a desire to treat others with compassion and concern. Even if these desires turned out to be extensionally equivalent—if they were in the end, desires to do the very same thing—the former invokes moral concepts in a way in which the latter does not. In philosophical jargon, both might count as moral desires in virtue of their distinctive subject matter, but (i) is a *de dicto* moral desire, whereas (ii) is a *de re* moral desire (Smith 1994, pp.74-5). I return to this distinction in §2.3.3.

But rejecting the anarchy hypothesis seems to land us in a different sort of trouble: perhaps *none* of these subjective concerns depend upon our moral beliefs at all. Perhaps we value helping others and telling the truth quite independently of whether we take such things to be morally good or required. If this were so, then our believing the moral error theory may seem to be of little practical significance. Perhaps we would continue to be caring and truth-telling people because these are the kinds of people that we want to be.

To my mind, neither of these predictions survives closer scrutiny.[77] It is clearly wise to steer clear of the slightly melodramatic sentiment that our believing the moral error theory would effect a radical upheaval of society as we know it. But we should also be cautious of maintaining that our believing the moral error theory would have *no* practical implications whatsoever—that our subjective concerns would remain wholly unaffected once we came to believe that nothing is morally right or wrong. Instead, we ought to occupy a plausible middle ground on this question: we should expect that our believing the moral error theory would have *some* practical implications. In support of that expectation, we can appeal to the following, plausible thesis:

> **(2) Covariance thesis**
> Our subjective concerns (and consequent motivations and behaviour) covary fairly closely with our moral beliefs, in the sense that these concerns, motivations, and behaviours are often influenced by our moral beliefs.[78]

The covariance thesis should strike us as eminently plausible. After all, people usually do make some efforts to perform those actions that they take to be morally right, and try to refrain from those actions that they take to be morally wrong. This is not to deny that some people might knowingly engage in wrongful behaviour fairly often. Committed utilitarians, for instance, may think that they're frequently failing in their duties to help the global poor. Nonetheless, they would still seem sufficiently sensi-

[77] I appreciate that the question as to what would happen if we all believed the moral error theory is an *empirical* question—the answer is not one that ought to be "..pronounced upon with any confidence from one's armchair" (Joyce & Kirchin 2010, p.xv). Nonetheless, given what we do know about our own motivational dispositions and behaviour, I think we can with at least some confidence deny that either of the aforementioned predictions is plausible.

[78] I am grateful to Guy Kahane, who pointed me towards his parallel (2016) claim that our subjective concerns covary fairly closely with our evaluative and/or normative beliefs more generally.

tive to moral considerations, in the sense that that their behaviour tracks their moral beliefs fairly well—they're not out stealing or stabbing, for instance.[79]

Moreover, we usually expect a change in subjective concerns or motivation to accompany a change in moral beliefs—especially in decent, well-informed, and strong-willed people (Smith 1994, pp.71-6). If I change my mind, and come to believe that voting for the liberals (rather than the conservatives) is the right thing to do, then we would, *ceteris paribus*, expect a change in my behaviour to follow suit. This phenomenon need not be restricted to those who are morally perfect (as we would ordinarily say). Moral conviction can sometimes counteract even implicit and deeply entrenched biases (Paxton & Greene 2010).

On the flipside, a denial of the covariance thesis should strike us as eminently implausible. Moral beliefs certainly don't seem epiphenomenal. It is odd to think that they have no causal impact whatsoever upon our subjective concerns and our actions.[80] Moral beliefs do seem to figure in our explanations of other people's behaviour. We often account for someone's helpfulness by appealing to their *belief* that helping was morally required of them. And indeed, people themselves sometimes cite their moral beliefs to explain their own actions. ('I thought it was the right thing to do'.) That moral beliefs figure in such explanations is decent evidence that they have causal effects upon the world—in particular, upon our motivations and behaviour.[81]

Before concluding, I should note that the covariance thesis does not entail that there is any deep or tight conceptual connection between moral belief and motivation—that is to say, it doesn't entail the truth of motivational internalism. Whether covariance is a contingent fact about human beings or a brute fact about agency is unimportant. For the purposes of establishing the claim that believing the moral er-

[79] Thanks to Daniel Nolan for pointing this out.

[80] Some come close to supporting this suggestion. Jonathan Haidt (2001) argues that moral judgments are the product of "intuitions" (roughly, emotionally-laden, gut responses), and that it is these intuitions (rather than our moral beliefs) that have the most significant effects upon our motivations and behaviour. Yet Haidt doesn't claim that our moral beliefs have *no influence whatsoever* upon our behavioural or motivational dispositions. Not even he thinks that they are wholly epiphenomenal.

[81] Admittedly, such evidence is defeasible. But absent any reason to think otherwise, there seems to at least be some presumption here in favour of the commonsense claim that our moral beliefs do sometimes explain our motivations and our behaviour (when working in concert with particular desires, of course).

ror theory would have *some* practical implications, it will suffice that there actually is some such covariance between our moral beliefs and our subjective concerns, motivations, and behaviour. And that seems difficult to deny.

§2.3.3 *Morality (huh): what is it good for?*

So far, I have argued for the following two claims: (1) if we believe that the moral error theory is true, then we no longer believe that anything is morally right or wrong, and (2) our motivations and behaviour covary closely with our moral beliefs. This suggests that a significant practical consequence of our believing the moral error theory would be the loss of particular motivational and behavioural dispositions. But would this really be *a loss*? I shall now argue that it would be. More specifically, I shall motivate the following two claims:

> **U1**: Acting in characteristically moral ways has many non-moral benefits at both the individual level and the social level.
>
> **U2**: Moral language and moral judgments are especially well-suited for disposing and/or prompting us to act in these beneficial ways.

As I will now proceed to explain, characteristically moral behaviour tends to have positive social consequences (U1). A world in which people help one another and keep their promises is valuable to us in non-moral terms; we do better if we can rely upon others to help us and keep their word. Put differently, we do better to live in a society in which people are *co-operative*, and take the interests of others into account when deciding what to do.

The idea that morality enhances social co-operation is not new.[82] It is often argued that one important function of morality is that of "crowd control" (West 2010, p.184; see also Copp 2009, p.27); that of helping us to overcome inclinations that are likely to get in the way of prosocial behaviour. Few people are self-sufficient. Most stand to benefit from co-operating with others. But it is difficult to secure the benefits of co-operation; for it can often be in an agent's immediate, short-term interest to

[82] It is no mere matter of armchair conjecture either. According to a popular and well-respected empirical hypothesis, our capacity to make moral judgments was a biological adaptation that provided a much-needed boost to our co-operative dispositions (Ruse 1986, Joyce 2006, Kitcher 2011, Lahti & Weinstein 2005, Haidt 2012).

defect. Given that moral considerations are characteristically other-regarding, they tend to dissuade individuals from pursuing self-interest at the cost of the social good, serving as "checks on…natural inclinations or spontaneous tendencies" (Mackie, p.106).

A related thought here—one owing to David Gauthier—is that moral requirements may be viewed as mutually beneficial constraints upon the pursuit of self-interest in strategic interactions. On Gauthier's way of seeing things, morality is "a system of principles such that it is advantageous for everyone if everyone accepts and acts on it" (1967, pp.461-2).[83] Simply put, it to an individual's overall rational advantage to comply with moral rules, provided that everyone else does so as well. (Though it might not be to her advantage in each and every instance; she will sometimes have to "perform disadvantageous acts".[84])

Moreover, social-coordination is arguably easier to achieve with a currency of shared values in place. Such values can serve as a social adhesive, helping individuals to identify as a community, and cultivate common ends. Of course, it is debatable just how much overlap there is among individuals' *moral* values. (There is certainly moral disagreement.) But even if the overlap here is not perfect, talk of rights and duties forms part of a "well-established framework" of "tacit understandings"—something that can be helpful when we need to navigate our way through practical disputes (Nolan, Restall, and West, 2005, pp.312-3).

Still, one might ask why morality in particular is important for the purposes of securing co-operative behaviour. Why, that is, should we believe U2? I think we should

[83] Gauthier's project bears some interesting similarities to Bernard Mandeville's (1714) *Fable of the Bees*. For Mandeville, what appears to be virtuous and other-regarding conduct is often self-interested behaviour in disguise.

[84] When Gauthier claims that it is to one's overall rational advantage to act in accordance with moral requirements, he is invoking a 'constrained maximising' conception of rationality rather than a 'straightforward maximising' conception. The thought here is that it is rational for an agent to act upon a decision-making policy that disposes her to make the most advantageous choices *overall*, even if some token choices are disadvantageous (Gauthier 1986, pp.170-77). It is of course possible to take Gauthier's view to be *a vindication* of the idea that we necessarily have reasons to act as morality requires. But this would, I think, be the wrong inference to draw. If Gauthier is correct, then when particular circumstances obtain, we may often have normative reasons comply with moral rules. But what must be shown (if we are to vindicate morality) is that individuals *necessarily* have normative reasons to act as morality requires. On this point, see Sayre-McCord (1989).

concede that morality may not *indispensable* for these purposes. But there are good reasons to think that it is nonetheless especially helpful. For one thing, moral judgments are arguably more effective elicitors of co-operative motivation than mere preferences or desires. Our resolve to be to help others is plausibly stronger when we not only desire to help them, but believe that we *must* or *morally ought* to help (Joyce 2006, pp.110-1). My desire to keep my promises might be overridden by the benefits that defection offers. But if I am morally required to keep my word, then I take there to be a *categorical demand* upon me to do so—I don't think that I can rid myself of any such obligation by citing an interest in defection.

A related thought here is that moral judgments function as *deliberation-stoppers*; they prevent competing considerations that would interfere with prosocial motivation from entering into the deliberative sphere (Joyce 2006, p.111). When an agent judges that she morally ought to φ, she takes it to be the case that she ought to φ, *period*; her other desires and ends are sidelined in the decision-making process. Daniel Dennett (1986, p.123) pushes a similar line, suggesting that public moral judgments function as *conversation-stoppers;* they block any further negotiations from taking place when making interpersonal decisions—especially when time and resources are scarce.

Moral practice also makes possible an array of powerful sanctions. Many of these are external.[85] Moral misdemeanours invite criticism and reproach, and those who earn reputations as cruel or inconsiderate cads tend to have less social capital. Since very few of us wish to be on the receiving end of such hostility, we usually do our best to avoiding acting in ways that are likely to elicit it.

Blame deserves special mention here; that is, the characteristic and forceful ways in which we negatively appraise others in response to perceived wrongdoing.[86] Of course, blame does have a rather bad name (Fricker 2016, p.168). (And the 'bad' here

[85] There are internal sanctions as well, of course (see West 2010, p.189). I omit these purely for considerations of space. (I cannot hope to list all of the potential benefits of moral practice here.)

[86] I assume here that blame is an aspect of moral practice that we may risk losing if we believe the moral error theory. More specifically, I assume that blame entails judgments with *moral content*; that it involves a judgment of wrongness (Sher 2006), or a judgment that another's "moral standing" has been diminished, for example (Zimmerman 1988, p.38). Not everyone thinks as much, of course (see Scanlon 2008). My suggestion here is merely that those of us who do take attributions of blame to have moral content should also regard blame as useful aspect of moral practice.

need not only allude to *moral* badness.) But it is my contention that blame is useful precisely on account of its distastefulness. Social interactions are underwritten by a common knowledge of our mutual susceptibility to evaluation—and that knowledge has motivational force. We care deeply about avoiding others' reproach. Blame, being punitive and at times, uncompromising, is therefore a particularly effective mechanism with which to keep one another in check.[87] Our deep awareness of our moral exposure, together with our keen interest in avoiding blame, shapes our behaviour, moving us to treat others with due concern.

Moreover, we have desires with moral content. Many of us desire to do what is right *de dicto*—to do what is right, whatever that may be. Admittedly, this motive has earned its fair share of bad press. Those motivated to do right *de dicto* have been branded "moral fetishists" (Smith 1994, pp.75-6). I am inclined to agree that desiring to do what is right *de dicto* is not a hallmark of the morally best sort of person. (Though *qua* error theorist, I must intend for this as a purely conceptual claim.) But I am also inclined to think this desire is particularly helpful to have around.

Though we may have cause for worry if someone were only ever motivated to care for others by a *de dicto* desire to do what is right, I think we would also have cause for worry if she had no such desire. The desire to do what is right (*de dicto*) plays a valuable role in securing prosocial behaviour. Hallvard Lillehammer seems to be tracking the role I have in mind when he invites us to imagine a woman who

> …goes to a party during a phase when she is tired of her husband. At the party she meets a very charming person and is tempted to have an affair. She judges that it would be wrong to have an affair on account of her husband's feelings. But she is temporarily indifferent to her husband's feelings. However, she has a standing *de dicto* desire to do what is right which, together with her moral judgment, causes her to do the right thing, in spite of the absence of a *de re* desire to do the right thing and the presence of a *de re* desire to do the wrong thing. (1997, p.192)

The basic suggestion here is that *de dicto* moral desires can *over-determine* prosocial behaviour (e.g., spousal loyalty). These desires provide a motivational safety-net of

[87] I make no claim to originality here; it is commonly recognised that blame has the function of modifying future conduct. See McGeer (2013), and Pereboom (2013).

sorts—one that is especially helpful when our *de re* desires falter.[88] Of course, one is tempted to reply to Lillehammer that it would be better if people's desires not to cheat on their partners were more robust. But as a matter of fact, not everyone has such robust desires; a desire to be a loyal spouse can wax and wane as the terrain of a relationship changes. Given the way things actually are, we seem to do better if our desires to act in characteristically prosocial ways are over-determined—if we have *de dicto* moral desires that can pick up the motivational slack.

In summary, there do seem to be a number of benefits to engaging in moral practice. Importantly, most if not all of these benefits of morality seem (i) to be linked to particular motivational and behavioural dispositions, and (ii) to be rooted in moral judgments (i.e., beliefs) and the language that gives expression to them. The considerations above would therefore seem to speak in favour of the following claim:

> **(3) Usefulness of morality**
> Moral beliefs and the discourse that gives expression to them provide us with a number of desirable practical goods in virtue of the behaviour and motivations that they facilitate.

In what follows, then, I will assume that our moral practices are useful to us. This is not to assume that they are *on-balance* useful to us. (That is a claim to be assessed in chapter 3.) An appeal to their *prima facie* usefulness will suffice for the time being.

§2.3.4 *The tension: why we need to answer the WNQ*

So far, we have the following package of claims:

> **(1) Devastating result**
> We no longer believe that anything is morally right or wrong.
>
> **(2) Covariance thesis**
> Our subjective concerns (and consequent motivations and behaviour) covary fairly closely with our moral beliefs, in the sense that these concerns, motivations, and behaviours are often influenced by our moral beliefs.

[88] It is admittedly an empirical question just how often these *de re* moral desires falter. But even if they only falter now and then, it would still seem useful to have some sort of safety-net in place, given that the stakes can often be high in interpersonal relationships.

> **(3) Usefulness of morality**
> Moral beliefs and the discourse that gives expression to them provide us with a number of desirable practical goods in virtue of the behaviour and motivations that they facilitate.

When these claims are put together, they carry a worrying implication. (**3**) tells us that morality is useful in virtue of the motivations and behaviour that it facilitates. But (**2**) suggests that these motivations and behaviour depend (to some degree) upon our moral beliefs—that is, upon our judging certain things to be right and wrong. The moral error theory therefore seems to endanger the desirable practical goods that our moral practices provide. For (**1**) reminds us that moral belief is no longer an option for us.

Or is it? It might be thought that the correct answer to the WNQ ought to strike us as obvious; we should simply hold onto our false moral beliefs! (This is not to assume any strong form of doxastic voluntarism. Even if we lack direct voluntary control over our beliefs, it is plausible that we have some measure of *indirect* control over what we believe. I discuss these issues at length in §7.4.) Of course, this strategy would involve holding onto beliefs that we take to be systematically false. But is there anything that counts significantly against our doing so?

Indeed, there is. The practice of intentionally preserving false beliefs is arguably the cardinal sin of philosophical inquiry.[89] As philosophers, we commit ourselves to maintaining a suitable level of epistemic hygiene. And philosophers are certainly not the only ones with truth-seeking goals; non-philosophers presumably have good reasons for maintaining true beliefs as well. To persist with beliefs that we take to be false would be contrary to various epistemic values that we hold.

Even those who do not value truth *per se* should be wary of being too epistemically cavalier. True beliefs are valuable resources. Typically, we're more likely to be effective at satisfying our desires if we have a stock of true beliefs to act upon rather than a stock of false ones (Joyce, pp.178-9). So we should be wary of racking up too

[89] The 'intentionally' is important here. Philosophers would be engaging in pretty risky behaviour if holding onto false beliefs were the worst thing one could do, philosophically! Thanks to Daniel Nolan for pointing this out.

many false beliefs. It therefore seems that we need to add a fourth claim to the package above:

> **(4) Epistemic values**
> There is a presumption against having false beliefs.

(4) clearly exacerbates the situation; for we can now see that our believing the moral error theory gives rise to a tension. On the one hand, epistemic considerations seem to speak in favour of doing away with our erroneous moral discourse; straightforward moral belief is no longer an option for us if we believe the moral error theory.[90] But purging ourselves of moral discourse does not seem all that appealing; we are likely to lose many of the benefits of moral practice should we do so. The project of answering the WNQ for moral discourse addresses this tension. In pursuing this project, we are attempting to navigate a middle path between these two extremes—to identify a means by which we can reap (at least many of) the benefits of moral discourse without throwing all epistemic propriety out the window.

So, what are our options?

§2.4 Answers to the WNQ for moral discourse

We can distinguish different answers to the WNQ for moral discourse by how they respond to claims (1)-(4) above.[91] Some resolve the tension by modifying or putting pressure upon one of these claims, whereas others dissolve the tension by rejecting one of them. In the remainder of the thesis, I will evaluate each of these suggestions in turn, and develop the proposal that (to my mind) has the greatest promise: conservationism. Here, I will also mention and set to the side a *propagandist* option.

[90] This is not to suggest that the right answer to the WNQ *cannot* involve intentionally preserving false beliefs. It is only to suggest that someone who thinks that we should preserve false beliefs must motivate overriding the strong presumption against doing so.

[91] The taxonomy is not airtight, given that modifying or rejecting one of the four claims can have implications for how one understands the other claims. But I do think that the heart of each proposal is nicely captured by the claim to which it directs its focus.

§2.4.1 *Abolitionism*

Moral abolitionists recommend that we do away with moral practice altogether; they advise us to cease using moral language, thinking moral thoughts, and invoking moral considerations when deliberating about what we ought to do.

Some abolitionists appeal to the epistemic value of doing away with our erroneous moral beliefs when motivating their proposal (e.g., Garner 2007, p.500; Burgess 2007, p.438). For the most part, though, abolitionists tend to be motivated by (what they take to be) the significant *harms* of engaging in moral practice. According to them, moral conviction breeds and perpetuates interpersonal conflict, and moral language all too often lends a helping hand to war and violence.[92] (See Hinckfuss 1987*,* §4.2; Greene 2002, pp.233-6 & pp.338-9; Garner 2007, pp.502-3; Ingram 2015, pp.240-1.)

As should be clear, the abolitionist dissolves the tension by targeting (3). She will concede that there are *some* benefits to engaging in moral practice. But she denies that these practices are useful *on-balance*. Indeed, she thinks that we would be far *better off* without them. In her view, any tension between our epistemic values and our practical interests here is merely apparent.

§2.4.2 *Revisionism*

Revisionists advise us to modify moral discourse such that our engagement in moral practice no longer commits us to the existence of categorical reasons. There are different ways to go about this task. We might change the way that we use moral language; moral utterances could come to express noncognitive attitudes (e.g., approval and disapproval) rather than beliefs (Köhler & Ridge 2013, Svoboda 2017). Since the resultant utterances wouldn't be truth-apt, they wouldn't be systematically false. Alternatively, we might modify the conceptual commitments of moral discourse, replacing our error-ridden moral discourse with an error-free, *schmoral* one (Lutz 2014).

Revisionists therefore propose to resolve the tension by directing their critical focus to (1). Some think that the moral error theorist *can* believe that some things are

[92] Given the nature of these latter concerns, abolitionism can be motivated quite independently of moral error theory—not all abolitionists are in the business of addressing our WNQ. But their arguments will still be relevant to our investigation.

right and wrong, albeit in a qualified sense; she can believe that some things are right* and wrong*. Others claim that there is no great worry if the error theorist cannot *believe* that some things are right and wrong; for she can simply replace these moral beliefs with non-cognitive attitudes.

The revisionist's guiding hope is that our beliefs about what is right* and wrong* (or the relevant non-cognitive attitudes) will influence our motivations and behaviour in the same (or at least in a very similar) way as our beliefs about what is right and wrong. It is this hope that grounds her expectation that schmoral discourse will deliver sufficiently similar practical goods to moral discourse. Revisionism would therefore seem to align with our practical needs. And since the revisionist does not ask us to hold onto our erroneous moral beliefs, her proposal would seem to align with our epistemic values as well.

§2.4.3 *Fictionalism*

The fictionalist's pitch is similar to the revisionist's in some respects; she too wishes to enjoy the benefits of moral discourse without incurring the relevant epistemic costs. The fictionalist proposes to do so by preserving moral discourse in the form of a useful fiction, replacing our moral beliefs with fictionalist attitudes (Nolan Restall, and West 2005, Joyce, ch.7 & ch.8). Unlike beliefs, these attitudes are not metaphysically committing.[93] The moral fictionalist only believes (or make-believes) that there are moral truths within some moral fiction or other. This does not commit her to the claim that anything is wrong *simpliciter*.

As I understand fictionalists, their strategy is to resolve the tension by putting pressure upon (2). According to the fictionalist, it is not only full-blooded moral beliefs that are capable of influencing our motivations and behaviour; our actions and our motives can be influenced by fictionalist attitudes as well—and in very similar ways. Given this, the moral fictionalist thinks that she can secure many of the benefits associated with moral practice.

[93] Of course, fictional attitudes may be metaphysically committing *in some respects*. (Though just what they commit us to is likely to depend upon one's metaphysics of fictions.) The point is that they are not metaphysically committing in the *relevant* respects; the fictionalist about morality, for example, is not committed to the literal existence of moral rightness or wrongness.

§2.4.4 *Conservationism*

Conservationists take moral discourse to be incredibly useful. But they part ways from the fictionalist and the revisionist in recommending that we *hold onto* our false moral beliefs. Interestingly, conservationists think that we ought to continue to believe the moral error theory as well. They usually recommend that we attend to our belief that the moral error theory is true in some contexts, and attend to our beliefs that particular actions are right or wrong in others (Olson 2014, pp.190-6).

Insofar as she thinks that the moral error theory merely prevents us from attending to our beliefs that some actions are right or wrong in particular contexts, the conservationist resists (1). But she is distinguished by how she responds to (4). Though the conservationist concedes that there is a presumption against having false beliefs, she argues that this presumption can be overridden in the event of a moral error theory. Our practical interests win out against our epistemic values under such circumstances; for the conservationist strategy is *the best*—or better, *the only*—means of effectively serving our real-world interests here. Other proposals that *are* in alignment with our epistemic values simply don't carry the same practical promise.

§2.4.5 *Why not propagandism?*

It might be thought that we philosophers would do best to keep uncomfortable truths hidden from non-philosophers or (as they are sometimes more affectionately known) 'the folk'. And the moral error theory strikes one as a very uncomfortable truth indeed. Why, then, do we not simply keep this inconvenient truth to ourselves, and allow the folk to continue moralising as before?

This suggestion often goes under the name 'propagandism'.[94] To my knowledge, it has received only one defence in the recent literature (that of Cuneo & Christy 2011). (Though this is not so much an outright defence as it is an argument to the effect that propagandism is superior to fictionalism in a number of respects.) Given consid-

[94] Strictly speaking, propagandism is not a response to the WNQ as I have formulated it; for I have presumed that *everyone* is *already* in the know regarding the truth of moral error theory. Propagandism would presumably do best if implemented before the truth of moral error theory had become well-known (though even then, its prospects would be uncertain).

erations of space, I shall not devote an entire chapter to assessing this proposal. Instead, I shall set propagandism to the side; for I do not regard it as a viable option.

There are a number of issues regarding how we might best go about implementing propagandism, and each of these, I believe, reveals a serious potential for instability. To begin with, there is the question regarding just who is to safeguard the truth about morality. Presumably, we would want to minimise risk, and entrust this dangerous secret to a small class of philosophers. To be on the safe side, we might even instruct them *not* to pass on this information to future generations. But even with these safeguards in place, propagandism still seems woefully unstable. After all, there is nothing to prevent contemporary philosophers not in the know (or those of the future) from discovering the truth about morality on their own accord.

This invites a second question: to what lengths we are prepared to go to in order to protect the public? If we are seriously committed to keeping the truth well-hidden, then we should presumably remove all error-theoretic texts (as well as commentaries and associated discussions) from circulation. But would we really be willing to commit the works of Mackie, Joyce, and Olson to the flames? I have argued that our believing the moral error theory has some potentially devastating consequences. But these consequences surely weren't so dire as to warrant doing away with an entire body of philosophical research. The more propagandism approaches stability, the more it risks becoming hysterical.

There are other reasons to be concerned about the stability of propagandism. As Köhler and Ridge (2013, p.438) point out, its success is likely to require "systematic intellectual dishonesty, deception and elitism" (see also Joyce, pp.214-5). And it seems unrealistic to expect that those in the know could keep up this ruse for very long. Moreover, doing so is likely to conflict with other values that they hold. The outright suppression of philosophical arguments (for fear of their dangerous consequences) is not something that we typically want to encourage.

Accordingly, I do not think that propagandism should even strike us as even initially appealing. The more we do to ensure its long-term stability, the more hysterical propagandism seems, and the more the proposal conflicts with too many of our other values. And the less we do to ensure its long-term stability, the more likely it is to leave us back where we started; everyone will be privy to the uncomfortable truth about morality, and we will still require an answer to our WNQ.

§2.4.6 *Evaluating answers to the WNQ*

The proposals outlined above are not unique to the moral domain. Fictionalism, for instance, has been developed in response to error theories about numbers (Field 1980, Balaguer 2009, Leng 2010), and colours (Boghossian & Velleman 1989), as well as agnosticism about unobservables (van Fraassen 1980). The moral fictionalist thinks that we can apply the same solution to moral discourse.

A recurring theme of this thesis will be that we must proceed with caution when importing such solutions. By this, I certainly *don't* mean to suggest that moral fictionalists themselves have not given the relevant issues any thought. But I do worry that they (along with others) have neglected to pay sufficient attention to the distinctive features of a discourse in virtue of which it is apt for abolition, revision, fictionalisation, or conservation.

As I will aim to show, particular solutions to WNQs are fitting for certain discourses in virtue of the distinctive features of those discourses. We must attend to the distinctive features of moral discourse more closely before employing dialectical moves that have been made in other error-ridden discursive domains. Only by doing so can we determine whether philosophically rewarding moves can be made in the moral analogue.

There are a number of qualities that will contribute to the attractiveness of any candidate proposal. I will argue that all else being equal, we ought to favour a solution to the WNQ for moral discourse insofar as it is (or can reasonably be expected to be) feasible, stable, in accordance with our epistemic values, and likely to preserve the desirable practical goods that our error-ridden moral discourse provides—or at least, a significant chunk of them. Different proposals may very well exhibit these virtues to different degrees; it is a comparative game that we are playing.

A small caveat is in order before proceeding. The proposals canvassed above are to my mind the strongest (and to my knowledge, the only) contenders. However, it is possible that they do not *exhaust* the options that are available to us. (There is more in heaven and earth than is dreamt of in philosophy.) If they do not, then my recommended solution should be understood as a conditional claim: *if* these are the best proposals on offer, *then* we ought to opt for proposal *x*.

§2.5 Philosophical dividends

Providing an answer to the WNQ for moral discourse is valuable in a number of respects. First, doing so responds to the worry that our believing the moral error theory endangers many of the desirable practical goods that our moral practices provide. Assuaging this concern may have implications for the acceptability of moral error theory as a meta-ethical position. It is not unlikely that at least some of the resistance to moral error theory is underwritten by a fear of the significant practical costs of embracing it. Indeed, some suspect that "…*much of the opposition* to the moral error theory is motivated by…an inchoate practical fear of what might happen should [it] be widely adopted" (Joyce & Kirchin 2010, p.xiv, emphasis added). Others have speculated that "…if a fully satisfying solution [to the WNQ] can be found, the error theory will become a more respectable position within metaethics" (Lutz 2014, p.370). By providing a principled and well worked-out answer to the WNQ, then, I hope to show that our believing the moral error theory need not entail the loss of many desirable practical goods; the position ought not be rejected on account of its feared practical consequences.

Second, answering the WNQ for moral discourse can be useful even for those who are currently unpersuaded by the arguments for moral error theory. Perhaps there being a non-negligible chance that the moral error theory is true would suffice (Köhler and Ridge 2013, p.430). As things currently stand, we certainly can't rule out the possibility that the moral error theory is true. So it may be wise to formulate a contingency plan.

Third, we might be interested in what the effects of our moral practices are, and in what way(s) beliefs in moral truths or a commitment to moral realism contributes to them. Reflecting upon what might happen if people came to endorse the moral error theory might be illuminating in this respect.[95] We might also wonder whether some abolitionists are correct in thinking that the costs of moral discourse—whether erroneous or not—are too high.[96] Even a moral realist may be interested in these issues.

[95] I thank Daniel Nolan for this suggestion.

[96] I thank Ben Fraser for this suggestion.

Fourth, providing an answer to the WNQ for moral discourse fills an important gap in the current philosophical literature. The issue concerning what to do when 'discourses go bad' is a question currently occupying the minds of philosophers of mathematics, mind, and even those who have embraced an error theory about colour. Yet while many answers to the WNQ have been offered for these other error-ridden discourses, few proposals have been properly developed in the moral case. The philosophy of mathematics, for example, offers a far richer and well worked-out array of answers to the WNQ that confronts mathematical discourse if there are no numbers. (See for example, Field 1980, Putnam 1967, Yablo 2001, and Balaguer 2009.) When compared to their analogues in other philosophical domains, post-moral-error-theory proposals are relatively unchartered philosophical territory. This is unfortunate. A world without moral goodness seems to me to be equally concerning (if not more) than one without the number seven. This research project goes some way towards filling that gap, and has the potential to enrich the dialectic within a meta-ethical context.

Finally, broader lessons can be drawn from my discussion of the conditions under which revisionism, abolitionism, fictionalism, and conservationism would be fitting solutions to the WNQ for moral discourse. A recurring theme of the discussion will be that different error-ridden discourses invite different solutions—and for principled reasons. By pointing towards such reasons, I hope to point towards some domain-general criteria for answering WNQs more broadly. The project therefore has implications for how we ought to answer the WNQ in other discursive domains as well. And as will become clear, this is a question that rears its head surprisingly often.

§2.6 TAKE AWAY

The purpose of this chapter was to clarify and motivate the project of answering the WNQ for moral discourse. I have specified that the WNQ is a *collective*, *normative* question; it is the question regarding what we, as a linguistic community (or, at least most of us with similar concerns), ought to do with our moral discourse if we believe the moral error theory.

It was also important here to motivate the project of addressing the WNQ. To this end, I have suggested that the moral error theory gives rise to a tension between various kinds of values and interests that most of us share. If we are to address this tension, then we must answer the WNQ.

The purpose of these first two chapters has been to set up the remainder of the thesis, and to motivate the question that will form its central focus. I hope that these preliminaries have convinced the reader that moral error theory is a position to be taken seriously, and that the WNQ is worth addressing. With that background out of the way, we can now embark upon the project of answering the WNQ for moral discourse. Let's get started.

Chapter Three

Moral Abolitionism

Our first proposal to consider is moral abolitionism. Abolitionists advise us to cease using moral language, thinking moral thoughts, and invoking moral considerations when deliberating about what we ought to do. Initially, this may seem like peculiar advice; for (as was suggested in §2.3.3) our moral practices seem to provide us with a great number of desirable practical good. But abolitionists think that we have underestimated *the harms* of these practices. In their view, we'd be better off without them.

The chapter begins by singling out the variety of moral abolitionism of interest (§3.1). Abolitionist proposals can differ along a number of dimensions, such as their intended scope. Abolitionists do, however, tend to be motivated by the same—if not very similar—concerns. The majority are motivated by considerations having to do with both (i) the practical costs of moral practice, and (ii) the epistemic disvalue of holding onto our false moral beliefs. For the most part, (i) seems to be load-bearing; quite a number of abolitionists think that the significant practical costs imposed upon us by moral practice are sufficient to justify doing away with it—that we have ample reason to do so quite independently of moral error theory.

The critical discussion takes a closer look at the case for abolitionism. There are two arguments in particular that we shall consider. I will refer to these as The Argument from Conflict (§3.2), and The Argument from History (§3.3). Both are intended to support moral abolitionism and both, I believe, fail. What these arguments do establish is that morality can be of great benefit or great harm, depending upon the manner in which it is used. However, the right response to this problem isn't necessarily to do away with morality. A better response is to seek a means by which we can reap the relevant benefits while avoiding the associated costs. I will argue that there are promising means by which we could control for the costs that abolitionists bring to our attention. If I am right, then the abolitionist's arguments do not so much support abolitionism as they support reform.

If my arguments in §3.2-§3.3 are successful, then abolitionism is radically under-motivated. In §3.4, I will suggest that we also have good reason to worry about the feasibility of the proposal; ridding ourselves of morality may very well be something that we cannot do.

As one can glean from the blueprint above, my considered verdict will be that abolitionism is the wrong response to the WNQ for moral discourse. The discussion concludes with a diagnosis of where abolitionism goes wrong (§3.5).

§3.1 WHAT IS MORAL ABOLITIONISM?

Abolitionist proposals differ along a number of dimensions, including the scope of their ban on moral vocabulary. And while they tend to differ less in their guiding motivations (most are driven by both epistemic and practical concerns), there is some disagreement as to how much weight each sort of motivation ought to carry. Since I will not be taking issue with every proposal in the abolitionist neighbourhood, an important task for §3.1 will be to distinguish the species under investigation from other varieties.

§3.1.1 *Motivation*

The majority of abolitionists are motivated by (what they take to be) the significant practical costs of engaging in moral practice. According to them, our moral systems breed and perpetuate interpersonal conflict, too often lend a helping hand to war and violence, and render our societies authoritarian and elitist. (See Hinckfuss 1987, §4.2; Greene 2002, pp.233-6 & pp.338-9; Garner 2007, pp.502-3; Ingram 2015, pp.240-1.)

Of course, abolitionists don't think that there are *no* benefits to engaging in moral practice. That morality can be useful is something they concede. (See for instance, Garner 2007, p.504.) But though it will be admitted that morality can be used for (non-morally) good purposes, abolitionists think that it can be used for (non-morally) bad purposes as well. And though the good here may be very good, the bad is especially horrid.[97]

These practical concerns should give us pause even if we are not moral error theorists. Quite a number of abolitionists regard the truth or falsity of our moral judgments as an orthogonal issue here; they think that the practical costs imposed upon us by moral discourse are *sufficient* to justify its elimination (Hinckfuss 1987, §1.5). Stephen Ingram has boldly suggested that even moral realists have good reason to do away with moral language (2015, pp.242-3). Many of the arguments that support abolitionism are therefore worthy of consideration quite independently of our metaethical tastes.

That said, the truth of moral error theory may very well be further grist for the abolitionist's mill. In the event that morality can boast neither truth nor usefulness, perhaps we would have doubly good reason to get rid of it. Indeed, abolitionists sometimes speak as though moral error theory and abolitionism is a package deal. Garner, for instance, describes the conversion to abolitionism as a two-stage process—one that in the first instance requires the initiate to take her moral beliefs to be systematically false (2007, p.500; see also Burgess 2007, p.438).

Historically, then, abolitionists have tended to disagree upon the role that error-theoretic considerations play in motivating their proposal—some take it to be a necessary first step, whereas others merely regard it as a helpful supplement. But the majority do seem to be moved by both practical and epistemic considerations, even if these considerations move them to different degrees. I shall therefore take both sorts of considerations to be important motivations for the proposal. (Though for reasons mentioned in §3.2, I will devote most of my critical focus to the former.)

[97] Of course, many abolitionists take issue with the fact that moral practice is often put to beneficial use for *a limited subset* of the population; for the elite who want to preserve inequality, say. (See for example, Hinckfuss 1987, §§2.3-2.4.) I attend to these subtleties in §3.3. For the time being, 'bad purposes' can be read as purposes that are harmful to most people.

§3.1.2 *Scope*

The species of moral abolitionism with which we shall be concerned is something of a minority view. However, more tempered varieties are not so uncommon. It has been suggested that we ought to eliminate the term 'evil' from our vocabulary on account of the social harms that it facilitates (Held 2001, p.107; Cole 2006, p.21). Others have argued for the abolition of particular moral terms on largely error-theoretic grounds. Jeremy Bentham (1792/1843, p.501), for instance, was in favour of eradicating talk of natural rights, which he regarded as "nonsense upon stilts".

Our abolitionist is different. She is not merely picking and choosing among bits of moral vocabulary. It is her contention that *all* of moral language ought to be consigned to the scrap heap.[98] Indeed, the net that she casts is even wider still. Our abolitionist recommends that we do away with *moral practice* in its entirety; she does not only advise us to cease using moral language, but also to stop thinking moral thoughts, using moral concepts, judging others along moral dimensions, and invoking moral considerations when deliberating about what we ought to do (Ingram 2015, p.231 & p.236; Garner 2007, p.500 & p.504; Fraser 2017, p.158). To be clear, the advice here is to do away with moral practices *of any kind*—revised, fictionalised, or otherwise.[99] And it is advice for our *moral* practices in particular—not our normative practices more generally. Most if not all abolitionists think that we should and can in good conscience continue to speak of what we ought (non-morally) to do. (See for example, Hinckfuss 1987, §1.3.1; Ingram 2015, p.236.)

§3.1.3 *A new hope?*

It is no exaggeration to say that abolitionism fails to make a good first impression. That the proposal is *prima facie* unattractive is something even its supporters concede:

[98] Some abolitionists do express a particular distaste for deontological language (e.g., Greene 2002; Hinckfuss 1987, §4.6). However, none think that it follows from the fact that deontology makes for particularly good target practice that the rest of moral discourse is off the hook. It's also worth noting that the abolitionist's recommendation that we purge ourselves of 'moral language' is plausibly restricted to moral language within the context of first-order moral practice. Moral abolitionists don't necessarily think that we should stop using the terms 'right' and 'wrong' when doing meta-ethics.

[99] This advice is most explicit in Garner (2007) and Ingram (2015). Both are concerned to motivate abolitionism over rival solutions to the WNQ which involve preserving moral discourse in some form.

> If someone told you that we should get rid of moral discourse and moral judgement, you'd probably raise an eyebrow or two…the idea that we should eliminate them seems strange. Indeed, it seems repugnant. What sort of person would want to eradicate the right and the good from their conceptual repertoire? Not a very nice one, you might think. (Ingram 2015, pp.227-8)

But perhaps our initial aversion to abolitionism rests upon a misunderstanding. We must take care not to conflate two senses of 'moral' (Greene 2002, pp.19-21):

> **Moral$_1$**
> Of or relating to the facts concerning right and wrong, etc.
>
> **Moral$_2$**
> Of or relating to serving (or refraining from undermining) the interests of others.

As abolitionists are especially concerned to emphasise, their claim is that we ought to do away with morality$_1$—to stop classifying certain actions or persons as morally good or bad, say. Their claim is *not* that we ought to do away with morality$_2$—to cease caring about others or their interests. Abolitionists are strongly in favour of cultivating morality$_2$. Indeed, the cornerstone of their practical proposal is the idea that we'd have *far more* morality$_2$ without morality$_1$ (Garner 1994, p.3; Greene 2002, pp.47-8; Marks 2013a, p.2). In their view, purging ourselves of morality may very well be "…an essential step in achieving many of the goals well-meaning moralists…have always cherished" (Garner 2007, p.51; see also Hinckfuss 1987, §5).

§3.1.4 *A caveat*

Most arguments for abolitionism are accompanied by an important concession: they are speculative. It seems that we simply "cannot know whether a thorough cost-benefit analysis of moral discourse and judgement would favour…abolitionism" (Ingram 2015, p.242). The question as to whether we would be better off without morality, is, after all, an empirical question—and perhaps "one to which we are not likely to find a definitive answer" (Garner 2007, p.506).[100]

[100] Indeed, Hans-Georg Moeller (2009) thinks that we ought to be *agnostic* as to whether morality does more harm than good, and recommends that we be more sparing with our use of moral language on account of this.

I attach a similar cautionary note to my critical discussion of abolitionism as abolitionists do to their defence of the position; many of my arguments will be speculative. That said, I am optimistic that we can get at least *some* traction on the relevant issues. Though the case may not be completely decisive either way, we do seem to have enough resources to make a principled assessment.

It's worth noting here that it is not merely the moral error theorist who may need to engage in such speculation. Some abolitionists, recall, think that the harms of moral practice are sufficient to justify its elimination. If they are right, then other meta-ethicists may also need to ask themselves whether they'd be better off without morality.

§3.2 The argument from conflict

We are now in a position to take a closer look at the case for moral abolitionism. Our focus will be restricted to the purely practical justification for the proposal. Moral abolitionism does of course have epistemic benefits as well—for enacting it involves doing away with our false moral beliefs. But it is the practical justification for abolitionism that is typically thought to be load-bearing. Targeting that justification should therefore suffice to cast doubt upon its plausibility.

The practical justification for abolitionism is premised upon the claim that our moral practices do more harm than good. Abolitionists have developed a number of arguments in support of this claim. But I shall pick my battles carefully here, restricting my focus to those arguments that strike me as most promising: The Argument from Conflict (AFC) and The Argument from History (AFH). (This section is devoted to the former.)

It's worth clarifying my ambitions before proceeding. It will not be my intention to establish that the abolitionist is wholly off the mark. She is perfectly right in thinking that there are disadvantages to engaging in moral practice. (Though, as we shall see, she does have a slight penchant for exaggeration.) What I want to put pressure on is the move from these disadvantages to abolitionism. The move is slightly too swift; for there are conceivable means by which we could minimise these disadvantages. If I am right, then abolitionism does not seem to be the right response to the problems that abolitionists raise. A better response would be to improve ourselves as interlocutors, empathisers, and critical thinkers.

§3.2.1 *Morality and interpersonal conflict*

A popular argument in favour of abolishing morality appeals to the role that moral considerations play in exacerbating interpersonal conflict.[101] This isn't to say that we don't butt heads over non-moral matters as well; we might disagree, for instance, over how to best divide an inheritance, or how to budget for an upcoming wedding. But at least in cases of conflicting preferences and interests, there seems to be a light at the end of the tunnel. In such cases, we can hope to meet one another half way. Compromise isn't typically off the table.

Matters seem different in cases of moral disagreement. If I think that we morally ought to refrain from genocide, and you think that we ought to go for it, then ought I to meet you half way by agreeing to wipe out just half of the relevant population? I don't think so. It seems that we often feel entitled to stand our ground when we butt heads over moral issues. (See Enoch 2011, ch.2.)

Yet it is precisely this feature of moral disagreement that abolitionists find problematic. Though compromise may not be either party's preferred option in a practical dispute, it is often better than a long-standing impasse. But instead of compromising on moral issues, we often find ourselves locked into interminable debates. The relevant worry isn't merely that disagreements regarding abortion, parricide, and marriage equality seem unlikely to be settled any time soon—it's that they seem unlikely *ever* to be settled. When these practical disagreements (i.e., disagreements about what to do) become *moral* disagreements, they become frustratingly persistent, and no resolution seems forthcoming (Hinckfuss 1987, §4.3; Garner 2007, p.502; Marks 2013b, p.446; Ingram 2015, p.240).

It's worth noting why this is *a problem*; for the persistence of a disagreement shouldn't always trouble us. There may very well be no conceivable end to the debate

[101] For ease of expression, I will sometimes speak loosely of moral considerations being harmful, or helping along harmful agendas. What I really intend by this is *the invocation of* moral considerations in discourse and deliberation.

as to whether *The Empire Strikes Back* is the best movie in the *Star Wars* canon. But no one thinks that we ought to stop watching *Star Wars* on account of this. However, sometimes the persistence of a dispute is cause for concern. An unresolved disagreement can have far-reaching consequences upon people's lives. It is yet to be decided in some sectors of the world whether (certain forms of) assisted suicide should be legalised. The longer this dispute persists, the longer people may suffer unnecessarily. Generally speaking, unresolved disputes are undesirable, since we often stand to benefit from collective action. Longstanding disputes can also entrench existing social divides, making large-scale co-operation more difficult in the long-run.

Abolitionists think that practical disagreements will be far easier to resolve following a wholesale purge of moral language. They don't of course think that this will solve all of our problems. Nonetheless, abolitionists do think that having stripped practical disagreements of their moral overlay, all that will remain are mere conflicts of interest—something that we (supposedly) have a far better shot at resolving (Garner 2007, p.502; Ingram 2015, p.241).

Of course, *qua* moral error theorists we cannot understand moral disagreements or their resolution in the usual terms. We cannot, for instance, understand a moral disagreement as always revolving around a set of claims which are mutually exclusive and jointly exhaustive, where exactly one party to the dispute will be speaking the truth. We need to allow that there will be genuine moral disagreements even where no party's position is correct. This should not strike us as *that* unusual. Atheists, for example, can plausibly make sense of two theists being engaged in a genuine disagreement (Sturgeon 1994, p.82).[102]

Moreover, and for similar reasons, we cannot take the resolution of these disputes to involve convergence upon the moral *truth*. I therefore suggest that we take resolution to consist in all parties reaching a justified consensus by epistemically respectable

[102] Matters may be slightly complicated by our having taken the existence of moral facts to constitute *an impossibility*. But I am inclined to regard this as a technical complication. The most plausible account of (im)possible world semantics should be able to make sense of reasoning about the consequences of views that are necessarily false. (Some metaphysical views currently on the market may turn out to be necessarily false, but we still seem capable of reasoning about what follows from them.) I leave it to the philosophers of language to do so. For a helpful discussion, see Nolan (1997).

means.[103] (I assume here that justification is fallible; one can be justified in believing false propositions.) I attach the qualifications 'justified' and 'by epistemically respectable means' so as to rule out as cheap and uninteresting ways in which consensus could be achieved (brain-washing, intimidation, and what-have-you). What is ruled in is another matter. What I have in mind, broadly speaking, are epistemically respectable procedures; those that involve evaluating arguments, taking one's evidence into account, trying to avoid inconsistency, and the like.

§3.2.2 *The argument from conflict: an assessment*

In order to properly assess the argument from conflict, we must ask *why* abolitionists think that the moral overlay makes practical disputes frustratingly persistent. Some (e.g., Burgess 2007) seem to put this down to the fact that moral disagreements are irresolvable in principle. Given this, the thought seems to be that a practical dispute whose resolution comes to depend upon the resolution of a moral dispute will likewise come to be irresolvable in principle. Others (e.g., Ingram 2015) think that moral disagreements are irresolvable in practice. Thus, as soon as the resolution of a practical dispute comes to hinge upon the resolution of a moral dispute, the practical dispute comes to be irresolvable in practice as well. We will examine each of these ideas in turn.

Before we do though, I should note from the outset that I am not going to try to settle here whether or not moral disagreements are resolvable in principle; whether they would persist even when all parties were free of "anything worth calling a cognitive shortcoming" (Enoch 2011, p.209). What I want to do instead is emphasise that this is a *highly* contentious issue, and to point out that abolitionists have done very little to motivate going one way or the other.

People often disagree upon particular moral issues—upon whether "plants or animals or zygotes have rights", say (Burgess 2007, p.436). Burgess suggests that such disagreements are in principle irresolvable because their resolution would require an-

[103] It seems possible for parties to reach a justified consensus by epistemically respectable means, even when they converge upon a falsehood. Members of The Vienna circle arguably did so when they came to agree upon the truth of logical positivism (a view which is today widely regarded as false).

tecedent agreement upon fundamental moral values or “general principles”. Given that there is substantial *divergence* among our fundamental moral values, we cannot—according to this pessimistic line of thought—ever hope to achieve convergence on specific moral issues.

But there is also the following, *optimistic* position: people do in fact share a common set of core moral values—moral disagreements are simply owing to differences in their non-moral beliefs (Rachels 2012, p.155). Generally, the optimist’s strategy consists in reducing putatively moral disagreements to non-moral disagreements. If we could only supply everyone with full empirical knowledge and equip them with the necessary powers of reasoning, then there’d be nothing left to quibble about. People wouldn’t continue to disagree upon whether zygotes had rights if they knew a little more about biology. Some are confident that this strategy can be applied to all (or almost all) instances of moral disagreement (e.g., Brink 1984, pp.116-7; Boyd 1988, p.213). Others are more hesitant. They suggest that the general applicability of this strategy is questionable (Enoch 2011, p.191), or that we cannot make any reasonable assessments of its promise without a good deal more empirical research (Loeb 1998, p.284).[104]

Matters are more complicated still. *Pace* Burgess, it is not even clear that the resolution of moral disputes would require antecedent agreement upon fundamental moral values. As Brink points out, this assumes “a one-way view of moral justification and argument”, on which people’s core moral values are foundational, justifying “particular moral judgments but not vice versa” (1984, p.116). And this, of course, is something that coherentists about moral justification will deny. (Achieving coherence among moral commitments may sometimes require giving up general moral principles rather than specific moral beliefs.) But even given coherentism, it still seems to be an open question whether or not the relevant idealised agents would converge

[104] This is not to suggest that no empirical research has been brought to bear on the issue. Doris and Plakias (2008), for example, draw upon Nisbett and Cohen’s (1996) research on ‘cultures of honour’ when discussing the prospects of resolving moral disagreements. It is only to suggest that there is arguably more work to be done.

upon the *same* set of moral commitments. It is difficult to know either way.[105] Indeed, I am not even confident that the debate as to whether moral disagreements are resolvable in principle *is itself* resolvable in principle. (On this issue, see Shafer-Landau 2003, p.223.)

None of this is to suggest that moral disagreements *are* resolvable in principle. I have only suggested that we shouldn't be at all confident either way. The matter rests upon highly contested claims regarding the nature of moral justification, and the outcomes of hypothetical disputes among idealised agents. This does not establish that the abolitionist is *wrong*. But it is nonetheless an interesting result, and one that I believe counts against her. If the entire case for the AFC were to rest upon the in principle irresolvability of moral disagreements, then the argument would be far from decisive.

That said, the abolitionist may happily grant to us that ideal agents would reach a consensus on moral issues. She need only point out that *we* are not ideal agents. Given the way we are, and given the ways in which we are prone to use moral discourse, perhaps introducing morality into the equation *does in fact* prevent us from resolving practical disagreements. Perhaps these disputes are irresolvable *in practice*.

At this stage, the abolitionist owes us an explanation as to why this is so. Stephen Ingram has risen to the occasion, emphasising the *obstinance* that moral conviction tends to encourage. This, he suggests, is owing to

> … the categorical authority of moral ascriptions. Although this authority might help combat weakness of will, it can also lead to opposing parties becoming entrenched in their positions, making it harder to bring conflicts to a satisfactory conclusion. (2015, p.239)

If Ingram is correct, then the persistence of moralised disagreements is the result of people having become "entrenched in their positions". Given the categorical authority of moral prescriptions, it is easy to become entrenched in one's views regarding the moral permissibility of abortion—far easier than it would be to become en-

[105] Indeed, criticisms of Smith's (1994) claim that our ideal selves would converge upon their desires suggests that it is very difficult to establish that norms of coherency will take rational agents from disparate starting points to full agreement. See Joyce (pp.75-7), Sobel (1999) and Bukoski (2016).

trenched in one's views regarding the aesthetic permissibility of wearing socks with open-toed shoes.

Interestingly, certain findings in moral psychology may back up what Ingram has to say here. It has been found that insofar as experimental subjects take moral truths to be objective (i.e., take their truth to be independent of any individual's or group's attitudes), they are prone to be intolerant of diversity and ineffective at dealing with conflict (Goodwin and Darley 2008, 2010, 2012; Wright et al. 2014).[106] Of course, the notion of objectivity is distinct from that of categorical authority. But inasmuch as people objectify morality as well, these findings may be further grist for the abolitionist's mill (see Fraser 2017).

That said, it shouldn't be obvious to us that it is morality (or, at least, *only* morality) that breeds obstinance in such cases. We should be wary of inferring from the correlation between objectivist moral intuitions and inefficiency in conflict resolution that the moral overlay is the most important contributing factor in ongoing practical disputes. Goodwin and Darley (2010) also found that moral objectivists (as opposed to non-objectivists) tend (i) to have little interest in identifying or understanding the source of their disagreements—i.e., where the other person is coming from, and (ii) to perform worse on disjunctive reasoning tasks (which tests for abilities to entertain different possibilities, and approach problems from different perspectives). (See also Feltz & Cokely 2008.) It is not unlikely that the difficulty that moral objectivists have resolving moral conflicts is owing (or partly owing) to *these* factors. If someone is *already* unwilling or unable to understand others perspectives, then resolving disagreements is likely to be difficult in any case.

Thus, it seems to me that we can grant much of what Ingram has to say while resisting the inference that moral disagreements are irresolvable in practice. We have good independent reasons to consider other perspectives, and to avoid becoming too deeply entrenched in our own views (moral or otherwise); for we often have good reasons to resolve interpersonal conflicts, and we are unlikely to do so if we are not sufficiently open-minded. The right response to the problem that Ingram raises

[106] I say 'insofar as' because Goodwin and Darley's (2008, 2010, 2012) findings suggest that people differ in the extent to which they are objectivists about morality, and that some moral claims are more likely to be considered objective than others. (See also Wright et al. 2014; Sarkissian et al. 2011.)

would surely be to improve ourselves as interlocutors; to cultivate dispositions to be more open-minded, to be more friendly to the possibility that we could be mistaken, and to become more responsive to arguments for opposing views. Even if morality does make obstinance more common, it certainly doesn't make it inevitable. (Not every moraliser is so stubborn.) There are conceivable ways in which we could control for it

One might worry that I have conceded too much to the abolitionist in allowing that moral conviction can encourage obstinance. But it's important to appreciate just how small of a concession this is. The abolitionist's central claim is that our moral practices do more harm than good. From this, she quickly infers (far too quickly, I think) that we ought to do away with these practices. But there is another option here. We could make some efforts to minimise these harms instead. In doing so, we could preserve the benefits of moral practice while avoiding the potential costs.

It therefore seems to me that the real issue here is not whether morality has associated costs, (it surely does), but whether these are costs for which we could control. In the ensuing discussion, I will suggest that there are in fact a number of promising means by which we can minimise abuses of the moral overlay.

§3.3 THE ARGUMENT FROM HISTORY

Abolitionists are particularly fond of drawing attention to morality's bad track record. Ian Hinckfuss points towards

> … the massacre of the moral Catholic highlanders by the moral Protestants at Culloden and its aftermath, the genocide of the peaceful and hospitable stone-age Tasmanians by people from moral Britain, the mutual slaughter of all those dutiful men on the Somme and on the Russian front in World War I, the morally sanctioned slaughter in World War II… all this among people the great majority of whom wanted above all to be good and who did not want to be bad. (1987, §2.2)

Stephen Ingram pays special attention to the role that moral considerations have played in helping along social oppression:

> If your group is in the business of subjugating some other group, one effective way to help sustain that subjugation is to convince everyone that your group is more competent at moral judgement. …Plausibly, such methods

> have been used throughout history to help sustain oppressive social hierarchies. (2015, pp.238-9)

These are the sorts of claims that characterise *The Argument from History* (AFH). The purpose of drawing our attention to these historical samplings is to motivate the idea that our moral practices are on-balance harmful. Moral considerations are pliable, and so, they can be used to further harmful agendas as well as desirable ones. Social conflict and oppression are bad enough quite on their own; any mechanism that helps them along is surely (the thought goes) something we can do without.

Unfortunately, abolitionists are seldom explicit regarding the finer details here. It isn't clear, for example, who is included in the scope of this 'we'; just whose interests are frustrated by these moralised agendas? Oppression and war are not certainly contrary to *everyone's* interests; profiteers and warmongers, for instance, often stand to gain. I will assume in what follows that the 'we' here is intended to apply to *most* of us, who presumably want to avoid an untimely demise, see to it that others are well, and live in a stable and co-operative society. Not everyone has these ends, of course. But we need not require that abolitionism be sound advice for everyone. It need only be sound advice for most of us, who share a broad variety of interests and concerns.[107]

We have specified to whom abolitionism is addressed. But other questions remain. It's not obvious what *sort of role* abolitionists take morality to have played in these samples from our history. Addressing this question will be helpful for the purposes of distinguishing the more plausible varieties of the AFH from those that can be dispensed with rather quickly.

[107] I will also assume that the 'we' here is restricted to people of the present day. The AFH doesn't necessarily appeal to human history in its entirety; abolitionists rarely discuss (what might be called) the distant past. They need not think that moral belief-systems were on-balance harmful for early humans. Indeed, it has been suggested that morality may very well have been on-balance useful in early human societies, but that it is more of a hindrance than a help to achieving our goals in the modern world (Fraser 2017).

§3.3.1 *The role of morality*

I should like my critical discussion of the abolitionist's arguments to be a discussion of her strongest arguments. So I will now proceed to single out what I take to be the most plausible variant of the AFH. Abolitionists argue that moral considerations have played an important role in their depressing sample of historical events. But there are a number of different roles that moral considerations may be thought to have played. One possibility is the following:

> **Strong role**
> Moral considerations were counterfactually responsible for these unfortunate events.

On this reading of the AFH, the unfortunate events in question would not have come to pass had it not been for the moral justifications that were offered in support of them. Certain remarks from abolitionists suggest that they intend to put forward this strong claim. Joshua Greene (2002) seems to think that nations would not be capable of garnering the necessary support for aggressive foreign policies if they did not have moral language at their disposal:

> One might go so far as to say that nations *require* the language of moral realism to marshal popular support for aggressive actions. Has a military aggressor ever not claimed a moral right to carry out its plans? Has a nation ever been moved to war by leaders who said, "It would be good for us economically, and we can get away with it, so why not?" (2002, p.238, emphasis in original)

I think that we can safely dispense with this strong variant of the AFH; for our answer to Greene's question ought to be a resounding *yes*. I take it that Genghis Khan did not have to tread carefully around his marauders' moral sensibilities, ensuring that they felt morally justified in riding off to rape and pillage.[108] Desires for glory and conquest appear to have been sufficiently strong motivators. Of course, identifying the full range of causal factors involved in any particular historical episode is difficult. But identifying unwelcome social agendas that succeed without morality is not. Wall Street profiteers do not seem to require any moral justification to advance their own

[108] I thank Ben Fraser for the example.

financial interests at others' expense. *Pace* Greene, 'we can get away with it, so why not?' can sometimes suffice.

Let us therefore set to the side the proposal that moral considerations were needed for war efforts (etc.) to gain a foothold. This is not to deny that moral considerations had any role to play. Perhaps their contribution is better characterised as follows:

> **Moderate role**
> Moral considerations were causally sufficient for these unfortunate events.

One finds this suggestion in Hinckfuss (1987). In maintaining that moral societies (i.e., those that participate in the institution of morality) are "elitist, authoritarian" and "inefficient in the resolution of conflicts", he does not intend to suggest that they would not be this way were they not *moral* societies. His claim is that that "the way morality perpetuates itself within a society is *causally sufficient* for the perpetuation and aggravation of these aspects of society" (1987, §2.2, emphasis added).

As I understand Hinckfuss's suggestion here, it is perfectly possible that an *amoral* society could be elitist and authoritarian; morality is certainly not necessary for things to go awry. This is plausible. But the flipside of the sufficiency claim is not. Hinckfuss seems to think that introducing moral practice into a peaceful and well-functioning society would be enough to send it on the path to rack and ruin. And that seems false. Just how morality manifests itself within a society will presumably depend upon the nature of that society; its members, its social organisation, and the like. I defer further development of this idea to §3.3.3. If what I have to say there is right, then the moral overlay is certainly not enough to render a society conflict-ridden, authoritarian, and elitist.

But all is not lost for the AFH. Even if morality was neither necessary nor sufficient for these atrocities, it is still possible that things would not have been nearly *as bad* were it not for moral considerations. The abolitionist may accord the following, weaker role to moral considerations:

> **Weak role**
> Moral considerations made things worse than they otherwise would have been.

Many abolitionists seem to have this weaker role in mind. Ingram (2015, p.238) concedes that moral considerations do not themselves generate social hierarchies, but argues that they help to perpetuate them. Richard Garner does not think that moral conviction is what leads us to war. But he does think that moral considerations have often made things worse, since they "...can be used to justify inflicting any cruelty deemed necessary for victory" (2007, p.507). I take this weak variant of the AFH to be the most promising. It is not my intention here to attack straw dummies. So I will assume in what follows that it is this argument that must be reckoned with.

§3.3.2 *Other culprits*

According to abolitionists, moral considerations have played an important role in the massacres and oppressive social structures of history, making matters worse than they otherwise would have been. I should acknowledge from the outset that this claim is not easy to assess. We can speculate, of course. But it is incredibly difficult to state with any great confidence how history would look without talk of moral rights, duties, and obligations. In any event, it is certainly not something that we ought to be confident of coming to know merely as a result of armchair speculation.

Let me be clear: I am no historian. My strategy will not be to tirelessly tease apart every potentially relevant causal factor at play, finally arriving at a principled conjecture as to whether things would have been just as grim were it not for morality. Instead, I shall simply grant to the abolitionist that moral considerations exacerbated the atrocities that she invites us to consider.

Even granting this, I think there is something to be gained from turning our attention to the many non-moral factors as play. As we shall see (§3.3.3), doing so helps us to appreciate that the harms associated with moral practice are not *inevitable*, and to identify some promising avenues for minimising the misuse of morality. I will also suggest that these non-moral factors may very well have been sufficient for war efforts and subordination to gain a foothold. If I am right, then we should be especially thankful for the *positive* role that moral factors have played in helping us to counteract undesirable social agendas.

Religion and Intolerance

We can now devote some space to examining the non-moral factors that have plausibly helped along war and oppressive social structures. Needless to say, religion has often been an important contributing factor. The Crusades in Jerusalem were driven by a Christian objective to reclaim the Holy Land from Muslims; the Thirty Years

War was spurred by Protestants' refusal to comply with Ferdinand II's attempt to impose Catholicism upon them; and The French Wars of Religion were, at least in great part, the result of Catholics' intolerance towards Huguenots.

As far as social subordination is concerned, the role of religion in facilitating oppression can hardly be overstated. The inferiority of women is enshrined in religious scripture—as is the validity of slavery. Religious teachings can also legitimise the status quo in the minds of both oppressor and oppressed. Kevin Bales notes that for many slaves in Mauritania (which uses Sharia as its legal system),

> … freedom is a dismal prospect. Deeply believing that God wants and expects them to be loyal to their masters, they reject freedom as wrong, even traitorous. To struggle for liberty, in their view, is to upset God's natural order and puts one's very soul at risk. (1999, p.108)

I do not want to pretend here that morality is easily dissociable from religion. Religious teachings do, after all, prescribe and prohibit certain kinds of behaviour, and these directives are seldom free of moral language. Indeed, it is sometimes customary to interpret people of faith as (tacit) champions of a distinct kind of meta-ethical position, according to which moral obligations have their source in God's will (see Anscombe 1958). Religious wars and oppression may very well have been fuelled by a sense of moral duty.

That said, we shouldn't be so quick to infer that a sense of moral duty is what's doing the heavy lifting in these cases. Religious belief-systems tend to come prepackaged with threats to the non-compliant; a smite from above, for example—or an eternity of damnation, perhaps. And a fear of divine reprisal has motivational force. Whatever one's considered moral judgment on the matter of war and oppression, the threat of fire and brimstone can surely suffice to motivate supporting the religious agendas of the day. The promise of avoiding divine reprisal was all the more tangible for those participating in The Crusades, who were promised absolution from their sins.

Religious conviction also has the potential to breed intolerance. The religiously affiliated have been found to be more intolerant of ethnic minorities than the unaffiliated (Allport & Kramer 1946, Hall et. al 2010). Regular churchgoers also tend to be intolerant of nonconformists (Stouffer 1955). Religious fundamentalism in particular seems highly correlated with prejudicial attitudes (McFarland 1989, Altemeyer &

Hunsberger 1992, Kirkpatrick 1993, Hunsberger 1996).[109] And intolerance can certainly help along social oppression. The intolerant feel no need to refrain from exercising their power to interfere with others.[110] Even if religions do prohibit and prescribe certain kinds of behaviour, then, it is far from obvious that a sense of moral duty is the dominant force at play in the cases to which the AFH appeals. Many of these atrocities were likely helped along by prejudicial attitudes as well.

Harmful ideologies

The considerations above notwithstanding, we do need to be wary of using religion for target practice. The two world wars were not carried out in the name of faith. Nor has oppression always been rooted in religious belief-systems, for that matter; other sorts of ideologies can also help to sustain social hierarchies.

That ideologies can breed contempt comes as no surprise—Nazism is a commonly cited example. Nazi ideology combined an especially toxic nationalism with an emphasis upon racial purity. Added into the mix was a deep-seated anti-Semitism; many of the country's social and economic woes were (wrongly) put down to the scheming interference of German Jewry. Once again, the boundaries here can be murky; it is not unlikely that Nazi ideology may have been moralised along the way. Nonetheless, it seems implausible to claim that Nazism wouldn't have gained *any* traction had it not been for moral considerations. The Nazis abhorred Jews, and regarded them as a threat to their nation's prosperity. It is not obvious that they wouldn't have done them any harm if they hadn't taken it to be their moral obligation to do so. Many facets of Nazism can plausibly be taken to have represented people's non-moral values and preferences: values attached to racial purity, ambitions for territorial expansion, and the like.

[109] Though, as Allport and Ross (1967) point out, the positive correlation between religious affiliation and intolerance seems to be stronger for those with an "extrinsic religious orientation" (who use religion to further their other ends) than those with an "intrinsic religious orientation" (who internalise the values of their faith). Batson and Ventis (1982) propose that another orientation which involves searching for answers to existential questions—what they call "quest"—is associated with greater tolerance and sensitivity towards others.

[110] I draw upon Cohen's (2004, p. 69) understanding of intolerance here.

Epistemic complacency

The list above should not be surprising; religion, intolerance, and ideological factors are the usual suspects. But there is a further suggestion that I want to develop here. It seems to me that there is one explanatory factor in particular that we find in almost all of these cases. Aside from religious fervour, wayward ideologies, and intolerance, many historical atrocities seem to have been marked by a staggering sort of *epistemic complacency*; an utter failure to carefully reason through the relevant issues, or to question and challenge the empirical beliefs of the day. Many of the beliefs that these people held—beliefs regarding the legitimacy of slavery, the inferiority of women, or the contribution of the Jewish people to Germany's loss in the first world war—were, by all appearances, simply taken for granted.[111]

It is my contention that epistemic complacency has played a substantial role in these samples from our past. This is, of course, an empirical hypothesis—one that would require further research and reflection before it could be pronounced with any greater confidence. But the hypothesis seems to me to be eminently plausible; it is difficult to maintain that these undesirable social agendas would have been just as successful had the parties involved been epistemically *vigilant*. Showing that Nazism would have gained the traction that it did even if theories of racial purity and the stab-in-the-back-myth had been subjected to further scrutiny seems like a tall order.

The epistemic complacency hypothesis certainly seems to have something going for it. However, let me say a little more to motivate it. It is important to appreciate that a culture's (dominant) moral belief-systems are seldom—if ever—divorced from its stock of non-moral beliefs. So-called 'caste societies' attach moral significance to hierarchy and social order. But their moral systems have long been intertwined with mystical beliefs about purity and pollution (Stevenson 1954, Haidt 2012). For much of history, moral justifications were offered for the enslavement of people of colour.

[111] This is not to suggest that such matters were taken for granted *by all*. Coalitions of willing dissenters have certainly not been lost in the annals of history, and I shall have more to say about them in §3.3.3. What is important to appreciate is that these coalitions seldom seem to have constituted the majority of those involved. As Konrad Adenauer once observed, it is unlikely that Nazism could have gained the momentum that it did had it not "… found, in broad strata of the population, soil prepared for its sowing of poison…Broad strata of the people, of the peasants, middle classes, workers and intellectuals did not have the right intellectual attitude".

But again, we cannot ignore the influence of the non-moral beliefs held by the subordinators. They regarded the inferiority of their slaves as scientific fact; studies from phrenology, for instance, suggested to them that certain races were more 'advanced' than others (Hanlon 2003). The moral justifications that were offered in support of oppression may have made matters worse. Yet it's difficult to shake the impression that people's inaccurate *non-moral* beliefs were the primary source of harm.[112] If they hadn't held these false beliefs, then it is far less likely that they would have been in a position to offer a moral justification for their oppressive practices.

None of this is to suppose that there was anything *epistemically special* about those involved. As Gideon Rosen points out when discussing the sexism of the 1950s, a failure to "…see through a pervasive and well-protected ideology need not be a sign of culpable negligence or recklessness…It might just be a sign of ordinariness" (2003, pp.67-8). Epistemic complacency isn't merely a quality of the epistemically challenged; few wonder through life reflecting upon or questioning the presuppositions of everyday thought.

But ordinary or not, epistemic complacency very much seems to have played a key role in the cases under consideration. Members of oppressive groups don't tend to be very open to the possibility that they might be mistaken about the legitimacy of their status; that women might have had equal intellectual potential to men is a consideration that seems to have given pause to few who lived before the 20th century. But if that's right, then the real root of the problem would appear to be bad reasoning, and false empirical beliefs. If the abolitionist is concerned to prevent war and oppression, then, it seems to me that she needs to be casting a much wider net; she ought to be taking issue with human stupidity as well.

A lack of human feeling

Though I regard epistemic complacency as an important contributing factor in many of the abolitionist's examples, it would be naïve to think that cognitive shortcomings were the only culprit. Perhaps Khan's marauders could have been more critical in

[112] To be clear, I am not here endorsing the claim that *all* moral disagreements come down to disagreements over non-moral matters. I am only endorsing the more modest claim that people's moral beliefs plausibly covary to some degree with their non-moral beliefs.

thinking about whether they ought to have looted some nearby town. But I take it that this wouldn't have been of much help. We may fault Khan and his posse for their lack of human feeling. But it is difficult to fault them with shoddy reasoning. It is unlikely that things would have been any better had Khan's followers been paragons of epistemic rationality (indeed, they may have been far *worse*).

It is therefore not merely epistemic complacency, but a lack of human feeling that can help along undesirable agendas. I intend to refer to something quite broad here; everything from shortcomings in empathy to unbridled selfishness and acute hatred. It is remarkable just how little thought Khan and his marauders seem to have given to the suffering of their victims. But, of course, it is not only them who exhibited such insensitivity. A strong willingness to be self-serving seems to have been a significant contributory factor in many of the abolitionist's examples. It is not in the least bit surprising that the institution of slavery was favoured by those who stood to gain economically, nor that the elite had a penchant for social stratification.

My tentative hypothesis, then, is that epistemic complacency and a lack of human feeling (especially when working in concert) were important and underappreciated contributory factors in the sorts of cases that the abolitionist brings to our attention. This is not to suggest that all warmongering and oppression can be put down to human stupidity and selfishness. But we should be careful not to underestimate the damage that ignorance and insensitivity can do.

Moving forward

The non-moral factors singled out for mention here are not exhaustive. But they will suffice for my purposes. When put together, these factors seem to form a large part of the explanation for the atrocities to which the abolitionist appeals. It is not implausible that they sometimes would have been sufficient. If anything, our dark past would seem to be overdetermined. The case certainly isn't knock-down; it's difficult to tease apart the many causal factors at play here—be they moral, religious, or otherwise. Nonetheless, we do seem to have good grounds for doubting that moral considerations were the only—or the even most important—culprits.

But just what is to be inferred from all of this? At this stage, not much. I have only suggested that many non-moral factors plausibly operate in tandem with the moral overlay in the cases that abolitionists invite us to consider. I have not denied that our moral practices play some role in generating harm. That said, emphasising the contribution of non-moral factors is important in our argumentative context. Doing so helps us to challenge the assumption that the relevant harms are unavoidable.

§3.3.3 *Minimising the misuse of morality*

I have suggested that abolitionists tend to significantly downplay the role that non-moral factors have played in our dark past.[113] But I have nowhere denied that moral considerations have played a role in helping along wars and oppressive social structures. I am certainly willing to grant to abolitionists that morality has something to answer for here. Perhaps moral justifications weren't necessary for women's subordination. But it seems difficult to deny that they helped.

My arguments would therefore seem to leave the abolitionist in a very comfortable position. It may very well be true that war and systematic oppression have largely resulted from a lack of human feeling, bad reasoning, and false empirical beliefs. But those who engage in faulty reasoning and disseminate false information do walk among us. So long as they do, it might be imprudent to hand them any tools that would serve to make the consequences of their behaviour even worse.

What is important to appreciate at this stage, however, is that the abolitionist doesn't think that morality can only be used for bad purposes. She concedes that morality can be and has been used for good purposes as well. What the AFH really shows, then, is that morality can be of great benefit or harm, depending upon the manner in which it is used. But it doesn't follow (at least not straightforwardly) from this that we ought to do away with morality. The real question is whether or not we can reap the relevant benefits while avoiding the associated costs.

If we could not reap the benefits of moral practice while avoiding the costs (costs of the kind that feature in the abolitionist's historical examples), then the AFH may lend considerable support to abolitionism. Morality might be said to carry far too much baggage; we could not hope to enjoy the relevant benefits without opening the door to (or worsening) oppression and war. But suppose that we could reap the benefits of moral practice while avoiding (or, at least, substantially minimising) the costs.

[113] One exception is Garner (1994), who emphasises the role of religion as well.

If this were so, then abolitionists wouldn't so much have motivated abolitionism as they would have motivated the need for reform.[114]

In what follows, I will argue not only that that there are tangible benefits of moral practice, but that we can conceivably reap these benefits while controlling for the costs. Simply put, there are means by which we can *minimise* the misuse of morality. I won't, however, rely upon the abolitionist's concession that morality can do some good to make my case. Instead, I shall introduce a new player into our dialectic—*the moralist*. The moralist thinks that our error-ridden moral practices are useful to us on-balance. And she has an AFH of her own.

According to the moralist, our moral systems have often been instrumental in overthrowing oppressive regimes, and have often helped us to put an end to war and violence. In support of this claim, she draws our attention to morality's *good* track record: the role of moral values in the eradication of slavery, the importance of women's rights discourse in their liberation from domestic servitude, and the like. The moralist does not pretend that moral considerations can only be used in service of these desirable ends; she acknowledges that they can be put to harmful use as well. Yet she insists that we have ample evidence that moral considerations can be put to very good use indeed. And though the bad here may be rather bad, the good is especially helpful.

Moreover, and as the moralist is keen to emphasise, moral considerations offer us a particularly effective means of *counteracting* harmful uses of morality. Moral conviction might reinforce oppressive social structures. But it is also of great help in overturning them. As Caroline West notes:

> Ideas such as that women have a moral *right* to be treated with equal concern and respect, that current unequal social arrangements are *unjust*, that sexual discrimination is *wrong*, that men *ought* not be differentially advantaged, and so on, function as a check on the behavior of the powerful, protecting the comparatively powerless from suffering further at their hands.

[114] Lenman anticipates this line of reply when he writes that "…vile things are done in the name of moral ideals. But that is not a good objection to morality any more than the existence of bad music is a good reason to dislike music … Rather it is an objection to bad morals and to the stupid, twisted and pathological forms that moral motivation, like any kind of motivation, can sometimes take. It may sometimes favour reform but it hardly favours abolition" (2013, p.397). See also Joyce (p.181).

(2010, p.192, emphasis in original)

Thus, morality doesn't just get us into these unhappy situations—it's also often what gets us out. Slavery in the US might have been maintained by twisted moral values, but the abolitionist movement was driven by moral conviction as well. On first appearances, this might not seem like much of a defence; for we presumably could have done without the years of systematic oppression in between. But recall the suggestion advanced in §3.3.2. If we are right in thinking that much of our dark past was overdetermined—that religious differences or intolerance would, in many cases, have sufficed—then we should count ourselves lucky that there were folk around whose moral convictions moved them to challenge the status quo.

Of course, the moralist must say something more about the role that she takes moral considerations to have played. It is not completely implausible that moral conviction may have been necessary for the eradication of slavery and the enfranchisement of women. But a weaker claim will serve her purposes here. She need only maintain that moral considerations played a key supporting role; that they amplified prosocial tendencies, and added momentum to campaigns for positive social change. Morality may not always be necessary for social progress. But it can often make progress easier to achieve.[115]

Yet why is morality in particular useful for such purposes? The moralist's answer is that moral demands have a distinct kind of practical import. When people judge that women have a *moral right* to be treated with equal concern, or consider their society *unjust*, they take themselves to have a reason to work against these oppressive social structures. And they take themselves to have such a reason independently of whether these arrangements happen to be to their benefit. Moral requirements are invested with *categorical authority*; they present themselves as inescapable demands. One cannot evade their force by citing an immediate interest in non-compliance.

Of course, the moralist is an error theorist, and so, regards the authority of morality as mere illusion; at the end of the day, there are no categorical reasons. What she

[115] As should be clear, the moralist does not claim that moral conviction is the *only* driver of positive social change. On this issue, see Sterelny (2012), who mounts a forceful attack against Kitcher's (2011) proposed history of social progress, which puts moral cognition at the centre. As he notes, we shouldn't discount the role of prosocial emotions, nor that of various circumstantial factors.

wishes to draw attention to is the instrumental value of a conceptual framework that presupposes their existence. When people conceive of oppressive social structures as the sorts of things that must be opposed independently of their ends, their motivation to work against them is likely to be stronger.[116]

Admittedly, there is still the question as to whether morality will be capable of playing this role once it is seen for the farce that it is. This will depend upon whether a fictional, revised, or conserved moral discourse will be capable of recouping the benefits of its error-ridden predecessor. But absent any reason for thinking that none of these proposals are capable of delivering the goods, the moralist thinks that we have good reason to favour them over abolitionism.

Having made her case, the moralist concludes that it's not all doom and gloom when it comes to our moral past. She concedes that the abolitionist has identified a significant problem. But she regards that problem as surmountable; for there are ways to *minimise* the misuse of morality. At this stage, the moralist draws upon the considerations raised in §3.2.2. There, it was claimed that epistemic complacency and a lack of human feeling have played an important role in many unfortunate incidents of history. Bad reasoning, selfishness, and false empirical beliefs, though not the only culprits, have certainly helped along harmful ideologies and wayward social policies, allowing them to go about unchallenged.[117]

The moralist takes these considerations to point towards a promising means by which we might minimise the misuse of morality. To begin with, we might cultivate people's dispositions to seek further evidence for their beliefs, to challenge existing dogmas, to question the ideologies to which they are exposed, and to carefully reason through the arguments in favour of competing social policies. In short, we ought to furnish ordinary folk with a philosophical toolkit of sorts—one that nurtures and encourages epistemic vigilance.

It's worth clarifying what this epistemic vigilance entails. It is certainly not the moralist's contention that oppression and violence can be curtailed by way of fur-

[116] See Joyce (2001, 2006), for a number of convincing arguments in support of this claim.

[117] These were not the only factors singled out for mention. But it is not my business here to establish that we ought to do away with religion. That issue has received its fair share of attention already, so I shall restrict my focus to a more novel suggestion.

nishing ordinary folk with the resources to acquire true *moral* beliefs. (She is, recall, a moral error theorist.) Her foremost ambition is to ensure that morality is not put to harmful use. And she thinks that we might be able to circumvent this by preventing false *empirical* beliefs—beliefs in the inherent inferiority of other groups, say—from gaining a foothold. By encouraging people to challenge and to reflect upon the information that comes their way, we can hope to stop at least many harmful agendas in their tracks.

The moralist acknowledges that epistemic vigilance is somewhat demanding. As things currently stand, it certainly doesn't appear to be a standard at which ordinary agents aim. Following Gideon Rosen, we might suspect that a more plausible standard is one that places us "… under no obligation to rethink the uncontroversial normative principles that form the framework for social life" (2003, p.65). But if we want to enjoy the benefits of moral practice while minimising its misuse, then more stringent epistemic standards may very well be required.

It should be noted that the case for epistemic vigilance doesn't *only* rest upon our interest in securing these practical benefits. We have good independent reasons to challenge what we take to be uncontroversial. To borrow a phrase from Peter Godfrey-Smith (1998), true beliefs are "fuels for success". Playing fast and loose with our evidence is rarely a winning strategy. Most of us care about acting in a way that furthers our ends, and we're generally better positioned to further those ends—whatever they may be—if we track truth effectively (Kornblith 1993, p.371; 2001, pp.158-9). Importantly, none of this suggests that *obvious beliefs* are fuels for success; for what strikes us as obvious may very well be false. We are not epistemically infallible, and so, we often have good reason to subject the obvious to further scrutiny if we want to discover the truth.[118] The justification for epistemic vigilance is thus purely hypothetical; it is conditional upon our interest in securing the benefits of our moral practices while avoiding their potential misuse, together with our general interest in acquiring a stock of true beliefs—interests which, I take it, the majority of us do hold.

[118] I borrow here from Mill's justification for freedom of expression (1859/1977).

The moralist therefore concedes to the abolitionist that morality can be put to bad use. But she holds that the problem is surmountable. Though morality is pliable, we can, through encouraging epistemic vigilance, prevent moral considerations from becoming attached to harmful agendas—such agendas are less likely to gain a foothold if they have been subjected to further scrutiny. When our empirical beliefs fall in line with the facts—facts regarding the equal intellectual potential of other groups, say—we are less likely to be a position to offer a moral justification for warmongering and oppression.

But we are not done just yet. Though epistemic vigilance will be of considerable help, it is unlikely to put an end to all of our troubles. As was noted earlier, a few weeks in the epistemology classroom is unlikely to be an effective antidote to rape and pillaging. This becomes especially evident once we distinguish the participants from the promoters of conflict (e.g., the poor folk dying in the mud from those giving the orders). It may be in the interests of those in charge to encourage going to war (say). And they may be effective at manipulating others to do so (via skilful rhetoric or misinformation, perhaps) even with such epistemic safeguards in place.

The problem is that we have restricted our attention to people's *beliefs* (or belief-forming methods). We have neglected to consider their *desires*; in particular, the degree to which they care for others. So long as there are cruel and inconsiderate folk around, undesirable social agendas may gain a foothold. And when combined with a willingness to be self-serving, moral considerations may very well make matters worse.

This suggests that better standards of evidence and critical thinking will need to be supplemented with particular emotional dispositions. We might, for instance, work to cultivate our capacities for empathy. In coming to empathise more with others, we would be in a better position to understand and identify with their needs. This would by no means guarantee that we would quickly transition to a peaceful and loving society. But it could certainly encourage us to take others' needs into account more often when deciding what to do.[119]

[119] The idea that empathy (or, at least, many of the empathetic emotions) encourages prosocial behaviour has decent empirical backing. For a review of the evidence, see Eisenberg (2014).

I hasten to add that we should not expect this strategy to be foolproof. I certainly do not mean to suggest that morality would never be put to harmful use if most of us grew to be more empathetic and epistemically responsible.[120] But we do not require a foolproof strategy to justify preserving our moral practices (in some form or other); it is not necessary to establish that these practices have no costs. We need only provide grounds for thinking that the cost-benefit analysis favours their preservation. So long as our moral practices confer distinctive benefits, and the relevant costs can be minimised (to a suitable degree), we would seem to do better to hold onto them.

It is admittedly difficult to know just how costly *enacting these changes* would be. But I have suggested that we have good independent reasons to improve ourselves in various respects—indeed, this is a point on which the abolitionist and I agree. Abolitionists think that we must work to promote mutual trust (Hinckfuss 1987, §4.5), cultivate tolerance (Greene 2002), and enhance empathy (Garner 2007, p.501) if we are to make do without morality. Either way, then, such changes will have to be made. But the costlier option is surely the one that combines these changes with the mammoth task of eradicating our moral practices, and deprives us of their many benefits in turn.

In conclusion, then, I think that abolitionists have plausibly established that morality can support war as well as peace, breed fanaticism as well as fellow feeling, and drive violence as well as positive social change. But they have not thereby established abolitionism. These arguments merely suggest that morality is something to be used with greater caution—not that it is something that shouldn't be used at all.

§3.4 A FEASIBILITY WORRY

When we object to an ideal on the grounds that it is infeasible, we are effectively objecting that it is not a state of affairs that we could plausibly bring about. This isn't necessarily to say that the relevant end would be *impossible* to achieve (Gilabert &

[120] Indeed, empathy may not be quite enough to promote other-regarding attitudes towards *everyone*; for the empathetic emotions are vulnerable to a number of well-known biases (Hoffman 2015, p.81 & p.94; Ugazio, Majdanžić & Lamm 2015, pp.169–170). This is precisely why I think that empathy would function best when supplemented with moral judgments (along with reasonable empirical beliefs).

Lawford-Smith 2012). In requiring that a recommended course of action be feasible, we are, I take it, requiring that there be a reasonable probability of success, conditional upon trying (Brennan & Southwood 2007). Presumably, we should want a solution to our WNQ to be *feasible*. The abolitionist's proposal would be terribly unhelpful were an amoral society something we couldn't plausibly bring about.

Yet it is not clear that the abolitionist's proposal *is* feasible. It is commonly recognised that purging ourselves of morality would be incredibly difficult. William Lycan suspects that "to produce a genuine freedom from moral intuitions…" one would need "…a steady diet of hard drugs, or some other very powerful alienating force" (1988, p.211, fn.10). Peter Singer goes so far as to say that we would "…find it impossible to prevent ourselves inwardly classifying actions as right or wrong" (2011, p.xv). Others have likened giving up moral talk to ceasing to speak of people having beliefs or desires; "…possible, perhaps, but not an easy thing to do" (Nolan Restall, and West 2005, p.311).

These claims seem to me to be well-founded. Presumably, doing away with our moral practices wouldn't be akin to putting an end to just any old habit—it wouldn't for instance, be like quitting smoking, or cutting off contact with an ex-lover. Moral thinking is deeply entrenched in our ordinary patterns of deliberation and evaluation. Indeed, we often have very little control over whether or not we assess the world in moral terms; it is quite difficult *not* to think that child labour is morally bad when reading about corruption in the global clothing market.

Moreover, many aspects of our environments prime us to think in moral terms—from stories conveying the rewards of virtue to laws that are justified partly on moral grounds. As Sterelny puts it, "the narrative life of a community—the stock of stories, songs, myths and tales to which children are exposed—is full of information about the actions to be admired and to be deplored" (2010, p.289). Given this, we should expect that purging ourselves of any moral thoughts will be incredibly difficult.

A common abolitionist rejoinder here is to point out that though eradicating widely held false beliefs can be difficult, doing so is not without precedent. Garner notes that we are better off for having abandoned beliefs in geocentrism, even though achieving this was somewhat difficult (2007, p.504). Yet to what extent are such cases analogous to the moral case? They seem utterly disanalogous to me. To the extent that getting rid of geocentrism was difficult, I suspect that it was difficult for different reasons. (The fact that heliocentrism was thought to conflict with religious scripture, for instance, seems to have been one significant obstacle.) Generally speaking, scientists are capable of revising their beliefs without undergoing too much hardship. (In chapter 4, I will draw attention to just how often they do so.) As Nolan, Restall, and West note, eradicating our moral practices is not plausibly akin to eradicating "…a relatively isolated and rarefied concept in a scientific theory" (2005, p.311). Consigning talk of lumineforous ether to the scrap heap may have taken some time. But it was surely not infeasible; there was no need to make sweeping changes to our cultures, or to override deeply entrenched patterns of thought. Making moral judgments, by contrast, is something that laymen and moral philosophers alike do almost every day. And our environments are richly structured in a way that primes us to think in moral terms.

An abolitionist might object that these considerations only speak against her proposal when it is addressed to fully developed moralising agents (i.e., us); the same problems would not arise for an abolitionist proposal that was intended as advice for *future* generations. To some extent, this is true; a child raised in an amoral environment would not face the monumental task of ridding herself of any moral thoughts. But this response still seems to underestimate the large-scale changes that would have to be made. For one thing, *we* would still have to refrain from using moral language while raising children. This, I submit, would be very difficult—and perhaps even unwise in the absence of a promising alternative strategy for encouraging prosocial behaviour in young children.[121] Raising amoral children would also require us to make significant changes to our environments. This is not a task to be taken lightly; everything from our popular culture to our legal systems is saturated with moral influence. It's not clear what abolitionists would have us do with these moral artefacts. (Though images from Bradbury's *Fahrenheit 451* do spring to mind.)

Either way, then, the eradication of our moral practices is likely to require large-scale changes to our current ways of life. Thus, even if abolitionists *were* correct in

[121] Indeed, some abolitionists support continuing to instil moral concepts in children for this reason. Greene argues that morality may be helpful for raising prosocial children, and permits the use of moral language "with those too young to handle the meta-ethical truth" (2002, p.266; see also pp.258-9).

thinking that our moral practices are on-balance harmful, doing away with them altogether might not be an option that is open to us. A more feasible remedy would be to change ourselves and/or our moral systems for the better—something which, I have suggested, we have good independent reasons to do.

§3.5 WHERE DOES MORAL ABOLITIONISM GO WRONG?

As should be clear, I don't think that abolitionism is the right response to the WNQ for moral discourse. For one thing, the proposal doesn't seem to present us with a feasible option going forward. It also strikes me as under-motivated. Abolitionists might have shown us that our moral practices incur some costs. But that doesn't supply us with a sufficiently good reason to do away with them. There are ways in which we could control for these costs, while continuing to enjoy the benefits that these practices afford us.

That said, abolitionism may very well be the right response to WNQs in other discursive domains. The preceding discussion has been instructive in suggesting to us the conditions under which it is likely to make for an appropriate policy. Abolitionism is likely to be a fitting response to a WNQ when:

1. It is feasible.
2. Preserving the discourse is likely to be on-balance costly.

The feasibility condition does not require much in the way of elaboration; it is arguably an important desideratum for any practical proposal (at least if it is to be useful). Regarding (2), I have claimed that *preserving* the discourse must (or be likely to) be on-balance costly. Notice that this is distinct from the claim that *currently employing* the discourse must be on-balance costly. It may be possible for a discourse that is currently on-balance costly to become on-balance useful—if we make some changes to how the discourse is used, say. If the costs of enacting these changes are not too great, then preserving a hitherto on-balance costly discourse may be better (all-things-considered, and in the long-run) than doing away with it. Appreciating this possibility was especially important in the preceding discussion. Even if the moral abolitionist were correct in thinking that morality does more harm than good (and here there is certainly room for doubt), she does little if anything to show us that moral practice cannot be modified so as to do more good than harm. Nor does she provide any reason to think that enacting these changes would be infeasible or excessively costly.

Even if both conditions are not plausibly met in the moral case, they do seem to be met in others. Consider the example of *phlogiston*, a substance that was thought to be released during combustion. Scientists of the 16th and 17th centuries appealed to phlogiston to explain various phenomena. But in the 18th century, Lavoisier showed that such phenomena were instead (or better) explained by appealing to *oxygen,* a substance that was *consumed* during combustion. Perhaps ceasing to appeal to phlogiston was slightly difficult, and it may have taken some time; but doing so doesn't seem to have been infeasible. There might also have been some practical costs associated with abolishing phlogiston-talk. But none of these were sufficient to justify continuing to speak of something that by all accounts wasn't there. It arguably would have been worse for scientists in the long-run to have continued working with a faulty paradigm.

I have not established in this chapter that morality does more good than harm; as was noted at the outset, that is a difficult matter to adjudicate—especially from the armchair. But I have argued that even if it did, we could plausibly control for the associated costs. In the remainder of this work, then, I will assume that our best option going forward will be to preserve our moral practices in some form. Our WNQ therefore becomes a policy question: what is the likely to be the best means of doing so?

Chapter Four

Revisionism

This chapter evaluates the revisionist option. In response to moral error theory, the revisionist suggests that we modify moral discourse such that our engagement in moral practice no longer commits us to the existence of categorical reasons. One means of doing so would be to change the way that we *use* moral language—by becoming expressivists, say. Another would involve changing the *conceptual commitments* of moral language, replacing our error-ridden moral discourse with an error-free, schmoral one. (The latter strategy will take centre stage in what follows.) The revisionist assures us that schmorality will deliver very similar practical goods to morality. And since schmoral discourse is no moral discourse, she promises to rid us of our false moral beliefs as well.

Getting clear on the recommendation to revise the conceptual commitments of moral discourse requires getting clear on what exactly we should take concepts to be. And so, the chapter begins with a discussion of concepts and conceptual change (§4.2). Clarifying these matters also helps us to distinguish our revisionist from a certain kind of success theorist: the reformist. As I shall explain when building a case for the revisionist's proposal (§4.3), there are some important connections between these two projects.

The critical discussion advances a number of arguments against revisionism, all of which are intended to show that schmorality is likely to be a rather poor stand-in for morality (§4.4). I begin by voicing some initial suspicions regarding the revisionist's proposed schmoralities. These seem unlikely to be very useful to us. Following that, I suggest that there is a principled explanation as to why any candidate schmorality is likely to fall short of giving us what we want. The explanation that I shall develop is premised upon some important disanalogies between moral and scientific concepts. These disanalogies are significant in our argumentative context; for a common reformist move—one that is also available to the revisionist—consists in enlisting partners in innocence. Reformists are keen to point out that conceptual change and reform are commonplace in science. Far from being a hindrance, changes to scientific concepts are often changes for the better; for they help scientists to achieve their distinctive goals. On this basis, it is often suggested that modified *moral* concepts are also likely to play a similarly valuable role in our lives, even after having undergone some renovations. This move becomes far less plausible once we appreciate substantial *disanalogies* between moral and scientific concepts. In my view, the latter are far more amenable to (fruitful) modification than the former, and this is owing to the distinctive functions of scientific discourse. I conclude the chapter by drawing attention to the implications of my arguments for revisionist projects in other discursive domains (§4.5).

As should be obvious, this chapter raises a host of complicated issues. Prior to proceeding, then, let me include a few qualifications for the reader to keep in mind.

§4.1 HOW I'M GOING TO GO ABOUT THINGS

A proper assessment of revisionism requires precisifying the proposal. To this end, we must take a stand on some thorny philosophical questions regarding the nature of concepts. There is a slight difficulty here. On the one hand, paying attention to the relevant subtleties and surrounding issues is needed if we are to give revisionism a fair hearing. However, too much attention and this chapter would quickly transform into a defence of a systematic account of concepts and conceptual change—a Herculean task that I have no intention of undertaking.

At times, then, I will need to gloss over some important controversies. There are three in particular that may sound some alarm bells, the first of which concerns the view of concepts that I shall be assuming. On the account for which I have the most sympathy, concepts are identified with *A-intensions*—functions from possible worlds to extensions. (I shall of course do quite a bit of work to spell out what this means in §4.2.) Some may worry that this account does not enjoy such widespread support outside the capital of Australia. It is of course no objection to a position that it is distinctively Canberran. Nonetheless, it may be objected that this position is not the orthodoxy.[122] To appease this worry, I will point towards some benefits of this Canberran way of seeing things, and anticipate some potential concerns. That said, the reader should not anticipate anything approaching a wholesale defence.

Secondly, I will be applying this broadly Canberran approach to scientific concepts as well as moral ones. Unfortunately, discussions of scientific concepts are often slightly disconnected from scientific practice; such debates are often found in the philosophy of language rather than the philosophy of science. Some may worry that these discussions are radically—and to their detriment—divorced from the contexts in which scientific discourse takes place. (As do Brigandt (2013, p.72), and Stotz et. al (2014, p.648).) I take this concern to be well-founded, and so, I will not allow the discussion to be wholly disconnected from scientific practice. Still, I expect that it will be more disconnected than a philosopher of science should like.

Thirdly, I will mostly operate upon the assumption of scientific realism. I have no intention of defending scientific realism here. (Though at least in this case, orthodoxy *would* seem to be on my side.) But it is not as though the *entire* discussion hangs upon realist assumptions. I will point out where I think others (e.g., a constructive empiricist) could happily agree.

To be clear: none of this is to suggest that I will be making assumptions that I do not think can be defended. (I think that what I have to say is right!) It is only to suggest that I will be picking my battles carefully, so as not to take us too far afield.

§4.2 CONCEPTS: WHAT THEY ARE AND HOW THEY CHANGE

I begin this first portion of the chapter by saying a little more about what I take concepts to be (§4.2.1). I will then proceed to discuss the ways in which they can change (§4.2.2), before adding some caveats and qualifications (§4.2.3). Having equipped myself with this philosophical toolkit, I put it to good use, distinguishing our revisionist from the reformist (§4.2.4).

[122] It may also be objected that I am an Australian who is quick to assume an Australian position. But I take it that insofar as this is a problem, it is not merely a problem for *me*. G.A. Cohen (2000) once observed that both himself and his contemporaries at Oxford had sympathy for the analytic–synthetic distinction, whereas Harvard graduates did not. See also White (2010).

§4.2.1 *Concepts*

There are different ways to approach the question as to what concepts are. It is perhaps unsurprising that we see these differences across disciplines, which differ in their explanatory and investigative aims. But even within philosophy, it is not easy to identify anything close to a standard way of approaching the question. As Margolis & Laurence (2014) observe, "disputes about concepts often reflect deeply opposing approaches to the study of the mind, to language, and even to philosophy itself". Given this inter- and intra-disciplinary diversity, it will be helpful to distinguish a distinctively philosophical understanding of concepts from the understanding most commonly adopted by psychologists, before introducing the specific philosophical account that we shall be working with.

Concepts, as I will be understanding them, can be initially characterised as the constituents of propositions. We can call this the *philosopher's* understanding of a concept. I will make it more precise in a moment, but first it is worth distinguishing it from (what might reasonably be called) the *psychologist's* understanding of a concept. On the latter understanding, a concept is an *explanans* in cognitive science that is invoked to explain certain kinds of higher cognitive processes—among them, categorisation and inductive reasoning (Machery 2009). It is natural to expect that there are important connections between philosophers' concepts and psychologists' concepts (see Weiskopf 2010, Lalumera 2014). But given widely recognised differences between the two, I do not assume as much here.[123] I will restrict myself to the philosophical understanding in what follows.

There are many ways in which we might try to precisify the characterisation of concepts as the constituents of propositions—at least as many ways as there are theories of propositions more generally. Much of what I have to say will be neutral with respect to the latter, but it will be helpful to have a concrete approach with which to frame the discussion. Hence, on the approach that I favour—that of Jackson (1998a)—concepts can be identified with so-called *A-intensions*. (Although this partic-

[123] Among these widely recognised differences are: (i) philosophers, unlike psychologists, are concerned with the semantics of concepts (Machery 2009), (ii) philosophers, but not psychologists, take concepts to determine their extensions (Margolis & Laurence 2007), and (iii) psychologists adopt a broader understanding of conceptual capacities than philosophers (Machery 2009).

ular way of spelling out the details is Jackson's, the view that I am about to sketch (broadly construed) does have an impressive fan base, including Lewis (1981, 1994) Chalmers (1996), Pettit (2003), Braddon-Mitchell (2003), and Smith (2004).)

An A-intension in general is a function from (centred) possible worlds to extensions (in this case, either individuals or classes of individuals).[124] We can distinguish between the A-intensions of singular terms and general terms. The A-intension of a singular term always picks out a unique individual at a world, if anything at all. For example, the A-intension of 'Hume' singles out whomever Hume is at a given world (if anyone is). For general terms, the A-intension singles out a class of individuals for each world. For example, the A-intension of 'philosopher' picks out the class of all philosophers at each world considered as actual.[125]

According to Jackson, the A-intension that is associated with an expression is determined by the *categorisational dispositions* of competent users of that expression—specifically, by how they are disposed to respond when presented with hypothetical cases (1998a, pp.31-42). Jackson takes competent speakers to have a kind of discriminating capacity: they can determine the extension of an expression in any fully-described hypothetical scenario.[126] If, for example, I am a competent speaker of English, and I possess the concept <water>, then I will be capable of picking out the extension of the term 'water' in any scenario that you present to me. More specifically, and in Jackson's jargon, I will be capable of picking out the extension of 'water' at any world *under the supposition that that world is actual*. Were you to invite me to consider the hypothesis that a world just like our own is actual (i.e., to provide me with a

[124] The subtleties relating to centred (or *de se*) content will not play a role in what follows, so to simplify the discussion I will generally just speak of possible worlds rather than centred possible worlds. I have also neglected to say anything about Jackson's C-intensions, and how they relate to his A-intensions. Within the pluralist framework that two-dimensionalists such as Jackson operate, it is natural to identify concepts with the epistemologically primary A-intensions: they come closer to representing the cognitive significance that attaches to a concept. See also Chalmers (2002a).

[125] For sentences, the A-intension spits out a truth-value; the extension of 'Hume is a fantastic philosopher', for example, is 'true' at all worlds where Hume is a fantastic philosopher—i.e., all words where Hume exists! My discussion will be restricted to singular and general terms.

[126] These claims come attached with some important caveats, of course. For one thing, we're supposing idealisations of the judger such that they can receive and comprehend a(n infinite number of) fully described possible world(s). For another, we're supposing that the hypothetical scenario isn't described in such a way as to make the judgment trivial.

complete description of it) then I would be able to tell you, ‘if things are actually thus-and-so, then water is H_2O’.

The underlying idea here is that there must be *something* that guides our responses to possible cases.[127] When someone asks us whether there is any water on Putnam’s Twin Earth (considered as a hypothesis about how the actual world is), we don’t just deliver a random guess. Our verdict is shaped by our (often tacit) assumptions as to which factors are relevant in determining whether or not something counts as water (Jackson 1998a, pp.29-42). We might like to think of these tacit assumptions as a kind of “implicit semantic rule” or “*internal reference-fixing template*…that guides [our] verdicts no matter what the actual world turns out to be like” (Schroeter 2012, emphasis in original).

One concern with this approach is that different speakers might have different linguistic dispositions, and thus, different concepts. If that is so, then successful communication quickly becomes a difficulty. However, we have good grounds for taking members of a linguistic community to have incredibly similar linguistic dispositions to one another.[128] If widespread similarities in our responses to thought experiments are any indication, then we do seem to associate the same (if not remarkably similar) intensions with expressions like ‘water’ (Jackson 1998a, pp.38-9). Moreover, the hypothesis that speakers share these dispositions provides us with a powerful explanation as to why linguistic communication is typically so successful (Jackson 1998b).

One might also worry that the approach quickly runs into epistemological troubles. Jackson understands our concepts in terms of our linguistic dispositions. But we clearly lack the resources needed to present real agents with *all* scenarios that might be relevant for determining what their linguistic dispositions are. Can we ever be sure, then, that we have gotten a proper handle on our concepts? Perhaps not. But we can at least have a reasonably good stab at it. In particular, we can consider—and likewise invite others to consider—a range of possible cases, and take note of the

[127] This is not necessarily to revert to the psychologist’s understanding of a concept; we need not be committed to any claims regarding the underlying psychological processes and mechanisms.

[128] To be sure, we can allow for *some* variation—so long as it is not typically so great as to pose a threat to successful communication (Jackson 1998b, pp.214–5).

regularities (if any) that emerge from our investigation. Should we identify any such regularities, then we can hope to have gained some insight into which features are relevant for determining the extension of a candidate expression.

Of course, we shouldn't always expect our investigation to yield a neat and tidy description, or a list of necessary and sufficient conditions. The features in question may often be highly disjunctive, complex, and difficult to state explicitly (Jackson 1998b, p.212; see also Chalmers 2012, ch.1). That being said, neat and tidy descriptions can still be helpful for offering an approximate characterisation of our concepts (Chalmers 2008, p.593). For example, and given what we know about subjects' responses to possible cases, we might say that the intension of 'water' is something close to "the watery stuff our acquaintance in [a] world" (Jackson 1998a, p.49). In putting forward this description, we offer a rough characterisation of the concept <water>—one that captures (as best we can) speakers' dispositional patterns of response to possible cases.

In summary, then, we can say that to grasp the concept <water> is to grasp (at least tacitly) the criteria which determine, for any possible scenario, what the extension of 'water' is in that scenario. Since these criteria are identified by examining our linguistic dispositions, our justification for choosing some characterisations of the concept <water> (e.g., 'the watery stuff') over others will appeal to our responses to relevant possible cases.

I have done my best to ward off some initial concerns with the approach that I will be assuming. But let me briefly rehearse some considerations in its favour as well. To begin with, Jackson's account vindicates the importance of conceptual analysis and appeals to intuition—the bread and butter of philosophy (Jackson 1998a, Bealer 1998, Plunkett 2011, Margolis & Laurence 2014).[129] Since concepts are something that can be investigated from the armchair, the account also sits well with the well-domesticated idea that philosophy is an *a priori* discipline (Jackson 1998a). Moreover, Jackson's view vindicates the importance of *the philosopher* in constructing an account of the world, suggesting a division of labour between her and the scientist. The philosopher is needed to "define the subject" (Jackson 1994)—to tell us

[129] This is not to suggest that conceptual analysis is only useful for determining what our concepts are. See Nolan (2009), who discusses a number of other benefits.

what (our concept of) beliefs, say, *is*—and hence to tell us whether empirical findings are findings having to do with beliefs or something else. The scientist's task is to deliver these findings—to ascertain what the actual world is like (Chalmers & Jackson 2001). Though the extent to which these advantages are *advantages* is something that is likely to differ with one's philosophical tastes, I hope they will suffice to convince the reader that the approach I have chosen is well-motivated.

§4.2.2 *Conceptual change*

Let me demonstrate how I will be understanding *conceptual change* by way of example. Consider the concept <Santa>, which many children acquire early in life. Their parents tell them that there is a jolly fat man living in the North Pole whose elves make presents that are distributed to nice children at Christmas time—naughty children receive coal. What might it mean to say that we could change these children's concept <Santa>? Three possibilities come to mind.

The first is *renaming*. We could simply change the meaning of the word 'Santa'. That is to say, we could change the intension associated with this term. We might say to these children,

> *From now on, the word 'Santa' refers to a miserable, old man who lives in Maine and hardly ever gives presents to anyone.*

Suppose that we held *everything else fixed* about these children. Would there be any significant change to their behaviour? Probably not—not unless we were merely concerned with their verbal behaviour. Their linguistic beliefs may have changed, but their other beliefs need not have; for all that we have said, they still believe that there is a jolly fat man living in the North Pole, who goes by another name.

We can call the second possibility *reforming*. This involves changing the children's beliefs about what Santa is like. We tell them,

> *Just about everything that you believe about Santa is mistaken. Santa is not a jolly fat man who lives in the North Pole. Talk of 'Santa' originated with an old man, 'Montgomery Santa', who lives in Maine and used to give presents to children at Christmas time, but no longer does. So when you talk about Santa, you're actually talking about Montgomery, who never gives presents to anyone.*

These children have now been told that the person whom their Santa thoughts are about has different properties—we have changed their *beliefs* about Santa. (We have not merely changed their linguistic beliefs regarding what refers to what.) Presumably, this will produce a change in their behaviour (and not merely their verbal behaviour). Perhaps they will no longer go to great efforts to be nice if there is no jolly fat man who only bestows gifts upon those who are well-behaved. Or perhaps they will start asking some questions about the origins of their presents.

We can dub the final possibility *revising*. Here, we change the children's beliefs regarding Santa's very existence, but offer a substitute to play a similar role. We say to them,

> *Santa does not exist. There is no jolly fat man who lives in the North Pole. This has always been a complete and utter lie. Your parents are the real Santas; it is they who buy you presents at Christmas time.*

Here again, we have changed these children's beliefs, and we would expect a change in behaviour to follow suit. They would not, for example, continue to send their wish lists to the North Pole, but would simply hand them to their parents instead.

Of our three phenomena, renaming seems the least interesting. It amounts to little more than terminological stipulation. I mention it largely to set it to the side. Our main focus in what follows will be reforming and revising. On my understanding, however, only the latter counts as genuine conceptual change. Let me explain why, proceeding, once again, by way example.

Some time ago, we came to appreciate that there is no such thing as species in an Aristotelian, essentialist sense.[130] That is to say, we came to discover that what we called 'species' did not have intrinsic biological traits that occur in all and only their members (Hull 1965, 1978).[131] However, we also came to discover that there was an-

[130] I thank Edward Elliott for the example. I should note that it is controversial whether Aristotle ever subscribed to the essentialist view that is so often attributed to him. See Wilkins (2013) for discussion.

[131] I should acknowledge that in principle, there *could* be traits like this. (We could perhaps ensure as much through genetic engineering.) But this does not, I take it, invalidate all post-Darwinian insights about species. The point is that an essentialist trait such as this is not what makes a species *a species* (i.e., it is not the important unifying feature). I thank Ben Fraser for helpful discussions on this point.

other biological category—genealogical lineages—which overlapped quite a bit with what we thought species were, and validated a lot of the inferences that we were disposed to make about them. Just as essentialists had thought, there were groups of creatures who tended to look rather similar, and produce fertile offspring. And so, we decided to use the term 'species' to refer to something slightly different—roughly, continuous genealogical entities on the tree of life.[132] Though there were no Aristotelian species, there was something that came damn near close. And it seems that it was close enough.

Has the concept <species> undergone genuine conceptual change? I myself am inclined to answer in the negative. Admittedly, this verdict is premised upon a particular way of understanding theoretical terms—one that has affinities (both historical and philosophical) with the Jacksonian view of concepts outlined above. But the verdict is principled and well-supported. I will now proceed to explain why, leaning heavily upon the work of Lewis (1970) and Braddon-Mitchell (2005) while doing so.

According to the approach for which I have the most sympathy, the meaning of a term like 'species' is intimately bound up with the postulates and presuppositions of the broader scientific theory in which it is embedded. Theoretical terms, as they are often called, get their meaning in virtue of the role that they play within the context of a theory (Ramsey 1931, Lewis 1970).

However, and as many have argued, we should not understand the meaning of a term like 'species' (merely) in terms of the *original* theory T in which it first appears. There is an important sense in which the meaning of 'species' is (partly) *deferential*. The meaning of scientific terms such as these embody a deference to future scientific findings and developments.[133] Their meaning is given by the "nearest near-realisation of T" (Lewis 1970, p.446). Braddon-Mitchell offers a helpful schematisation of this idea. Following his lead, we can say that what an initial theory T1 tells us is that 'spe-

[132] I speak here of how *we* chose to use the term species. But of course, matters are not so simple. Scientists are often engaged in different endeavours, and so, some may think of species differently than others. I will continue to speak of *the* species concept, (and in what follows *the* species-role) for the time being, returning to this complication in §4.2.3.

[133] There is, of course, another important sense in which scientific terms are commonly thought to be deferential; language users can defer to current experts on the correct application conditions for a particular scientific term (Putnam 1975a; cf. Burge 1979). I return to this idea in §4.4.2.

cies' refers to whatever is described by a possibly unknown true theory TT that is a development or refinement of T1. More specifically,

> T1 contains a term P1 (['species']) and a clause which associates with P1 whatever properties are associated with the term Pt of some true theory TT that explains the nature of what plays the T1-role actually. (2005, p.160)

Let's apply this approach to <species>. Assuming (as I will) that Braddon-Mitchell's proposal is roughly correct, what the Aristotelian theory T1 told us was that 'species' refers to whatever plays the species-role in another theory, TT, that is a development and refinement of T1. What is crucially important is this: the concept <species> effectively remains *constant* throughout. 'Species' always refers to whatever plays the species-role within a possibly unknown true theory that is a development of our original one.

Initially, this suggestion may come as a surprise. But it is a lot less surprising once we remind ourselves that it is speakers' linguistic dispositions that are of primary importance. Suppose that the essentialists were to have considered (as actual) a world in which scientists had concluded that there are *no* (i) groups of creatures (S_1) with common essences that interbreed and produce fertile offspring, but there *are* (ii) groups of creatures (S_2) that share a lineage and tend to (but do not always) have many traits in common, interbreed, and produce fertile offspring. Call this scenario W. How would we expect the essentialists to have answered the question: 'If W is actual, then is S_2 species?' It is very plausible that they would answer 'yes'; for it is very plausible that ordinary speakers are disposed to defer to scientists on such matters. But if that is correct, then their linguistic dispositions (and thus, their concept <species>) are the same as our own. They simply had false beliefs about species.

As I understand it, then, our species case is an instance of *reforming*. Upon coming to appreciate that the groups of organisms that we called 'species' did not have biological traits that occur in all and only their members, we did not throw our hands up in the air and declare that there were no species. Instead, we changed our beliefs about what species are like. But since we did not change the intension associated with the term, this was not an instance of genuine conceptual change.

Yet there was indeed *a* change when we left the Aristotelian conception behind. If it was not conceptual change, then what was it? I suggest that we understand reforming to involve a change in *conception*—that is, a change in our beliefs about the properties had by the things that the concept picks out.[134] This seems to be the right result. When we let go of Aristotelian species, we didn't cease believing that species were whatever played the species-role in a future scientific theory. What we ceased to believe was that every member of a species had a certain sort of intrinsic essence common among all and only members of that species. Taking the Aristotelians to share our concept of <species> has an additional virtue; doing so allows us to straightforwardly accommodate the intuition that they got things *wrong*. One is inclined to say that the Aristotelians had false views about *species*—not that they were perfectly correct about species*, and merely wrong as far as species is concerned.

In summary, then, we can say that reforming only involves making some changes to our conception(s), whereas revising and renaming are instances of conceptual change; for both involve changing the intension associated with a term. Moreover, both revising and reforming differ from renaming in that they involve a non-trivial change in beliefs—the beliefs that change are not merely our linguistic beliefs about the application conditions of a term.

§4.2.3 *Some qualifications*

Before proceeding, it will be helpful to introduce some further subtleties to this broad brush-strokes picture of conceptual change.[135] Firstly, I have not denied that concepts *ever* undergo change. I have only suggested that we should not be too quick to infer that they have. New discoveries often signal a need for conceptual reform rather than a need for radical changes to our concepts.

Empirical discoveries can signal a need for conceptual refinement as well. There is often some indeterminacy in our concepts, and empirical findings may move us to precisify their application conditions.[136] Consider the example of <rheumatism>. It is

[134] I borrow the concept–conception from John Rawls (1999).

[135] I thank Ben Fraser for pressing upon me the need to do so.

[136] Given such indeterminacy, there may be borderline cases as well. It may very well be indeterminate what 'mass' referred to before Newtonian mechanics was replaced by the theory of relativity, for example (see Field 1973). It is also not obvious what we would say when considering a scenario in which all of the creatures we call 'cats' turn out to be robots controlled by Martians (Putnam 1962). Would this amount to the discovery that there are no cats, or the discovery that cats are really robots? It's not at all clear. In these case and others, our dispositional patterns of response may not be suffi-

possible that its initial extension was indeterminate between applying to (i) the syndrome and (ii) the most common underlying disease (Chalmers and Jackson 2001, pp.344-5). But over time, the concept of <rheumatism> has been refined (it now applies to the syndrome), and its extension has become more determinate as a result.

Secondly, in some cases it may be incredibly difficult to judge whether or not conceptual change has taken place. Consider the concept <gene>. Talk of genes has varied quite substantially alongside empirical developments. Griffiths and Stotz (2007) map this variation by distinguishing "the traditional gene", from the "post-genomic molecular gene" and the "nominal gene". As they note, these different senses of 'gene' do come apart. But whether it follows from this that <gene> has undergone radical conceptual change over time is debatable. As Stotz et. al (2004, fn.2) observe, it is somewhat controversial to speak of different 'gene concepts', as opposed to different conceptions.

The issue is further complicated by the fact that talk of genes also varies among scientists working in *the same* time period. Developments biologists, for instance, (predictably) emphasise the role of genes in developmental pathways, whereas evolutionary biologists tend to concentrate more upon the effects of genes on particular phenotypes (Stotz & Griffiths 2004, p.16). Given these different understandings of 'gene', some may be inclined to say that there has been a proliferation of different gene concepts. However, this is not obvious. The concept 'gene' may very well be disjunctive. This possibility comes close to a suggestion once made by the biologist Thomas Fogle (2000), who suggested that biologists use a "consensus gene" concept; something counts as a gene if it has *enough* of the features of a stereotypical gene—an RNA transcript, a TATA box, and so on. (Stotz & Griffiths (2004, p.16) note that this is consistent with their findings.) None of this is to deny that <gene> *has* undergone conceptual change. (It is a difficult case, and one on which I am hesitant to take a strong stand.) My goal here is merely to acknowledge that deciding these matters is not always straightforward.

Finally, we should distinguish the goals of the folk from those of the scientist. Earlier, I claimed that *we* ceased to use the term 'species' to refer to groups of crea-

ciently determinate to fix the relevant A-intensions, and so, the A-intensions in question may be somewhat fuzzy.

tures who have common essences. Given some plausible assumptions about linguistic deference (recall §4.2.2, and see §4.4.2), I stand by this claim. However, it's worth clarifying that this is consistent with an essentialist *conception* of species being commonplace among non-experts, and sufficient for their goals; distinguishing edible from poisonous plants, and dangerous from benign snakes, for example.

§4.2.4 *Reformists and Revisionists*

I want to conclude this introductory portion of the chapter by distinguishing our revisionist from a particular sort of success theorist: the reformist. (These differences will become important in §4.3.) The revisionist and the reformist part ways on the matter of whether the existence of categorical reasons is a *non-negotiable* commitment of our moral concepts. The revisionist, *qua* error theorist, answers in the affirmative; to rid moral discourse of this commitment would amount to a change of subject. The reformist denies this. A world without categorical reasons may very well prompt us to transform our *conception* of rightness (and the like). But she thinks that our moral *concepts* would remain very much intact.

(Much of) the disagreement between our revisionist and the reformist therefore takes place at the level of conceptual analysis. To be sure, the reformist need not deny that there is *something* right about the error theorist's analysis. Perhaps the *best* deserver of 'moral rightness' would be a property which was such that if an action had that property, there would be a categorical reason to perform that action. But deservers need not be perfect. The reformist may argue that our moral concepts are considerably more flexible and disjunctive than the error theorist assumes. Perhaps the property that she proposes to identify with 'moral rightness' isn't as perfect a deserver of the name—but it may be deserving enough.

Of course, it is sometimes customary to interpret reformists as recommending that we revise our moral concepts. (See for example, Loeb 2008, p.377.) However, I would caution against this understanding (especially given the view of conceptual change that we are here assuming). This interpretation effectively takes reformists to be recommending that we change the subject. As such, it misconstrues their theoretical ambitions. Presumably, that ambition is vindication—not changing what it is that is to be vindicated.

Taking reformists to be recommending only changes to our conceptions would seem to cohere better with what they have to say. Railton, for example, emphasises that a reformist analysis, if it is to be "vindicatory", must retain sufficiently many "essentials" and secure an appropriate degree of fit with ordinary moral discourse—it cannot amount to a change of subject (1989, p.172). Indeed, Brandt's (1979) reforming project is sometimes criticised on the grounds that it *does* change the subject, and so, fails to vindicate *morality* rather than something else (Sturgeon 1982, p.418; see also Velleman 1988; Loeb 2008). These remarks would be strange if it were fair game for the reformist to change our moral concepts.

§4.3 THE CASE FOR REVISIONISM

It is common for reformists to draw optimism from the enterprise of scientific theorising. Reforming seem commonplace in science, and scientific discourses continue to serve their important functions in spite of (sometimes radical) changes to our conceptions of scientific phenomena. This is often thought to justify a promising prognosis for the reformist project in meta-ethics; we have good grounds for thinking that moral discourse will likewise continue to play a sufficiently similar and valuable role in our lives, even after our conceptions of moral phenomena have undergone some changes. (See Lewis 1989, Finlay 2008, Jackson 1998a.)

Though the revisionist is not explicit in drawing optimism from scientific theorising, she shares the reformist's confidence that significant changes to moral discourse would not prevent it from fulfilling its core functions. We might doubt that this is so. But a natural response for the revisionist is to invoke the reformist's argument from precedent, drawing optimism from the enterprise of scientific theorising. Here, I examine this reformist strategy, and explore how the revisionist could enlist it in defence of her own proposal.

§4.3.1 *The reformist strategy*

There is a long tradition of advancing a sort of 'partners in innocence' argument that likens reforming our conception of moral rightness (say) to reforming our conceptions of scientific phenomena. Many reformists who adopt this broad strategy concede to the error theorist that strictly speaking, our moral terms fail to refer. But they hold that some conceptual reform is all that is needed to resolve the issue. The resultant departure from the 'strictly speaking' is often accompanied by a battery of apologies or excuses, and it is here that scientific analogues usually come into play. The reformist reminds us that we are perfectly content to allow our conceptions of scientific phenomena to undergo important changes. And she suggests that matters ought to be no different in the moral case. My goal in what follows will be to put pressure upon this move. But it will be helpful to consider some examples prior to doing so.[137]

Our first case study is Stephen Finlay. Finlay (2008) agrees with the error theorist that there are no absolute moral values (that is, values that are not relative to particular human ends). But he thinks that she is mistaken in believing that there need be any in order for our moral terms to successfully refer. Our belief in the importance of absolute value is misguided; this is simply the wrong conception of moral value. The right conception, according to Finlay, is a relational one; in his view, moral concepts have an inherent relativity—albeit one that "hides in plain sight" (2008, p.361). When we declare that something is morally wrong, what we are really declaring is that it is wrong relative to a particular evaluative standpoint. We fail to recognise this—and find the absolutist conception irresistible—because we very often find ourselves in societies that share a moral code. In these homogeneous environments, the relational character of moral value goes unnoticed.[138]

Finlay's suggestion that ordinary speakers' conceptions mistake a relational phenomenon for an absolute one might sound surprising, and perhaps even uncharitable. But he thinks that a scientific analogue lends plausibility to his case:

> We find analogous tendencies to absolutism in people's assumptions about

[137] I should note that the list to follow is far from exhaustive. Harman (Harman and Thompson 1994, p.4) make a similar point to Finlay, and also draws upon the absolute–relative motion distinctive when doing so. (As does Dreier 2010, p.79.) Railton's reformist project appeals to the example of water, which is a compound according to our current conception, but was once believed to be an element (1989, p.157). Sterelny and Fraser (2013) discuss ancient mariner's astronomical judgments, which were partly correct.

[138] Interestingly, some empirical studies may partially support what Finlay has to say. Sarkissian et. al (2011) found that subjects have more 'objectivist' moral intuitions when assessing individuals from their own culture, and more 'relativistic' moral intuitions when assessing people from different cultures.

> motion... Motion, we now accept, is a relational matter: there can be motion only relative to a frame of reference. But this is something that needed discovery: for most of history, motion has been taken to be absolute. (2008, p.361)

Finlay claims that it would be "preposterous" to take the motion discourse of ancient mariners to have been infected with error (2008, p.362). We can certainly criticise their absolutist *conception* of motion. But since they make more or less the same motion judgments as the rest of us, we should take them to have been tracking the very same thing—what we now know is motion relative to a frame of reference. Finlay thinks that a similar diagnosis ought to be offered in the moral case. The absolutist's conception of moral value is mistaken. But it need not follow that her (or our) moral concepts are riddled with error. Since she makes more or less the same moral judgments as the rest of us, we should take her concept of moral value to be tracking the very same (instantiated) thing—actions that are morally right or wrong relative to a particular set of standards.[139]

Lewis also enlists the analogy with the theory of relativity when defending his view that values are what we are ideally disposed to desire to desire. This proposal carries the unwelcome consequence that value is contingent:

> Our dispositions to value things might have been otherwise than they actually are. We might have been disposed, under ideal conditions, to value seasickness and petty sleaze above all else. Does the dispositional theory imply that had we been thus disposed, those things would have been values? That seems wrong. (1989, p.132)

Lewis considers what seems a promising fix: amending the proposal so that 'value' refers to whatever it is that we are necessarily disposed to desire to desire. Yet this fix leads to an even more unsettling consequence. Though 'value' in this latter sense "would fully deserve the name", there is no such thing. "If a value, strictly speaking, must be something we are necessarily disposed to value, and if our dispositions to value are in fact contingent, then strictly speaking, there are no values" (1989, p.134). Lewis ultimately concedes that strictly-speaking there are no values; for there is noth-

[139] For criticism of Finlay's argument (and some complications and problems that arise from his invocation of the absolute–relative distinction), see Joyce (2011).

ing that perfectly earns the name. However, he maintains that loosely speaking, there *are* values; whatever we are ideally disposed to desire to desire is an imperfect deserver of the term.

It is here that we see the reformist strategy at work. Lewis likens his choice to settle for an imperfect deserver of 'value' to our choice to settle for an imperfect deserver of 'simultaneity'. Strictly speaking, there is no such thing as absolute simultaneity—there is only simultaneity in the frame-dependent sense. But we do not swiftly infer that "Nobody ever whistled while he worked!" Instead, we adopt a "calm and conservative response"; we conclude that simultaneity is not quite what we thought it was. Lewis suggests that the same conclusion is fitting in the moral case: we ought to say that value (like simultaneity) is not quite as we thought (1989, p.137).

Finlay and Lewis appeal to scientific concepts to motivate affording moral concepts a similar treatment. But many reformists simply *assume* that the two are to be handled in the same way. Jackson (1998a) adopts much the same approach for moral discourse as he does for folk psychology.[140] Regarding the latter, he proposes to find a home for the mental states that folk psychology posits (e.g., beliefs, desires) by first gathering a set of subject-determining platitudes. Doing so helps us to systematise what say, beliefs, are according to our ordinary conception. This systematisation provides us with a kind of job-description for belief; it tells us what sorts of things beliefs (if any there be) must be. Jackson does not think that something can stray *too far* from this job description if it is to deserve the name 'belief' (1998a, p.38). But he does think that developments in neuroscience could move us to slightly modify our understanding of what beliefs are like. In collaboration with Phillip Pettit, Jackson hedges his bets on a mature folk psychology that has been informed by future scientific findings (1990, p.50).

Jackson thinks that we should adopt more or less the same approach in the moral case. Again, we begin by gathering a set of subject-determining platitudes about (say) moral rightness, and use these to systematise what 'moral rightness' is according to our ordinary conception. And here again, rightness is whatever plays the rightness-

[140] Braddon-Mitchell also treats the two cases in a similar manner. He develops a disjunctive (or conditional) analysis for both 'qualia' (2003) and 'moral rightness' (2006).

role in a *mature* folk morality—that is, "where folk morality [would] end up after it has been exposed to debate and critical reflection" (1998a, p.133). According to Jackson, then, we ought to go about finding a home for moral properties and mental entities in much the same way; we formulate a job description, find something that comes close to satisfying it, and settle for an imperfect deserver if need be.

The reformist strategy seems to me to be rather popular, though it manifests itself in different ways. Sometimes, scientific analogues are invoked to motivate the plausibility of the claim that our conceptions of moral phenomena can be radically *mistaken* (as in Finlay 2008). Other-times, they are used to lend support to the idea that we should be content to make do with imperfect deservers of our moral terms (as in Lewis 1989). And some theorists seem to think that when hunting for moral properties, objects or relations, we should proceed just as we do in the hunt for mental states (Jackson 1998a). What unites these different manifestations of the reformist strategy is that they amount to treating moral concepts and scientific ones very similarly—if not the same.

It is certainly no coincidence that reformists tend to be naturalists. Many of these theorists are concerned to solve the 'location problem' for ethics; they wish to determine where (if anywhere) moral properties might find a home in the natural world. Within this context, reforming our conceptions is near inevitable; for our beliefs about natural phenomena are very often mistaken. It is also not surprising that many reformists are champions of (or at least have very close ties with) the Canberra Planning approach to solving the location problem; systematising commonsense intuitions, building a job description of the phenomenon to be located, and hoping that something will satisfy it (or come close).

Two things should be noted before proceeding. First, I do not think that reformists are wrong to assume that *some* conceptual commitments of moral discourse are up for grabs. As Hussain (2004, p.160) observes, no complete meta-ethical theory can plausibly be expected to take *no* commitments of our current moral practices to be mistaken or confused. At least some conceptual tweaking seems inevitable. My goal will only be to suggest that we shouldn't overestimate just how much is admissible.

Second, I am in fact incredibly sympathetic to the Canberra Planning approach to solving location problems. (As one might have guessed from §4.2.) My arguments in what follows are by no means intended to show that this approach is fundamentally flawed. The worry to which I will soon give voice is that we might have been slightly too cavalier about treating moral rightness so similarly to how we treat mental states and motion.

§4.3.2 *Grist for the revisionist's mill*

There are different ways in which we might go about revising moral discourse. One option would be to follow the recommendation of 'revolutionary expressivists', who advise us to use moral discourse in an expressive manner if the moral error theory is true (Köhler & Ridge 2013, Svoboda 2017). Another would be to revise the conceptual commitments of moral language such that our talk of rightness and wrongness no longer commits us to the existence of categorical reasons (Lutz 2014).

As I have noted, the revisionist's project differs in important ways from that of the reformist. But reforming definitions can still be of use to the revisionist. Matt Lutz proposes that the revisionist can

> ...adopt this basic move of the reforming-definition theorist—to reorient our moral thought and language toward the salvaged concept...The salvaged concept is different enough from our moral concept that following this suggestion would not constitute a type of realism about morality. But it is similar enough that doing so looks like a promising way to capture enough of what we want... (2014, p.365)

The revisionist can, in other words, take (something close to) what the reformist regards as *analyses* as promising candidates for *schmoralities*. While these analyses were not quite enough to rescue moral discourse from an error theory, they may be enough to secure (many of) the benefits of moral practice. Just as the reformist is confident that a modified conception will get us what we want, so too is the revisionist confident that a modified concept will give us "…very good prospects of living a fulfilling normative life" (Lutz 2014, p.370).

(I should qualify my use of 'schmorality' in this context (and at this stage). I do not necessarily intend for it to refer to something that is not worth caring about; for if the revisionist is correct, then schmorality could be immensely useful to us. By 'schmorality', I intend to refer to something closer to what Joyce has in mind when he characterises a schmorality as "…something bearing a resemblance to a morality—enough, perhaps to be mistaken for the real thing by the inattentive—but which falls short of really being so" (2008, p.65; cf. Dennett 2006).)

We might be sceptical, however, that a *useful* moral practice could survive such radical conceptual change. A change of subject (accompanied by a relevant change in beliefs) seems bound to translate into a change in practice. And there is no guarantee that the resultant practice would be a desirable one. We should want some further grounds for placing our faith in the revisionist's schmorality.

A natural response for the revisionist at this juncture is to enlist the reformist strategy, gesturing towards the success of scientific theorising. Of course, the revisionist would need to point towards examples of conceptual change, rather than changes to our conceptions. I have suggested that the latter are arguably more common, so the revisionist may not be on quite as strong ground as the reformist here. But it is not as though there are no case studies to which she could appeal. By drawing our attention to these precedents, the revisionist could provide us with grounds for thinking that her proposed revisions to our moral concepts will not prevent them from serving their valuable functions. There have (let's assume) been substantial changes to the concept <gene>. But far from preventing us from engaging in a useful form of gene discourse, these modifications have been of considerable *help* in allowing us to make empirically meaningful inferences and generalisations. The same may very well be true of moral discourse; far from being a hindrance to practical success, conceptual change may be our best chance of securing it.

So a natural move for the revisionist is to parasitise not only the reformist's analyses, but her argument from precedent as well. In doing so, she substantiates her claim that a schmorality will be capable of serving the important functions of morality. In what follows, however, I will argue that this hope is misplaced. The revisionist's project is vulnerable to the very same (if not very similar) worries as the reformist project that it is parasitic upon.

§4.4 Against schmorality

I will begin my criticism of revisionism by taking issue with some ostensibly promising candidates for schmoralities (§4.4.1). Here, I will rehearse some (all-too-familiar) concerns about the promises of naturalistic reductions in ethics, and use them to my advantage. Following that, I suggest that these problems are in fact a symptom of a deeper issue: there is a principled explanation as to why any candidate schmorality is unlikely to get us what we want (§§4.4.2-3). Some may take the explanation on offer as further reason to doubt the prospects of reduction in ethics. Less ambitiously, it may provide us with a diagnosis as to why so many find these proposed reductions unsatisfying.

§4.4.1 *Against the letter of schmorality*

If we follow the revisionist's recommendation to reservice reformists' analyses, then our schmorality is very likely to be a naturalist one; we will identify moral properties like rightness (or, I should say, rightness*) with natural properties—increasing pleasure, or being what we are ideally disposed to desire to desire, say. But there are grounds to be wary of naturalist schmoralities such as these, and it is the same grounds that are often thought to count against naturalist analyses. Naturalists take moral properties to be run-of-the-mill natural properties that bear no essential link to normative reasons (Brink 1989, ch.3; Sturgeon 2006, pp.110-12); whether or not we have any reason to act as morality requires ultimately depends upon whether or not we care about the moral good.

Now, the error theorist agrees with the naturalist about the facts on the ground; she does not think that moral requirements *do* supply all agents with normative reasons for action. But she also takes moral discourse to play a valuable role precisely on account of the presupposition that it *does* have these normative credentials. In chapter 2, I appealed to the idea that moral judgments function as *deliberation-stoppers*; they prevent competing considerations that would interfere with prosocial motivation from entering into the deliberative sphere (following Joyce 2006, p.111). When an agent judges that she morally ought to φ, her other desires and ends are sidelined in the decision-making process. A related idea (one advanced by Dennett 1986, p.123) is that public moral judgments function as *conversation-stoppers*; they block any further negotiations from taking place when making interpersonal decisions.

Yet moral judgments plausibly play these valuable roles because we tend to conceive of our moral obligations as things that we ought to do, *period.* And it is precisely this problematic conceptual commitment that the revisionist wishes to purge. Were we to follow the revisionist and revise moral discourse such that it was an *open ques-*

tion whether or not we ought to pursue the moral good—the answer to which *depended upon* our ends—then it not at all clear that these ends will continue to be sidelined or blocked in our decision-making processes.

These worries for the revisionist are reminiscent of a well-known challenge to the reformist. You've heard it all before: reforming analyses seem unable to accommodate the distinct kind of *practical authority* with which we invest moral requirements. They cannot do justice to "the authority of reason" (Hampton 1998, ch.3). They "[drive] what to do out of ethics" (Gibbard 2003, p.13). The underlying and well-trodden idea here is that descriptive properties do not seem capable of performing the action-guiding function that morality must plausibly fulfil if it is to be useful to us. As Hinckfuss puts it,

> Jiminy Cricket is not of much use to his Pinnochio, if he is able to say only after the behaviour in question, that he has perceived it to be good or bad. Moral knowledge is for…preventing sin, not simply describing it. (1987, §2.6)

These old worries for the reformist transfer rather straightforwardly to the revisionist. The old worry is that an analysis that does away with practical authority is *intensionally* inadequate; it fails to do justice to important moral platitudes. The new worry (for the revisionist) is that such analyses are also likely to be *practically* inadequate. In divesting moral discourse of a commitment to categorical reasons, the revisionist divests it of its distinctive rhetorical and motivational force.

(Indeed, one may think that there is a sense in which these two worries are two aspects of a single worry. It seems that our answers to metaphysical questions (e.g., 'what are *x*'s?) sometimes *should* be intimately tied to our answers to practical ones (e.g. what concept of *x* should govern our practices surrounding *x*'s?). As Caroline West (2008, p.58) points out when discussing the concept of personal identity, the answers to such metaphysical questions are at risk of being uninteresting if they merely describe a "theoretical epiphenomenon" that is fundamentally disconnected from our real-world interests and concerns.)

Matters seem especially pressing for response-*dependent* schmoralities, which render moral judgments akin to judgments of taste. Lutz's proposed schmorality, for example, takes rightness* and wrongness* to be determined by an individual's subjective patterns of approval and disapproval. And expressivist schmoralities (e.g., those of Köhler &Ridge 2013; Svoboda 2017) refashion moral judgments such that they come to be expressions of non-cognitive attitudes. Insofar as these proposals forge a connection between our moral commitments and our preferences, they can secure a link between moral judgment and motivation. But they are likely to have trouble securing consensus and co-ordination of the right sort.

As was suggested in §3.2.1, individuals are usually willing to compromise on their preferences. I don't aesthetically approve of wearing red and green together, but I can bring myself do so if my friends want to dress up as Christmas elves come December. Generally speaking, some measure of compromise is taken to be appropriate when our (interpersonal) desires come into conflict (Enoch 2011). If you feel like ordering Vietnamese takeaway, and I had my hopes set on Indian, then it's surely not appropriate for me to insist upon my preference. I should instead try to meet you half way, perhaps by settling for Thai—or by agreeing to do Indian tonight, and Vietnamese next week.

But sacrificing some of our wants spells trouble when it comes to our moral wants. If I think that we morally ought to refrain from genocide, and you think that we ought to go for it, am I now to meet you half way? Perhaps I should agree to wipe out just half of the relevant population—or to go ahead with the purge this time, but insist that we refrain on a future occasion. Yet that seems like the wrong result—that is to say, the *practically* wrong one.[141] We surely do not all benefit from compromising upon undesirable social policies.[142] But if our schmoral judgments are mere preferences, then it seems that this is precisely what we should do. If schmoral facts are merely facts about our subjective patterns of approval and disapproval, then

[141] Compare Enoch (2011, ch.2), who uses these considerations to mount a *moral* case against response-dependent views

[142] Of course, losing half of the population would surely be better than losing it all. It would obviously be practically suboptimal to refrain from moral compromise when standing one's ground would lead to far worse outcomes. My point is merely that it would be bad if the first port of call were always to compromise and meet others half way in such cases even when we didn't have to—if we were ordinarily disposed to respond to the genocide case much as we were disposed to respond to the restaurant case. This isn't to deny that we should be reasonable interlocutors who are open to opposing views—being inclined to stand our ground when we believe we are right is consistent with being willing to hear out the opposing side.

the appropriate form of resolution would be to meet one another half way, as we do in other cases of conflicting tastes.

Though I take the options canvassed here to be the most promising candidates for schmoralities, their sponsors encounter very similar problems to those who put them forward as reforming analyses. Of course, reformists do have responses to these challenges, and the revisionist could very well enlist these in defence of her own proposal. Lutz could, for instance, build some resistance to compromise into the notions of rightness* and wrongness*. And expressivists are experts in philosophical jiu-jitsu, so we should expect that revolutionary expressivists will have more to say here as well. Whether such responses would be satisfactory is, of course, up for debate. But here is not the place to settle whether meta-ethical naturalism and expressivism do justice to moral practice. And picking off schmoralities one by one isn't the most principled way to go about objecting to the revisionist's project in any case. In what follows, I suggest that there is a deeper problem at hand.

§4.4.2 *Moral concepts and scientific concepts: some disanalogies*

The deeper problem in question is premised upon some important disanalogies between moral and scientific concepts. I will consider these in some detail before explaining why they spell trouble for the revisionist's proposal. It should be acknowledged from the outset that the differences to which I will draw attention may turn out to be differences of degree rather than kind. Moreover, I will not infer from these differences that the revisionist's project could not possibly be viable. I will only claim that they supply us with good reasons to doubt its prospects.

The differences are three. To begin with, scientific concepts embody an element of linguistic deference that seems to be absent in moral concepts. There are two kinds of deference at play here. The first is what we can call *horizontal deference*. In his seminal discussion on the topic, Putnam explains the phenomenon in terms of a "linguistic division of labour" (1975a, pp.227-8).[143] Only scientific experts can really

[143] It's worth assuaging a potential concern here. Throughout this chapter, I've put quite a bit of emphasis upon (what may reasonably be thought of as) descriptivist or internalist approaches to understanding meaning and content. So the reader might be suspicious of my relying upon what Putnam

determine what sort of stuff is gold; for only they are sufficiently familiar with the relevant theory. But this does not prevent laymen from speaking meaningfully of gold. Insofar as the folk meaningfully apply the term 'gold', their doing so depends upon a kind of implicit deference to scientific experts; they can defer to these experts on the correct application conditions of the term.

Moral concepts would seem to lack this feature of horizontal deference. When you ask me what I'm talking about when I speak of gold, I can say, 'I'm talking about that expensive yellowish substance that has the underlying structure that has been articulated by experts'. But it is difficult to say anything of the sort when you ask me what I'm talking about when I speak of moral goodness. For one thing, it seems strange to think that there is a unique, underlying microphysical structure at play here. The proposal that there are fundamental moral particles—"morons", as Dworkin (2011) calls them—to which we purport to refer when we talk of rights or duties is, to put it bluntly, laughable.[144]

For another, there don't seem to be any moral experts to whom we could plausibly defer. Moral philosophers or religious leaders might be prime candidates, but they're certainly not an uncontroversially recognised group of experts in the way that, say, physicians or chemists are. No one questions that chemists claim expertise with respect to gold, and that it is appropriate to defer to them on the matter. By contrast, there are widespread suspicions about the appropriateness of *moral* deference. Williams is particularly dismissive

> There are, notoriously, no ethical experts...Anyone who is tempted to take up the idea of there being a theoretical science of ethics should be discouraged by reflecting on what would be involved in taking seriously the idea that there were experts in it. It would imply, for instance that a student who had not followed the professor's reasoning but had understood his moral

(an arch-externalist) has to say. However, the descriptivist-style approach that I have been assuming is sufficiently flexible to incorporate Putnam's insights. Its sponsors can allow that some concepts are deferential; that their extension depends upon the way in which the relevant term is used within one's linguistic community (Chalmers, 2002b, §6; Jackson 2004). We might articulate the general idea by saying that the concept <gold> is captured by a description roughly (very roughly) of the form, 'the actual, valuable, yellowish substance of our acquaintance that is used to make jewellery, and which has the underlying microphysical structure that the experts say it has'.

[144] Finlay expresses a similar sentiment: "…it's absurd to suggest that [moral goodness] might have a complex molecular structure" (2008, p.363).

> conclusion might have some reason, on the strength of his professorial authority, to accept it...These Platonic implications are presumably not accepted by anyone. (1995, p.205)

Following Williams, it is not even obvious that we have reasons to accept *specific* moral verdicts on the basis of expertise. It should be less obvious still that the meaning of our moral terms depends upon the judgments of philosopher kings or some moral elite.

Still, not everyone is as dismissive as Williams. Some have proposed that we can gain knowledge from moral deference, though not *understanding*—the latter being what explains the appearance that there is something untoward about deferring to others on moral matters (Hills 2009). Yet even so, a disanalogy remains; no one thinks that I have to understand how the speed of light is calculated, lest there be something untoward about deferring to scientists that $E=MC^2$.

Others have suggested that even if there are no moral experts *tout court,* people may still claim expertise with respect to a specific moral issue (Hopkins 2007, pp.623-26). But it is one thing to concede that some people may be well-positioned to offer advice on particular moral issues. It is quite another to say that the meaning of our moral terms systematically depends upon what they have to say. That strikes me as a considerable leap. (At the very least, it is a new and interesting philosophical thesis in need of support.) I take the considerations above as decent (though no doubt defeasible) grounds for thinking that moral concepts are *not* plausibly akin to the "technical concepts" that one finds in science, which are partly "fixed by the usage of experts" (Finlay 2008, p.363).

A second sort of deference is what I will call *vertical deference.* This phenomenon was discussed in §4.2.2, wherein I followed others in suggesting that scientific concepts embody a deference to future empirical findings. Are moral concepts also vertically deferential? Some philosophers seem to think so. Jackson, recall, passes the linguistic buck to "mature folk morality". But it is far from clear that deferring to the future is appropriate in the moral case. As Stephen Yablo notes, it seems eminently possible that "… morality could take a direction that we would regard as quite misguided"; "it is no part of current folk morality to defer to whatever comes along" (2002, p.16).

Of course, Jackson thinks that deference is to be paid to future populations who are *reasonable*, and have thought through moral matters *carefully*. 'Rightness' picks out whatever plays the rightness-role in a folk morality that "…has been exposed to debate and critical reflection" (1998a, p.133). But one striking disanalogy between moral and scientific discourse is that the latter is subject to explicit and rigorous scrutiny. Folk morality, by contrast, is implicit at best, and is quite possibly not fully coherent or determinate (Loeb 2008, Sinnott-Armstrong & Wheatley 2012). The caveat that mature folk morality "…has been exposed to debate and critical reflection", renders the analogy with mature scientific theories easier, but at the cost of relevance to actual moral practice. Scientific concepts can survive quite radical modifications while continuing to play their role in theory and practice. But scrutiny and revision is business as usual in science. (I shall return to this point shortly.) It is far less clear that the *moral* folk are as sanguine about scrutinising and revising their theories. One worries that Jackson may be defining away an important difference between scientific theorising and moral practice.[145]

Yet another concern pertains to the sorts of justifications that we can offer for deferring to the future. Scientific enquiry has a good track record. Controversies about replication aside, developments and improvements are more or less routine. So it seems reasonable for us to place our faith in the scientific method as a reliable recipe for progress, and peg the reference of scientific terms to future theories. Some might be optimistic about moral progress as well, of course.[146] But it seems that whatever optimism we have here must be considerably more tempered. The method of careful reasoning is no guarantee that people won't be led morally astray (see Williamson 2001, p.630). Placing our faith in the future appears far more sensible and well-motivated in the scientific case.

(I do not want to pretend that matters are simple or uncontroversial here.[147] It is not out of the question that future science could take a wrong turn—perhaps even

[145] I thank Ben Fraser for helpful discussion on this point.

[146] *Qua* moral error theorists, it is unclear what exactly *we* should take moral progress to be. (Clearly, we cannot take it to consist in our beliefs more closely approximating the moral *truths*.) Perhaps moral progress might be said to involve developing moral (or schmoral or fictional) beliefs that better serve our common interests.

[147] I thank Ben Fraser and Daniel Nolan for pressing me on these points.

for similar corrupting reasons. It is also not obvious that scientific enquiry has a good track record of arriving at *the truth*. After all, we regard many scientific theories of the past as *false*. Perhaps future people will think the same of our own, and so, perhaps we should infer that our current theories are in fact false (Laudan 1981). I cannot hope to mount a systematic defence against such pessimistic meta-induction here. But the inference is certainly questionable. Since the best theories of the past plausibly differ in important ways from the best theories of the present (e.g., the latter enjoy greater success and are better supported by our evidence), an *optimistic* meta-induction does not seem to me to be off the cards. (For arguments for optimistic meta-induction, see Fahrbach 2011, pp.153-4; Park 2011, pp.77-80.))

There is a further respect in which moral concepts seem to differ from scientific concepts. I will call this final distinguishing feature *gross fallibility*. To say that scientific concepts are grossly fallible is to say that we acknowledge that our conceptions of scientific phenomena could be radically mistaken. We acknowledge, for instance, the in principle possibility that a wide range of phenomena could play the germ-role; germs could be harmful micro-organisms (as they plausibly are), miasma, or imperceptible demons who enjoy making life rather unpleasant for various organisms. If any one of these things were to the bill (that is, if further inquiry into the underlying causes of disease were to reveal as much) that then we'd be happy to call them germs.

Moreover, we're perfectly willing to allow that the role-realisers in such cases could surprise us. You might deny that the switch from absolute to relative motion revealed something deeply wrong with the concept <motion>. But you have to admit that there's something strange about asking a fellow passenger when London stops at your train. However—and this is my point—you're willing to shirk that discomfort. Your folk intuitions surely don't permit you to insist that Einstein got things wrong.

This feature of gross fallibility does not seem to be present—at least not to nearly the same extent—in moral concepts. This is not to say that people do not acknowledge (implicitly or explicitly) that some of their moral views could be false. I expect most will readily admit that they could be mistaken in *the content* of (some of) their moral judgments (perhaps some people more readily than others). But it seems far more difficult to admit that one could be radically mistaken about what (broadly speaking) *grounds* one's moral judgments. It is difficult, for example, to be open to the possibility that what grounds the wrongness of child pornography is merely social mores.[148]

Of course, we may acknowledge that our conception of moral goodness (say) could be *somewhat* mistaken. Yet it seems that it could not be *radically* mistaken. My folk intuitions don't permit me to object to Einstein's theory of relativity. But they surely *do* permit me to object to Stevenson (1937, p.14) when he suggests (albeit humorously) that goodness is pink with yellow trimmings. These intuitions also permit us to object—as philosophers routinely have—to more serious naturalist proposals. If the utilitarian's identification of goodness with increasing pleasure fails to comport with important moral platitudes, then, the thought goes, so much the worse for the utilitarian.

Our resistance to such analyses suggests that we do *not* acknowledge the in principle possibility that a wide range of phenomena could play the goodness-role. Nor do we seem all that happy to allow that the role-realisers could surprise us. Even if there were some unique set of natural properties picked out by the term 'good', it would be difficult shirk our discomfort if those properties didn't comport with our moral intuitions to a suitable degree.

(I don't wish to oversell the point. As Nolan points out, there are plausibly *some* platitudes about scientific phenomena that constrain our investigation; "… momentum couldn't have turned out to be a kind of cat, or be identical to the temperature 24°C" (2009, p.286; see also Joyce, p.97). So the difference here may very well be one of degree; a difference in *how* fallible we take ourselves to (potentially) be, and *the extent* to which we take our folk intuitions to license us to have a say in such matters.)

A complementary point here concerns the distinction between vindicating and tracking (Fraser 2014). To vindicate moral discourse, it does not seem sufficient to show that our talk of 'moral goodness' (say) tracks some property cluster *p*. What rather needs to be shown, it seems, is that our talk of goodness tracks some property cluster *p* that is *suitably close to* what competent speakers *mean to attribute* when they call

[148] See Southwood (2011), who suggests that moral judgments are distinguished from conventional ones by the *grounds* that one can offer for them.

something (morally) good. As far as moral discourse is concerned, a tracking explanation is not necessarily a vindicating explanation. Fraser draws an analogy with the term 'witch':

> It may have been—suppose it was—true that medieval witch hunters' judgments about who was and who was not a witch fit a particular pattern: all those judged to be witches were female, old, ugly, and outcasts. So, there was a property cluster shared by all the things judged to have the property of being a witch…. [but this] leaves untouched a further question: did those judged to be witches have the property (cluster) that the witch hunters meant to attribute? To answer that question, it is necessary to know what the witch hunters' notion of a witch was…it is clear that it will involve some supernatural element (practicing magic, or making deals with demons, perhaps). And, the tracking story made no mention of any such element. So, it is possible for a tracking story to be true of some class of judgments without a vindicating story being true of them. (2014, pp.796-7)

As Fraser suggests, the same lesson seems to apply to 'moral goodness'. Even if our talk of moral goodness were to track some natural property *p*, showing that this were so would not suffice to vindicate moral discourse. Competent speakers could acknowledge the availability of such a tracking story while denying that there is any such thing as moral goodness; for the property tracked may not deserve the name (cf. Copp 1990).

By way of contrast, we would expect tracking and vindicating to often *run together* in scientific theorising.[149] As things have turned out, talk of 'germs' reliably tracks facts about harmful micro-organisms. But no one is inclined to object to the epidemiologists that harmful micro-organisms are not properly deserving of the name. We laypersons presumably don't have much ontological say on the underlying causes of disease.

I suspect that the list above does not exhaust the disanalogies between moral and scientific concepts. But it will suffice for my purposes. In what follows, I shall argue

[149] My suggestion is not that tracking and vindicating *always* go together in scientific theorising—only that they often do. Following Joyce (p.4), it would have been odd for the phlogiston theorist to have said "Well, this stuff that Lavoisier is calling 'oxygen' just is what I've been calling 'phlogiston' all along".

these disanalogies put pressure upon the optimistic inference that the revisionist may wish to draw from conceptual change in scientific theorising.

§4.4.3 *Against the spirit of schmorality*

My intention in emphasising a number of disanalogies between scientific concepts and moral concepts has been to garner support for the following claim: the revisionist's (and similarly, the reformist's) import of optimism is unjustified. There is a principled explanation as to why the valuable role of scientific concepts is not thrown into jeopardy by revision or reform, and such an explanation does not seem available in the moral case. The explanation, simply put, is that revision and reform *coheres with* the distinctive function(s) of scientific discourse.

The idea that scientific discourse has 'distinctive functions' needs some spelling out. Let me briefly explain what I intend by this, prior to drawing some important connections with the discussion in §4.4.2. A *discourse*, recall (§I), is a domain of talk and thought that is structured around a particular set of concepts and associated beliefs. Santa discourse, for example, is structured around concepts such as <North Pole>, <elves>, <presents>, and <Christmas>. For children, this discourse also involves a characteristic set of beliefs; the belief that Santa rides a sleigh, the belief that Santa gives presents to nice children, and the belief that he heroically delivers these presents to hundreds of millions of children within the space of a single night.

Santa discourse, I want to propose, has a set of functions that explains why it is valuable to us. On the one hand, Santa discourse has what I will call a *basic linguistic function* (one that it shares with many other descriptive discourses): that of allowing us to denote certain kinds of things, and to assert particular kinds of propositions. But Santa discourse also has a number of what we can call *proper functions*—ones that some discourses have, but others lack.[150] Among these functions is that of social engineering; we introduce Santa discourse to children in order to encourage them to

[150] By 'proper function', I intend for something quite different from what teleosemanticists (e.g., Millikan 1989) discuss. In my usage, 'proper function' is not necessarily tied to evolutionary history, but concerns the distinctive practical purposes of the discourse beyond its basic linguistic functions.

behave well. If they don't, then come Christmas time, their stockings will be filled with coal. (Or so they believe.)

Like Santa discourse, species discourse plausibly has the basic linguistic function of allowing us to denote certain kinds of things, and to assert particular kinds of propositions. But unlike Santa discourse, the proper functions of species discourse aren't best understood in terms of social engineering. We value species discourse because it helps us to achieve distinctively scientific ends; to make certain kinds of empirical inferences and generalisations, for example. Talk of species helps us to pick out and examine things that exist, which have interesting and potentially regular connections to other things that exist. We use such talk, for instance, to make interesting and theoretically well-motivated claims in evolution and ecology.[151]

It's worth emphasising this contrast between Santa discourse and species discourse. The purpose of Santa discourse is not that of helping us to make meaningful empirical claims or discoveries about things that exist. Of course, getting children to *falsely believe* that Santa exists is important for encouraging good behaviour. The point to appreciate is that in order for Santa discourse to serve its proper functions, there is no need for these beliefs to be *true* (or remotely close to the truth). Matters are quite different in the species case. In order for species discourse to serve its proper functions, it is very important that scientist's beliefs and assertions about species are (at least close to) correct. If they are not, then species discourse is unlikely to fulfil these functions. If our theories about species are not even remotely close to the truth, this would hamper our ability to gain a more accurate picture of the world through scientific theorising.

It is here that the discussion in §4.4.2 becomes important. There, I argued that our conceptions of scientific phenomena are always (at least tacitly) regarded as negotiable and fallible; this is suggested by the important kinds of deference that scientific concepts embody, and our readiness to tolerate errors in our understanding.

[151] Once again, I am passing over some heated debates here. Following most scientific realists (e.g., Putnam 1975b, Boyd 1983), I assume that scientific theorising aims to supply true descriptions of the world, and that progress in science has consisted in further approximations to the truth. (Perhaps a constructive empiricist (e.g., van Fraassen 1980) could at least agree with this assumption as far as observable features of the world are concerned.)

What I now want to propose is that these dispositions to defer and tolerate error are closely tied to the *proper functions* of scientific discourses—for example, that of enabling us to make meaningful discoveries about things that exist. If we are to make progress in biology or physics, then we simply *must* afford scientific enquiry a suitable degree of open-endedness; we must acknowledge that species or motion may not turn out to be quite what we thought they were.

It is for this reason that I believe revising and reforming are often appropriate and helpful in scientific theorising; it is because they cohere with the proper functions of scientific discourse. But why does this spell trouble for the revisionist (or the reformist)? Note, to begin with, that the explanation as to why conceptual revision and reform pose no threat to scientific practice does not seem available for *moral* practice; for there does not seem to be any heavy-duty deference or error-toleration built into moral practice. Whereas scientific concepts are structured in such a way as to make room for the possibility of substantial modifications (and for good reason), the same does not seem true of moral concepts.

Indeed, I think that this points towards a further respect in which moral and scientific concepts differ. Given that the proper function of scientific discourse (or, at least, one important proper function of scientific discourse) is to help us to achieve distinctively scientific ends (e.g., to make certain kinds of empirical inferences and generalisations), we are very often willing to *sacrifice* intensional adequacy. We are willing to allow that the right account of species and motion might not cohere perfectly well with our essentialist or absolutist intuitions. Roughly put, we let the world have the final say. This is because *extensional adequacy* is typically of greater importance to us in scientific theorising; we want terms like ‘species’ to succeed in picking out empirically meaningful categories.

But as far as moral discourse is concerned, intensional adequacy seems just as important—if not more—than extensional adequacy. That is to say, having a theory that coheres sufficiently well with deeply held intuitions seems equally (if not more) important as having one that enables moral terms to refer, and moral predicates to be literally satisfied. Reformists will of course deny that this is true (or the whole truth) when it comes to philosophising about and understanding morality. But it certainly seems true as a practical matter—that is to say, as a matter of how well or poorly different revisionist schmoralities are likely to fare. To the extent that a moral system is built on concepts alien to the folk who are to use it, and delivers verdicts at odds with their strong intuitions about the nature of moral phenomena, it seems doomed to be practically irrelevant at best.[152]

To my mind, then, moral discourse is more like Santa discourse than species discourse. Like Santa discourse, the proper function of moral discourse is (at least in great part) *action-guiding*. Moral terms and predicates are not merely valuable to us because they allow us to assert particular kinds of propositions—they are valuable to us because they enable us to shape one another's behaviour. Unlike species discourse, the proper functions of moral discourse are not tied to successfully picking out empirical phenomena. Indeed, there is no need whatsoever for moral terms to successfully refer if moral discourse is to fulfil many of its proper functions. (If the error theorist is correct, then moral discourse *has* long fulfilled many of its proper functions in spite of systematic referential failure.)

The problem for the revisionist is that whether or not moral discourse *can* fulfil its distinctive action-guiding function(s) seems intimately tied to its commitment to categorical reasons. As was argued in §4.4.1, to divest moral discourse of a commitment to categorical reasons is to risk divesting it of its distinctive rhetorical and motivational force. Accordingly, it is highly doubtful that moral discourse could continue to serve its proper functions in the face of such radical changes. And it won't do to appeal to scientific theorising to convince us that it would. Whereas scientific concepts are structured in such a way as to make room for change, the same does not seem true of moral concepts. The proper function of moral discourse is not to pick out empirically interesting phenomena, whatever they may be, but to guide action in characteristic ways. And that proper function is thrown into jeopardy as soon as we propose to conceive of our moral duties as normatively optional, or dependent upon our ends.

§4.5 REVISIONISM: GENERAL LESSONS

The shortcomings of the revisionist's proposal are, I think, owing to her failure to appreciate the sense in which the action-guiding function of morality is intimately

[152] I thank Ben Fraser for helpful discussions on this point.

tied to its problematic conceptual commitments. We cannot simply go about our lives schmoralising and expect to reap the very same (or even sufficiently similar) practical rewards as before. These practical advantages plausibly depend upon conceiving of moral requirements in a particular sort of way, and holding a certain set of (at least tacit) beliefs about the nature of our moral obligations. In purging moral discourse of a commitment to categorical reasons, we rob it of its distinctive practical force.

There are other cases in which the action-guiding function of a discourse seems intimately tied to particular conceptual commitments. Indeed, I suspect that this is true of a number of myths directed at young children. Talk of tooth fairies reduces the unpleasantness associated with losing a tooth. Talk of Santa encourages good behaviour (at least leading up to Christmas time). Clearly, these discourses are shot through with error—suppositions regarding fairies and magical elves are problematic conceptual commitments if any are. But it is difficult to imagine them fulfilling their proper functions (that of shaping children's behaviour) without them. This, I suspect, is precisely *why* we continue to make use of these erroneous discourses. No one feels any need to revise our concepts of <tooth fairy> or <Santa> so as to make them successfully refer—and nor should they.

The suggestion developed in this chapter is that matters are different for scientific discourse. Far from permitting the toleration of outlandish myths, scientific theorising is intended to help us to discover truths about the surrounding world. It would be antithetical to the proper functions of scientific discourse were the fate of its posits to stand or fall with our folk intuitions. And clearly, they do not. We are perfectly prepared to allow that scientific discoveries may surprise us—to defer to the relevant experts, and to tolerate errors in our understanding. More often than not, reforming or revisionist proposals for scientific phenomena—germs, species, and what have you—are unlikely to come into conflict with the proper functions of scientific discourse. So it is here that they stand the greatest chance of success.

CHAPTER FIVE

Moral Fictionalism

Moral fictionalists hold that moral discourse is worth preserving on account of its usefulness. But they do not advise us to preserve moral discourse as it stands. Instead, fictionalists recommend that we preserve moral discourse in the spirit of a useful fiction. In doing so, we can to continue to reap the benefits of moral practice—and we can do so without incurring the epistemic costs associated with preserving our false moral beliefs.

This fictionalist manoeuvre is certainly not unique to the moral domain; it is a common, and increasingly popular response to many (allegedly) error-ridden discourses. The chapter begins with an overview of this rich fictionalist research program (§5.1). This will be helpful for the purposes of understanding the different ways in which fictionalism has been developed as a response to the WNQ for moral discourse. In §5.2, I consider two such developments—those of Richard Joyce (2001, 2005), and Daniel Nolan, Greg Restall and Caroline West (hereafter, NRW) (2005).

In §5.3, I assess a number of challenges that have been directed against moral fictionalism. These challenges do not strike me as devastating. They do suggest that moral fictionalism needs to be further refined. But they should not suggest to us that it ought to be abandoned.

§5.4 marks the end of my friendship with the fictionalist. Here, I shift my critical focus to the attitudes that the fictionalist intends to substitute for our (erroneous) moral beliefs. These attitudes, I will argue, are neither a stable enough nor a strong enough basis for securing the practical benefits of our error-ridden moral practices. I conclude with a diagnosis of where moral fictionalism goes wrong (§5.5).

§5.1 FICTIONALISM: A GENERAL OVERVIEW

I will begin this chapter with a general overview of fictionalism in philosophy. Following a condensed introduction (§5.1.1), I turn to some important distinctions in the current literature. The first concerns which kinds of attitudes we are to adopt and which kinds of speech-acts we are to make when we take on a fictionalist stance towards a particular subject matter (§5.1.2). The second distinction marks a divide between those who think that we *already* use a particular discourse *D* in a fictional spirit, and those who recommend that we *become* fictionalists with respect to *D* (§5.1.3).

§5.1.1 *What is fictionalism?*

Drawing inspiration from Chris Daly (2008, pp.425-6), we can formulate fictionalism about some discourse *D*, schematically, as follows:

> (1) The fictionalist does not believe or assert any of the propositions expressed by the sentences of *D*, but
>
> (2) She believes that partaking in discourse *D* is useful

Regarding (1), the fictionalist takes the sentences of *D* at face value; like other cognitivists, she holds that the relevant language has a representational semantics, and is in the market for truth and falsity (Friend 2008 p.14; Kroon 2011, p.791). However, the fictionalist parts ways from the success theorist—and sides with the error theorist—in maintaining that all of these sentences are strictly-speaking false.[153]

A fictionalist about some discourse *D*, then, does not *believe* any of the propositions expressed by the sentences of *D*—she adopts a different attitude towards them. I will use 'felief' (forgive me) as a placeholder for this characteristic attitude that a fictionalist has towards these propositions, and will consider different candidates for such an attitude in §5.1.2.

An important corollary here is that the fictionalist does not *assert* any of the propositions expressed by the sentences of *D* either. One who sincerely asserts that *p* typically has the communicative intention of conveying that she *believes* that *p* (Daly 2008, p.426). But as far as *D* is concerned, our fictionalist lacks any such intention; for she believes no such thing. She asserts that *p* with the communicative intention of

[153] Following our earlier discussion (§I), we might want to restrict this claim to *the D-sentences of D*. Presumably, the fictionalist about Sherlock Holmes will want to say that the sentence 'there was never really a famous detective named Sherlock Holmes who lived on Baker St' is *true*. See Brocke (2002) for discussion.

conveying that she *felieves* that *p*. I will use 'fassertion' as a placeholder for the distinctive kind of speech-act that a fictionalist makes when she asserts one of the propositions expressed by the sentences of *D*. The task of characterising this speech-act is, again, one that I defer to §5.1.2.

It should be noted that not anything goes as far as these feliefs and fassertions are concerned. As well as characterising the distinctive kinds of attitudes and speech-acts that her approach entails, the fictionalist also needs to explain the appropriateness of particular feliefs and fassertions and the inappropriateness of others (Nolan 2005, p.205). The fictionalist about Sherlock Holmes, for example, must explain why it is appropriate to felieve that Sherlock Holmes is a clever detective, and inappropriate to felieve that he is a timid meter maid. The (in)appropriateness of feliefs and fassertions might be explained by appeal to truth-conditions; feliefs and fassertions might be correct insofar as they are true, and incorrect insofar as they are false. Or they might instead be explained by appeal to assertability conditions. (The latter may be a fitting approach for those fictionalists who deny that feliefs and fassertions are bearers of truth and falsity.)

Regarding (2), the (alleged) usefulness of *D* will vary from discourse to discourse. Mathematical discourse might be strictly speaking false, but talk of numbers might nonetheless be useful in scientific theorising. And though there may be no such things as viruses, quarks, or electrons (suppose), talk of unobservable entities might be useful for the purposes of medical intervention, or progress in physics.

An important task for any fictionalist, then, is to motivate the claim that her candidate discourse can continue to deliver the goods even when it is used as a fiction. To this end, she does well to tell us how her fictionalist discourse is to interact with the *base discourse* (Nolan 2005, p.211; NRW 2005, pp.309-10). The base discourse encompasses our literal commitments, and so, it will not contain any sentences that commit us to the existence of those properties, objects, or relations which we intend to be purely fictional. The base discourse for a moral fictionalist, for example, will be free of sentences like 'x is morally wrong'.

A common strategy for explaining the interaction between the base discourse and the fictional discourse is to devise *bridge-laws*: "statements that connect what is literally true in the base discourse with what is true in the fiction and vice versa" (Nolan 2005, p.211; see also Field 1989).[154] These bridge laws allow us to infer certain literal truths from others, following a detour through the fiction. The bridge laws provided by a fictionalism about unobservables, for instance, could allow us to infer from a patient's pasty complexion and exhaustion (a literal truth) that she has contracted some virus (a fictional truth), and to infer, in turn, that she is likely to experience nausea and dizziness (another literal truth). Some claims that are true in the fiction will, of course, be literally true as well (Nolan 2005, p.212). It is, for example, true in the fiction of Sherlock Holmes that there is a bustling city in Europe called London.

In addition, the fictionalist should want to convince us that the use of these bridge laws will not be contrary to the interests in question (NRW 2005, pp.322-3; Nolan 2005, p.211; 2014, p.218). To demonstrate, consider one potential benefit of fictionalism about unobservables: its associated bridge laws enable us to infer certain literal truths (e.g., the optimal kind of medical intervention) from other literal truths (e.g., a patient experiencing nausea), following a detour through the relevant fiction (according to which the patient has a particular virus). To make the case that her fiction will be useful, then, the fictionalist about unobservables needs to show us that an inference from one literal claim to another that takes a detour through her proposed fiction will not lead us to endorse any falsehoods. The most effective way to achieve this is, perhaps, to provide evidence that the relevant fiction is a *conservative extension* of the base discourse, in which case the use of its associated bridge laws simply could not lead one from literal truths to literal falsehoods (Field 1980).[155] But it is also open to the fictionalist to provide other sorts of grounds for thinking that her proposal will generally be reliable (Nolan 2005, p.212).

[154] Common, but by no means universal. Walton (1990, ch.4) argues that paradigmatic fictional truths are generated in "complex and unsystematic" ways.

[155] Field (1980) famously motivates mathematical fictionalism by arguing that our mathematical theories are a conservative extension of our scientific theories. That is to say, these mathematical theories, when combined with our scientific theories, do not produce empirical consequences that could not have been derived from the scientific theories alone. Field takes these mathematical theories to be false on account of their falsely presuming the existence of numbers. However, he argues that since they (i) help us to more easily make derivations when practising science, and (ii) are conservative, and thus, do not produce any new empirical information, we can justify preserving mathematics in the form of a useful fiction.

A small clarification is needed before proceeding. For the purposes of this thesis, I regard fictionalism as a potential response to the WNQ for an error-ridden discourse. This is not necessarily misleading; error theory and fictionalism is a very popular package. However, some fictionalists are *agnostic*; they refrain from taking any stance on whether the sentences of the relevant discourse are true (e.g., van Fraassen 1980, Dorr & Rosen 2002). Agnosticism may be a sound motivation for fictionalism. But error theorists *cum* fictionalists will be our primary focus here.

§5.1.2 *Prefix fictionalism and preface fictionalism*

There is some disagreement regarding the characteristic attitude that a fictionalist has towards the sentences of *D*, together with the sorts of speech-acts that she makes when she utters one of these sentences. Getting clear on these details will be useful for the purposes of later discussion—specifically, for categorising different moral fictionalist proposals, and understanding the distinct problems that arise for them.

Fictionalists fall into two broad camps on the matter of how best to characterise feliefs and fassertions. According to *prefix fictionalists*, fassertions are assertions and feliefs are beliefs. On this view, an individual who utters some fictional sentence is indeed asserting something. However, she is asserting something other than its literal content through invoking (implicitly or explicitly) some kind of prefix (Eklund 2015). To borrow a well-rehearsed example, the person who utters the sentence

> **SH1** Sherlock Holmes lives on Baker Street

does not thereby assert that Sherlock Holmes lives on Baker St. Instead, she asserts that

> **SH2** According to Arthur Conan Doyle's fiction, Sherlock Holmes lives on Baker Street

Simply put, to *fassert* SH1 is to assert SH2, and to *felieve* SH1 is to believe SH2. Given prefix fictionalism, fassertions and feliefs are apt for truth and falsity; a fassertion or felief that *p* is true just in case according to the relevant fiction, *p*. Prefix fictionalism is a popular approach to discourse about paradigmatically fictional characters like Sherlock Holmes (see Lewis 1978, Brocke 2002). It has also been developed for modal language (Rosen 1990; Brogaard 2006).

Preface fictionalists, by contrast, take fassertions to be pretend-assertions, and feliefs to be make-beliefs. Instead of a prefix, there is instead effectively thought to be a kind of "disowning preface" in place that one that "rob[s] all that comes after of assertoric force"—for instance, 'let's make believe the Holmes stories are true, though they aren't' (Lewis 2005, p.315). The preface fictionalist who utters or thinks to herself 'Sherlock Holmes lives on Baker St', then, neither asserts nor believes anything. She simply pretends to assert and make-believes that Sherlock Holmes lives on Baker St.

Given preface fictionalism, fassertions and feliefs are not in the market for truth. But there can still be (in)appropriateness conditions for these attitudes and speech-acts. A preface moral fictionalist, for example, might specify through her bridge laws that it's appropriate to make-believe or to pretend to assert that something morally bad is taking place just in case something of such-and-such a sort is occurring (the infliction of great harm, say).[156]

Moreover, it is not only assertions (or something like assertions) that can be made in fictional contexts. Other speech-acts—for example, questions and commands—can also be expected to make an appearance.[157] A moral fictionalist would likely want to include questions regarding what is right and wrong and commands to do the right thing in her repertoire of fictional speech acts as well (amongst other things).

§5.1.3 *Hermeneutic fictionalism and revolutionary fictionalism*

Sometimes, fictionalism is intended as a descriptive claim—as an account of how a discourse is used. Other-times, fictionalism is a call for change; it is recommended that we modify our use of some discourse by using it in a fictive spirit. The former has been dubbed *hermeneutic fictionalism*, the latter, *revolutionary fictionalism*.[158]

[156] There are other candidates for feliefs and fassertions that do not fall neatly into either of the two categories discussed here. Some take fassertions to be assertions that that are made under particular presuppositions (Hinckfuss 1993), and feliefs to be attitudes of *acceptance* (van Fraassen 1980, Yablo 2001).

[157] I thank Daniel Nolan for bringing this to my attention.

[158] The hermeneutic–revolutionary distinction originates (at least to my knowledge) with John Burgess (1983). The distinction is sometimes described as one between 'descriptive' and 'prescriptive'

The hermeneutic fictionalist claims that we are already fictionalists about some discourse *D*. This proposal is particularly plausible as an account of paradigmatically fictional discourse. Presumably, we are fictionalists about Sherlock Holmes, Pegasus, and Luke Skywalker. (Few believe that they are relaying historical events when they rehearse the plot of *The Empire Strikes Back*.) A fictionalist approach allows us to explain why people who partake in *Star Wars* discourse are not thereby committed to the existence of Darth Vader, The Millennium Falcon, or Wookiees. In partaking in that discourse, they merely felieve and fassert that there are such things.

As one can likely surmise, hermeneutic fictionalism does not make for a particularly fitting package with an error theory; it is difficult (though perhaps not impossible) to hold ordinary speakers guilty of an error if they aren't committed to the existence of those properties, objects or relations over which their fictional discourse quantifies. Interestingly, hermeneutic fictionalism has been put forward as an account of our existing moral discourse. Mark Kalderon (2005) argues that our attitudes towards moral propositions are closer to attitudes of make-belief than belief. A form of non-cognitivism, this view evades moral error theory by denying that moral judgments are beliefs. (Kalderon motivates his proposal by arguing that moral judgments are not governed by the same epistemic norms as beliefs, and so, cannot be beliefs.)

Revolutionary fictionalism, by contrast, often goes hand in hand with error theory. The revolutionary fictionalist traditionally takes the sentences of her target discourse at face value; she interprets them literally, but holds that when so interpreted, they are systematically false. What makes revolutionary fictionalism *revolutionary* is that it recommends a radical change in the way that a discourse is used. Its proponents advise us to cease believing or asserting the propositions expressed by the sentences of the relevant discourse (on account of their falsity), and recommend that we replace these beliefs and assertions with feliefs and fassertions of some kind (on account of their anticipated usefulness).

varieties of fictionalism (e.g., Ingram 2015, p.232). But being a 'prescriptive' fictionalist seems to me to be consistent with being a 'descriptive' fictionalist; one might think both that a discourse *is* used in a fictive spirit and that we *ought* to continue using it in this way. Alternatively, one might think that it is indeterminate whether or not the discourse is used fictionally, but that it ought to be in any case (see NRW 2005, p.322). The label 'revolutionary' more faithfully signals that fictionalism is a call for *change*. I thank Matthew Hammerton for helpful discussions on this point.

Revolutionary fictionalism will take centre stage in what follows; I mention hermeneutic fictionalism largely to set it to the side. Any subsequent talk of 'moral fictionalists' should therefore be taken to refer only to fictionalists of the revolutionary variety, unless otherwise stated.

§5.2 Moral Fictionalism

I will now consider two varieties of moral fictionalism; those of Joyce (2001, 2005), and Nolan, Restall, and West (2005). These proposals differ along a number of dimensions, including what they take the greatest benefits of employing a moral fiction to be, and the distinct kinds of feliefs and fassertions that they recommend. These details will be important later on when considering the objections that have been brought to bear against each variety.

§5.2.1 *Joyce's moral fictionalism*

The motivations underlying Joyce's moral fictionalism are two-fold. The first is epistemic; moral fictionalism is thought to be attractive in virtue of ridding us of our false moral beliefs (pp.178-9). The second motivation is practical; moral fictionalism is also expected to secure many of the desirable practical goods that we associate with moral practice (pp.206-21).[159]

Among these practical goods is that of "combatting weakness of will" (p.228). Joyce regards moral commitments as especially effective mechanisms of self-control (p.181). Even if a course of action is in our practical interests, and even if we judge it to be so, that is no guarantee that we will pursue it; for we are frequently weak-willed, and fail to act in accordance with our considered judgments (p.211). We might, for instance, succumb to the temptation to go back on our word, or to pinch a few pennies from an unsuspecting friend.

[159] Insofar as having false beliefs is *practically* disadvantageous, the first motivation is to some extent practical as well (see Joyce, pp.175-80). But it does not strike me as unintuitive or misleading to characterise the first motivation as epistemic, inasmuch as holding onto false beliefs will often conflict with our epistemic ends (see §2.1.1).

These lapses may sometimes go unnoticed or unpunished, and so, it may occasionally be in our immediate interests to be unco-operative. Yet being unco-operative is likely to be contrary to our *long-term* interests—one seldom does well in life through coming to earn a reputation as untrustworthy.[160] Generally speaking, it is plausibly in our long-term interests to co-operate with others, and moral commitment is helpful in that regard (p.210). An agent who takes promise-keeping to be *morally required* of her is likely to be someone with a stronger resolve to keep her promises (recall §2.3.3). Thus, morality is useful in functioning as a bulwark against weakness of will; when we enlist moral concepts in our deliberations, we are more likely to act in accordance with our (long-term) practical interests.[161]

It is Joyce's contention that we can continue to enjoy these benefits—without believing moral propositions—by preserving our moral practices as a useful fiction. More specifically, he recommends that we adopt attitudes of "make-believe" (p.197) towards moral propositions, and "pretend to assert" moral sentences (p.291). As I understand it, then, Joyce's proposal is a variety of *preface fictionalism*; life as a fictionalist consists in acts of pretence, rather than believing that such-and-such is the case according to some moral story.

According to Joyce, make-beliefs have a number of important properties which distinguish them from other sorts of fictionalist attitudes. For one thing, make-beliefs involve *thoughts* unaccompanied by belief: when an agent make-believes that *p*, she is thinking the proposition *p* without believing it (p.197). Make-believing that *p* is also said to involve a disposition to withhold assent from *p* in "critical contexts"; contexts wherein we "…investigate and challenge the presuppositions of" ordinary thought (p.192). A moral fictionalist, for example, is disposed to attend to and express her be-

[160] Joyce acknowledges that it might be in an agent's *short-term* interests to (e.g.) break a promise when there is no possibility of being caught or earning an unfavourable reputation. But given the very real danger of miscalculating the risks and the high costs of error, he argues that it is likely to be in an individual's *long-term* interests to adopt a policy of keeping her promises. Though doing so may not benefit her on each and every occasion, it is likely to benefit her in the long run (pp.212-3).

[161] Joyce considers other potential benefits of our moral practices as well (though he does not emphasise them to the same degree). He suggests that thinking in moral terms might also provide "a strong foundation for "moralistic aggression" towards defectors", and that shared moral values may help to build social cohesion (p.228).

lief that nothing is really morally right or wrong when she finds herself in a philosophy classroom.[162]

On the matter of *what* we are to make-believe, Joyce suggests that a moral fiction "need consist primarily of a few general existential claims"—for instance, "there are obligations and prohibitions", and "people have character traits" (p.195). He adds some platitudinous claims (e.g., "torturing babies to pass the time is always wrong") to partially constrain what these obligations and character traits could be. Thus, Joyce's moral fiction is effectively "a conceptual framework"—one that leaves many first-order moral questions open. Two moral fictionalists could not come to agree upon "…whether a second-trimester abortion is permissible … simply by consulting the "story of morality," in the way that two Holmes fans may consult the canonical texts in order to settle a dispute about Watson's war wound" (p.195).

§5.2.2 *Nolan, Restall, and West's moral fictionalism*

NRW (2005) argue that moral fictionalism fares better than rival solutions to the WNQ for moral discourse. Like Joyce, they understand moral fictionalism to be superior to conservationism (i.e., preserving our false moral beliefs) on epistemic grounds (2005, p.314). Insofar as moral fictionalism promises to secure the desirable practical goods that we associate with moral practice, it is also said to be an attractive alternative to abolitionism (2005, p.307).

According to NRW, moral discourse is useful on account of the valuable role that it plays in directing our social lives; in helping us to co-ordinate our attitudes and regulate interpersonal conflict, for example (2005, p.314).[163] With moral considerations in hand, we have an established framework—one of rights, duties, obligations, and the like—with which to navigate our way through practical disputes.

Whereas Joyce recommends a relatively open moral story, NRW seem to have a much richer fiction in mind. They propose that a moral fiction can contain first-

[162] For further discussion, see Cuneo and Christy (2011, pp.87-8), who characterise a critical context as one where there is a strong presumption of truth-telling. They draw a comparison with the court room, wherein one swears to tell the truth, the whole truth, and nothing but the truth.

[163] Following Joyce, NRW acknowledge that moral discourse may also be useful for combatting weakness of will. The difference here is largely a difference in emphasis.

order moral claims, citing act-utilitarianism as an example (2002, p.15, emphasis in original):

> An agent *ought to* do *x* only if no *better* action incompatible with *x* is possible for that agent.

This fiction of act-utilitarianism could be connected to the base discourse via the following bridge law:

> The action *x* is *better than* the action *y* if and only if the action *x* would cause greater overall happiness than *y* would.

The latter would allow us to infer certain literal truths from others, following a detour through the moral fiction. For instance, from the (literal) truth that donating to charity would cause greater overall happiness than purchasing an expensive car, one could infer the (fictional) claim that one morally ought to donate to charity. Assuming that the fictionalist takes herself to have good (non-moral) reasons to act as her fiction prescribes, she could then infer that she (non-morally) ought to donate to charity (a literal truth).

Though they do explore different options, NRW are more or less neutral regarding what sorts of attitudes and speech-acts feliefs and fassertions are to be. This is understandable, given that their foremost intention is to provide a blueprint for would-be fictionalists (and to build a case for the position), rather than a fully developed proposal. For the sake of contrasting different species of moral fictionalism, however, my dialectical choice will be to associate NRW with prefix fictionalism in what follows.

§5.3 The Case against Moral Fictionalism

There are a number of objections that have been raised against moral fictionalism. But it would be intellectually dishonest of me to use all of these to my advantage, since I do not regard all of them as especially worrying. I am inclined to view many such challenges as problems that can be solved with some fancy footwork in the philosophy of language, rather than problems that go to the heart of the fictionalist's proposal.

Of course, it would also be intellectually dishonest of me to ignore these challenges altogether. But since it is not ultimately my intention in this chapter to *defend* moral fictionalism, I am not going to develop detailed responses to each of these challenges on the fictionalist's behalf. Instead, I will explain why we have good reason to regard them as surmountable.

§5.3.1 *An acceptance-transfer problem*

A key selling point of fictionalism is its potential to secure many of the benefits of (our erroneous) moral discourse. Among these is the valuable role that moral discourse plays in "…inter- and intrapersonal reasoning and deliberation" (NRW 2005, p.307). Presumably, then, fictionalists should want to preserve our practices of putting forward and evaluating moral arguments. However, some have argued that preface fictionalists are likely to have trouble doing so. Olson (2011c) and Lenman (2013) expect that the preface fictionalist will have difficulty licensing the rational transfer of acceptance from the premises of valid moral arguments to their conclusions.

To appreciate the problem, we can begin by noting how acceptance-transfer usually works: since the conclusion of a valid *modus ponens* argument (say) is a consequence of its premises, it seems that insofar as someone is rational, and they accept the premises, they have good reason to accept the conclusion.[164] This is just to say that acceptance, like truth, usually transfers from the premises to the conclusion of a valid argument.

Now, if moral acceptance is *belief*, we can offer a fairly straightforward explanation of this transfer of acceptance. The goal of truth-seeking is plausibly built into belief, and an individual with jointly inconsistent beliefs knows that they cannot all be true. Thus, someone who accepts the first two premises of a *modus ponens* argument will usually have a reason to believe the conclusion; she would be inconsistent not to do so. (See Oddie & Demetriou 2007, p.493.)

Yet the preface moral fictionalist's recommended attitudes towards moral propositions are attitudes of *make-belief*. These certainly don't seem to aim at truth. It's also not clear that make-believing incompatible propositions is inconsistent, nor that someone who make-believes the premises of a valid argument has a reason to make-

[164] My characterisation of this problem borrows heavily from Oddie & Demetriou (2007).

believe its conclusion. Preface moral fictionalists would therefore seem to have difficulty accommodating the inference-licensing properties of valid moral arguments. This challenge has been raised by James Lenman, who writes:

> If I believe that p and believe that if p then q, logic requires me to believe that q. But if I pretend that p and pretend that if p then q, logic doesn't require me to do anything. (2013, p.406)

The problem has also been noted by Jonas Olson, who claims that

> … pretending to believe the premises of a logically valid argument does not commit one to pretend to believe the conclusion. Acts of pretence and pretence attitudes, I submit, are not subject to norms of consistency. (2011c, p.191)

Yet the acceptance-transfer problem is not new—it has been brandished against non-cognitivists for quite some time.[165] So it would be surprising if there were no systematic attempts to address the problem. And there certainly have been such attempts. A wise move for the preface fictionalist would therefore be to turn to non-cognitivists for help; she might draw inspiration from their proposed solutions.

There are countless options to choose from here. Among these is Simon Blackburn's (1984, pp.192-5) development of the idea that there can be consistency constraints among non-cognitive attitudes. There is admittedly a common worry for Blackburn's proposal; many complain that the inferences that he licenses among moral judgments are not properly called *logical* inferences (Hale 1986, Schueler 1988, van Roojen 2005). What Blackburn rather shows is that someone must hold particular moral judgments if they hold others in order to be *pragmatically consistent* in their attitudes. (As Hale (1986, p.74) notes, "Don't do what you boo!" very much seems to be a requirement of pragmatic rather than logical consistency.) Yet this problem for Blackburn is not necessarily a problem for the preface fictionalist (see Svoboda 2017, p.70). The name of the game for the non-cognitivist is to provide an adequate descriptive account of moral practice. But the fictionalist's task is only to preserve something that will do the pragmatic work, and Blackburn's framework may very well suffice for these purposes.

[165] The *locus classicus* is Geach (1965). For more recent developments, see Dorr (2002) and Schroeder (2010).

Another option for the fictionalist would be to appeal to principles that seem to govern our engagement with fictions more generally.[166] Kendall Walton's (1990) account provides something in the way of guidance. Walton understands fictional truths (e.g., 'it is fictional that *p*') in terms of what a particular game of make-belief *prescribes us to imagine*. Within children's games of make-belief, these prescriptions to imagine might be understood as stipulations—for example, 'wherever there is a tree stump, imagine that there is a bear' (1990, p.38). Building upon this, the preface fictionalist could develop appropriateness conditions for moral make-belief—for example, by claiming that it is appropriate to make-believe some moral proposition *p* iff *p* is authorised by the relevant moral fiction, and by specifying which imaginings the moral fiction prescribes. To cover all bases here, she might even prescribe that it is inappropriate to imagine anything that is logically incompatible with what is stipulated by the fiction.

Admittedly, this is only a start.[167] But it is not my intention here to provide a fully-developed fictionalist proposal. My intention is only to give moral fictionalism a fair hearing, and I do not think those who raise worries regarding acceptance-transfer have done so. Though the fictionalist is perhaps not out of the woods just yet, she does have some promising strategies available to her.

§5.3.2 *A validity problem*

Joyce has suggested that prefix moral fictionalists will have difficulty accommodating the validity of moral arguments (2005, p.292).[168] To demonstrate the worry, he invites us to consider a prefix fictionalist, David, who approves of a valid argument of the following sort:

K1 If my cousin is an infant, then it is wrong to kill my cousin
K2 My cousin is an infant
K3 Therefore, it is wrong to kill my cousin.

[166] I thank Daniel Nolan for helpful discussions on this point.

[167] I acknowledge that further complications raised by Oddie and Demetriou (2007) would also need to be addressed.

[168] Joyce frames his discussion in terms of prefix *colour* fictionalism. But as I understand him, he takes the problem to generalise. So I will reframe things in terms of *moral* fictionalism here.

According to Joyce, David faces the following dilemma (2005, p.292). On the one hand, he might take only those sentences containing moral terminology (K1 and K3) to be prefixed by a fictional operator (e.g., 'according to the moral fiction…'). The problem with this move is that the argument above (which is surely valid) is no longer valid—it is no longer an instance of *modus ponens*. Alternatively, David might prefix *all three* claims with a fictional operator. This move maintains validity, but Joyce complains that it is utterly unmotivated; insofar as a moral fiction has any content at all, that content surely doesn't include any claims about the age of David's cousin.[169]

It seems to me that the moral fictionalist should simply take the second horn, which *pace* Joyce, I don't think is much of a horn. It is very common for some things that are true in fictions to be true in real life as well (Ryan 1980, Gendler 2000, Weatherson 2004, Nolan 2005). We often *import* truths about the actual world into fictional stories. We typically hold fixed, for example, the truths of logic and mathematics, the laws of physics, and perhaps even general facts about human psychology (Gendler 2000, p.78). Fictions are also governed by rules of *export*. That we export some truths from fictions should come as no surprise—people often learn about the past from great works of literature. One might, for example, learn how women wore their hair in nineteenth-century France by reading a story that takes place during that period (Gendler 2000, p.76).[170] It doesn't seem to be particularly problematic for the moral fictionalist, then, if some things that are true in the moral fiction—the age of someone's cousin, say—are true in real life as well. She simply needs to develop rules of import and export to account for this.

Joyce (2005, p.293) suggests that this would give rise to confusion. If David imports quite a number of literal truths into his moral fiction, then how shall we determine whether claims like 'my cousin is an infant' are intended as literal or fictional?

[169] It's worth noting (again) that I've slightly modified Joyce's original discussion of the colour fictionalist in order to explore the challenge as it applies to the prefix moral fictionalist. His original example is the following argument: (1) Fresh grass is green, (2) My lawn is made of fresh grass, (c) Therefore, my lawn is green. His complaint (which parallels the one above) is that "The fiction of a colored world, in so far as it has a determinate content at all, does not include claims about what anybody's lawn is made of" (2005, p.292).

[170] Of course, it is certainly possible to make *errors* in exportation. (See Mag Uidhir 2013.) This is perhaps more easily avoided in "distorting fictions" (Gendler 2000, p.77), which wear their departure from actuality on their sleeve.

The answer, it seems to me, is that we *are* very often able to know whether literal claims that have been imported into fictions are intended as literal or fictional. We often have ample evidence with which to distinguish someone who is speaking of *our* London from someone who is speaking of the London imported into the world of Sherlock Holmes. For the most part, context tends to do a lot of the work here. We take someone who claims that Sherlock Holmes lives nearer to Paddington Station than to Waterloo to have imported facts about London's geography into the story, and to be speaking of what is true of London within that story.[171] But in a context free of any mention of Holmes, Watson, or Moriarty, there seems to be no reason to take someone who says 'let's walk to Paddington station from Baker St, seeing as it's closer than Waterloo' to be speaking of the London described by Arthur Conan Doyle. It is open to the prefix moral fictionalist to argue that the context of her utterance will determine (or largely determine) whether or not a moral fictionalist is making a straightforward literal claim, or a literal claim that has been imported into the moral story.

§5.3.3 *A problem with preserving moral disagreement*

Fictionalists who takes moral discourse to be useful on account of its valuable role in coordinating interpersonal behaviour should want to make room for genuine moral disagreements. When we disagree over the moral permissibility of abortion, our claims seem to truly conflict with one another. Our impression that moral disagreements are genuine explains (in part) why it is that we can often hope to co-ordinate our behaviour through moral dispute and reasoned argument. If 'permissible' meant something *different* in your mouth than it did in mine, then we couldn't hope to co-ordinate our behaviour through coming to a(n apparent) consensus; if you agree to act under the supposition that abortion is permissible, and I agree to act under the supposition that abortion is permissible*, then what's to say that we've agreed to behave in the same way?

Yet some have argued that prefix fictionalists will have difficulty preserving genuine moral disagreement. Olson (2011c) makes the point by inviting us to imagine two

[171] The Paddington Station example is Lewis's (1978, p.41)

moral fictionalists: Con, a consequentialist, and Don, a deontologist. Con and Don (seem to) disagree about the right response to the trolley problem. Con claims that 'according to the moral fiction, one is morally required to flip the switch' (C), and Don claims that 'according to the moral fiction, one is *not* morally required to flip the switch' (D). On first appearances, C and D appear to be incompatible claims. But Olson argues that these claims are incompatible only if we take Con and Don to have subscribed to the *same* moral fiction. And he denies that the moral fictionalist is justified in assuming this:

> Since Con is, in the moral fictionalist discourse, a consequentialist, he is presumably referring to a fiction in which an action is morally required if and only if it maximizes happiness. And for parallel reasons, Don is, let's assume, referring to a fiction in which an action is morally impermissible if it treats a person as a mere means. (2011c, p.186)

Yet this strikes me as unfair. It is up to the moral fictionalist to decide how determinate the content of her fiction is to be. She could of course recommend a multitude of moral fictions, each of which specifies which actions (or kinds of actions) are morally right and wrong. But she might equally well recommend a single moral fiction, and hold that all that is true according to that moral fiction is that there exist such things as categorical reasons, rights, and duties. (Recall that Joyce (p.195) takes a moral fiction to be a "conceptual framework".) This more minimal fiction would not settle whether or not an action is morally required if and only if it maximises happiness, or whether or not an action is morally impermissible if it treats a person as a mere means.[172]

The latter option (adopting a more minimal moral fiction, that is) does not seem vulnerable to Olson's objection. It is possible for Con and Don to subscribe to the *same* moral fiction—one according to which there exists such things as categorical reasons, rights, and duties—but to genuinely disagree upon what their moral duties

[172] There is a slight complication here. Since we are operating under the assumption that the moral error theory is *necessarily* true, the moral fiction may very well describe an impossibility. How exactly we can reason about or from impossibilities is a good question, but it is not one that I will attempt to answer here. For discussion, see Nolan (1997). See also NRW (2005, pp.325-7), who suggest some ways in which a moral fictionalist could hope to get by with impossible moral fictions.

are. So long as the moral fiction is sufficiently indeterminate, it will be possible for two parties to disagree about which actions it prescribes.[173]

Of course, a moral fictionalist may want to avoid *excessive* indeterminacy. (If her fiction leaves *everything* open, then it may fail to be action-guiding.) But there are ways to constrain the content of a moral fiction. The platitudes that Joyce includes ("torturing babies to pass the time is always wrong") are not simply thrown in for good measure. They introduce conceptual constraints upon what sorts of things could be true according to the moral fiction. A moral fiction that included such platitudes wouldn't permit us to identify something's being morally good with its being a particular shade of blue.

Other constraints might arise from our psychological limitations. There is a well-known phenomenon of *imaginative resistance*. Though people can entertain many far-fetched fictional scenarios with ease, they find it difficult to comply with prescriptions to imagine deviant moral judgments— "in killing her baby, Giselda did the right thing; after all, it was a girl", for example (Walton 1994, p.37).[174] The moral fictionalist might therefore require that any candidate moral fiction be user-friendly (i.e., that its content be possible to imagine). This would place further restrictions upon what its content could be.

5.3.4 *The case against moral fictionalism: summing up*

I have argued that many criticisms of moral fictionalism should not strike us as especially troubling. What these criticisms should suggest to us is that fictionalists have some more house-keeping to do. It was not my intention here to get all of that house work done. But I have made some suggestions as to how the fictionalist might get started.

[173] Another option for the moral fictionalist would be to recommend multiple, more substantive moral fictions, and to explain how we can hope to achieve co-ordination through fictional moral disputes in the absence of genuine disagreement. To this end, she might borrow some resources from relativists (e.g., MacFarlane 2007).

[174] For discussion, see Walton (1994), Moran (1994), Gendler (2000), and Weatherson (2004). The phenomenon seems to apply to other sorts of deviant judgments as well—prescriptions to imagine mathematical falsehoods, for example.

§5.4 The real case against Moral Fictionalism

I will now proceed to develop what I take to be the *real* case against moral fictionalism, using Olson's (2014) critique of preface moral fictionalism as a launch-pad into the discussion. The moral fictionalist, recall, promises us the best of both worlds: she assures us that we can continue to enjoy the advantages of engaging in moral practice without holding onto our erroneous moral beliefs. According to Olson, there is a tension in this promise:

> One recommendation is to practice self-surveillance to make sure moral belief is avoided. This seems to involve occasionally attending to the belief that morality is fiction. A second recommendation is to suppress or silence belief to the effect that morality is fiction. This leads to instability in that while ways of thought and behaviour likely to prompt moral belief are recommended, moral belief is to be avoided. (2014, p.190)

Thus, the moral fictionalist seems to face the following dilemma. If she wants to make good on her epistemic promise, then she must avoid becoming too immersed in her moral pretence. Otherwise, she risks a slip into moral belief (horn one). To avoid this slip, she must exercise epistemic caution, regularly attending to her belief that morality is merely a fiction. Such caution is, however, in tension with her practical promise; reminding herself that morality is a fiction is likely to prevent her from taking the pretence suitably seriously (horn two).

I think that the fictionalist can avoid impaling herself on Olson's first horn; she is not plausibly guilty of any moral believing. In what follows, I will argue for this claim on the fictionalist's behalf by drawing attention to important behavioural and psychological differences between attitudes of belief and make-belief (§5.4.1).

However, I also think that avoiding Olson's first horn makes the second considerably *worse* for the fictionalist. Once we appreciate these important differences between belief and make-belief, we see why it is that the fictionalist cannot plausibly make good on her practical promise. Certain characteristic features of make-belief suggest that it would be a rather poor substitute for moral belief (§5.4.2). In §5.4.3, I will argue that parallel problems apply to the fictional *beliefs* recommended by prefix fictionalists as well. Nothing I say should suggest to us that fictionalist attitudes would be *utterly useless*. But they should lead us to doubt the extent to which fictionalist attitudes are likely to be as useful as full-blooded moral beliefs. This insight is, in large, part, what motivates the conservationist proposal developed in the following chapters.

§5.4.1 *Horn 1: the slip into belief*

Olson argues that the immersed moral fictionalist is very likely to slip back into moral belief, reneging on her epistemic promise. Yet not much is said to motivate this claim. In its defence, Olson does little more than point out that moral make-believers would have to adopt incredibly similar behavioural and psychological dispositions as moral believers (with respect to moral propositions):

> … someone who takes up a fictionalist stance to morality adopts certain behavioural dispositions and backs them up by moralizing her thoughts, i.e. by thinking of certain actions as wrong, unfair, or undeserved, etc.… But given successful adoption of the relevant behavioural dispositions, it seems difficult in many cases to avoid *believing* the relevant moral propositions, as opposed to merely *accepting* them or *thinking* about them. (2014, p.189, emphasis in original)

> … acquiring physical and psychological dispositions to behave in accordance with the fictional moral norms makes it all the more likely that one slips from moralized thought into moral belief.[175] (2014, p.189)

This at least gives us something to work with. In what follows, I will build a case for Olson's first horn on his behalf, suggesting two ways in which it might be argued that the immersed moral fictionalist risks a slip into moral belief. Neither, I will argue, is promising.

First, the idea that make-belief is likely to lead to genuine belief might be advanced as an *empirical claim*. Perhaps it is a psychological fact about ordinary agents that when immersed in a pretence, they usually find it difficult to avoid believing what they initially set out to merely pretend. Unfortunately, Olson doesn't provide us with any evidence to support this claim. And there is evidence that speaks against it.

[175] Garner briefly voices a similar suspicion when he suggests that if the moral fictionalist "…begins to have moral feelings, moral outrage, moral guilt, and moral arguments, then we have every reason to say that he has reverted to his former moral beliefs, and to the error he once identified and abandoned" (2007, p.509).

Experimental studies suggest that pretence-subserving representations are typically "quarantined" from representations of reality in games of make-belief (Leslie 1987). A child may pretend that a banana is a telephone, but she never loses sight of the fact that it is a banana. The events that occur within a pretence are only taken to have effects within that circumscribed domain; when the child pretends to talk to her father on the banana/telephone, she does not afterwards believe that she ever really had that conversation (Nichols & Stich 2000, p.120). These events are not treated as relevant to guiding action in the outside world—if someone breaks a leg, no one uses a banana to call an ambulance.

Acts of pretence might therefore be said involve a kind of dual-representation; the child simultaneously represents the banana as a banana, and as a telephone. What is important is that she never surrenders the capacity to distinguish what is real from what is imagined. Indeed, this distinguishing capacity is so commonplace that some have taken its absence as a sign of pathology; it has been suggested that some mental disorders might profitably be explained in terms of a failure to quarantine what is real from what is imagined. (See Currie 2000.) Absent any argument to the effect that immersed make-belief merits such a worrying prognosis, it seems uncharitable to deny to the fictionalist a distinguishing capacity that is reliably present in young children.

We also seem to have decent anecdotal evidence that make-belief isn't likely to lead to belief. Currie and Sterelny put the point particularly nicely:

> … the view that we believe in the fictions we encounter, even when we get deeply absorbed in those fictions, remains a non-starter. We do not rush on stage to Desdemona's defence, or seek to intervene in even the most naturalistically staged aggressions…If we say that those engaged by a fiction are limited believers of it, we are bound to say that being engaged by a fiction is an epistemic vice. If we say that those engaged by a murder on stage believe, to some degree, that a murder is about to be committed in front of them, we are bound to say that their unwillingness to intervene is a moral vice. But fictions are not entrapment tools, and responsiveness to fiction is an imaginative virtue, not a vice of any kind. (2000, pp.150-51)

Following Currie and Sterelny, the claim that an immersed make-believer risks a slip into belief would seem to wholly mischaracterise our engagement with fiction. Worse still, it casts those who partake in pretence as moral villains or epistemic failures—which is neither plausible, nor appealing.

As an empirical claim, the idea that make-belief is likely to lead to belief seems implausible. However, there is another kind of argument that could be developed on Olson's behalf. Perhaps we have simply been operating upon a faulty understanding of belief. Given certain assumptions about the conditions under which an agent believes that *p*, perhaps immersed make-belief would (as a *conceptual* matter) properly be classified as genuine belief after all.

Of course, this objection would require adopting a particular understanding of belief. Given Olson's remarks concerning the behavioural and psychological dispositions of moral make-believers, a natural route (at least for the sake of exploring this option) would be to operate upon a *dispositionalist* account of belief, which I shall understand as follows:

> **Dispositionalism**
> Beliefs are to be characterised in terms of an agent's dispositions.[176]

The common thread that runs through different varieties of dispositionalism is the idea that beliefs can be characterised in terms of their functional roles. The relevant roles here, being dispositions, are 'forward-looking' in that they tell us what kinds of states a belief that *p* typically brings about (Stalnaker 1984). One who believes that *p* will, for instance, typically be disposed to assent to *q* if shown that *p* implies *q*, to express surprise upon hearing that not *p*, and so on. An agent who has these and other relevant psychological and behavioural dispositions can be said to believe that *p*. (These claims plausibly come attached with tacit 'if' clauses; an agent who believes that *p* will typically be disposed to assert that *p* *if* she has not decided to deceive others about her beliefs, *if* she has not lost her voice, and so on. (See Schwitzgebel 2002, p.253.) I will return to this issue in §6.3, where drawing out some of the finer lineaments of dispositionalism will be needed. The basic idea should suffice for now.)

Here, then, is how we might go about arguing for the conceptual claim that immersed make-belief is properly classified as belief. We might begin with the idea that moral believers have particular dispositions in virtue of which they can be said to be-

[176] I've intentionally offered a schematic characterisation here that is neutral between varieties of dispositionalism that identify beliefs with the mental states that typically cause particular dispositions, and those that identify beliefs with the dispositions themselves. My arguments do not rest upon going one way or the other.

lieve some moral proposition *p*. We might then point out that moral *make-believers* do not seem meaningfully different from moral believers at all with respect to these dispositions. Indeed, it would seem to be more or less business as usual for the convert to fictionalism; she is so immersed in her pretence that she scarcely pays any attention at all to the fact that it but a mere fiction. Her moral thoughts come to be "well-rooted habits of thinking" (Joyce, pp.218-9). She continues to enter into moral disputes, and to speak of "goodness and badness, rightness and wrongness, duties, justice, and obligations" (NRW 2005, pp.311-2).

All in all, the psychological and behavioural dispositions of a make-believer seem to be *incredibly* similar to those of a moral believer. This charitable reconstruction is an alternative way to make sense of Olson's claim that an immersed fictionalist is a run-of-the-mill believer. A make-belief walks like a belief, talks like a belief, and sounds like a belief. So the onus is upon the fictionalist to tell us why it *isn't* a belief. (As Douglas Adams (2014, p.227) puts it, "If it looks like a duck, and quacks like a duck, we have at least to consider the possibility that we have a small aquatic bird of the family anatidae on our hands".)

This problem is not new. Distinguishing beliefs from fictionalist attitudes has long been a challenge for fictionalist proposals. Many suspect that the distinction here is bogus; that it is a "distinction without a difference" (Horwich 1991, p.3; see also O'Leary-Hawthorne 1997, Rosen & Burgess 2004). Since the psychological and behavioural profile of (certain) fictionalists attitudes seems indistinguishable from that of belief, it is tempting to think that there is no real difference between them. It is my contention that this burden can be discharged as far as make-beliefs are concerned. As I shall now argue, the fictionalist can meet the challenge by pointing towards important differences between believers and make-believers.[177]

One possible difference between the dispositions of a believer and those of a make-believer was considered earlier (§5.2.1). Joyce, recall, distinguishes his moral fictionalist from a moral believer by appealing to her disposition to *withhold assent from*

[177] Since I am responding here on behalf of moral fictionalists who recommend attitudes of make-belief, what I have to say may not double as a defence of the fictionalists targeted by Horwich (1991) and others, who recommend attitudes of *acceptance*. The slip-into-belief worry may very well be harder to avoid in this case. I will discuss acceptance in greater detail in §7.1.

positive, first-order moral claims in critical contexts. This suggests the following difference between belief and make-belief:

> **BEL$_1$**
> One who believes that *p* will typically be disposed to assent to *p* in critical and non-critical contexts.
>
> **M-BEL$_1$**
> One who make-believes that *p* will typically be disposed to assent to *p* in non-critical contexts and to withhold assent from *p* in critical contexts.

This distinction has some initial appeal. However, I worry that BEL$_1$ is false. As Joyce characterises a critical context, it is one in which an agent "investigates and challenges the presuppositions of ordinary thinking"—the philosophy classroom being a prime example (pp.190-94). But it is far from obvious that we are disposed to assent even to very firmly held beliefs in critical contexts, so understood. Plausibly, the average student believes that she has hands, that there are numbers, and that the sun will rise tomorrow. But she may not be so willing to assent to such claims in the philosophy classroom.

But no matter. There are other important differences between believers and make-believers. To begin with, make-beliefs can typically be resisted or acquired at will. Indeed, I would go so far as to say that it is *constitutive* of make-belief that it be tied to our will. We *choose* to participate in games of make-belief—we *elect* to imagine, and we *decide* to pretend. It seems close to being analytic that an individual could not pretend to do something that she did not set out to pretend. As Searle (1975, p.325) notes, 'pretend' is an intentional verb, one with the concept of intention built into it.

Not only are make-beliefs the sorts of attitudes that we can choose to dispense with—they are also especially liable to being dispensed with when the practical stakes are high. Oddie and Demetriou offer a nice illustrative example:

> Suppose it is true in [a] play that two people are chatting comfortably on a couch in their home, and that their home is not on fire. As we get into the play, we make-believe that that is true. Now, if smoke starts seeping onto the stage from backstage but it is clearly true, in the play, that there is no smoke in the room, we tend in such circumstances to abandon the make-belief (that there is no smoke in the room) and go with the belief (that there is smoke), and it is entirely reasonable to do so. (2007, p.487)

Following Oddie and Demetriou, make-beliefs are "highly overridable" attitudes (2007, p.487). When the practical stakes go up, we typically can and do dispense with them—and rightly so if they cease to be useful to us.

In this respect, make-belief is markedly different from belief. Indeed, some claim that an attitude so intimately tied to one's will could not possibly be a belief (e.g., Railton 2003). But a far weaker claim will serve the distinction that I am seeking to establish here. Whatever one has to say about the control that we may be able to exert over our beliefs, it certainly doesn't seem *constitutive* of belief that it be tied to our will. There are attitudes that urge themselves upon us that are properly called beliefs. The clearest cases are perhaps the beliefs that we form on the basis of perception; our coming to believe that there is a house in front of us after having perceived a house, say. It's hard to make the case that one has any real say on the matter here.[178]

Moreover, one does not typically cease believing that *p* when that belief is contrary to one's practical interests.[179] Suppose that I have excellent evidence that my partner has cheated on me, and that this has a significant impact upon my daily functioning; I feel hurt and betrayed, and am consumed by a desire for revenge. It might be in my interests to dispense with my beliefs regarding their infidelity. (Doing so might help us to go back to the way things were.) But it seems implausible to say that someone in my situation would typically be capable of doing so.

What I am proposing, then, is that we have comparatively little *control* over our beliefs. Make-beliefs are easily overridden—and rightly so when the pretence ceases to be useful to us. But we neither do nor can typically dispense with a belief that *p* after

[178] I am not claiming that we always acquire perceptual beliefs simply in virtue of seeing the world. We might, for instance, have independent evidence that what we are seeing is an illusion. I only claim that when one sees that the world is thus-and-so, one also typically believes that the world is thus-and-so. If one sees a certain object as a tree, then, absent any countervailing considerations, it is typically difficult not to believe that there is indeed a tree there. For further discussion of belief being the "default state", see Egan (2008, pp.55-8).

[179] There might be exceptional cases of, course. Under threat of torture, Winston (of Orwell's *1984*) comes to believe that 2+2=5. But such exceptional shouldn't pose too great a threat to the general claim that beliefs do not typically fall by the wayside when the practical stakes go up.

having come to appreciate that such a belief is contrary to our practical interests.[180] This suggests the following difference between belief and make-belief:

> **BEL$_2$**
> One who believes that *p* will typically persist with that belief even when the stakes are suitably high so as to render that belief contrary to her practical interests.
>
> **M-BEL$_2$**
> One who make-believes that *p* will not typically persist with that make-belief when the stakes are suitably high so as to render that make-belief contrary to her practical interests.

BEL$_2$ dovetails nicely with the well-known idea that beliefs aim at truth (Williams 1973, Velleman 2000, Wedgwood 2002, Boghossian 2003, Shah 2003).[181] To maintain that we are typically capable of believing whatever we like at will would be to suggest that we can typically believe propositions without *any regard whatever* as to whether they are true (Williams 1973, Velleman 2000). Yet that seems to run contrary to the popular idea that part of what it is to believe that *p* is to hold that belief that accountable to truth. Perhaps this is why it strikes many as odd to think we can choose beliefs at our fancy.

Yet another difference between belief and make-belief concerns the distinctive ways in which these attitudes integrate with the rest of our psychology and our behaviour. I will here focus upon the different ways in which belief, as opposed to make-belief, elicits particular emotional responses. As Shaun Nichols points out, our affective responses to fictions can differ strikingly from our responses to belief:

> At the end of Dr. Strangelove, we imagine that all human life is about to be destroyed, and we find this amusing in the context of the film. Presumably this is not how we would react if we had the real belief that all human life is about to be destroyed. Perhaps if we really believed that all human life was about to be destroyed, we could find some humor in the situation, but sure-

[180] Doxastic voluntarists might beg to differ, of course. But doxastic voluntarism is not especially popular, nor appealing—at least not in its most philosophically interesting manifestations. (More on this in §7.4.)

[181] I acknowledge that claims regarding the truth-directedness of belief are open to a number of different interpretations; such claims may be claims regarding what is *conceptually* constitutive of belief (as in Boghossian 2003) or what is constitutive of the *essence* of belief (as in Velleman 2000, Wedgwood 2002), or something else still. Since the basic point here does not rest heavily upon any single precisification, I will leave the matter open.

> ly this would not be the predominant emotional response. (2006, p.464)

We shouldn't oversell the point here. Our affective responses to fictions don't always differ *strikingly* from our responses to belief. (Few feel warm and fuzzy inside when watching man-eating spiders on their television.) Nonetheless, these affective responses can be expected to differ in a number of important respects. One who believes that there is a huntsman spider lurking beneath their bed (a reasonable belief to have when one lives in Australia), is apt to feel a very real kind of fear—a fear that someone who make-believes that a rock is a spider is unlikely to experience.

This is not to deny that our engagement with fictions can prompt emotional response.[182] Nor is it to deny that these responses can prompt *similar* kinds of behaviour. As Nichols and Stich note when reporting the results of a study:

> … in our burglar in the basement scenario, one subject picked up the phone that was available and dialed 9-1-1. However, she took precautions to ensure that the call did not really go through. She didn't want her behavior to be that similar…she wanted to be sure that the police didn't really come. (2000, p.129)

The important point to appreciate is that whatever affective responses make-beliefs are capable of eliciting, these can be expected to integrate with the rest of our psychology and our behaviour in importantly different ways than the affective responses triggered by belief. Someone who believes that Freddie Kruger exists is likely to fear encountering him in her dreams, and to take steps to ensure her safety; she might sleep less often, or bring a crucifix to bed. One who merely make-believes that Freddie Kruger exists while watching *A Nightmare on Elm St* is unlikely to take such steps to avoid him. (Though she may have more trouble than usual falling asleep.) This suggests another important difference between attitudes of belief and make-belief:

> **BEL$_3$**
> One who believes that *p* will typically be disposed to experience the emotions stereotypically associated with believing that *p*, and to engage in the behaviour stereotypically associated with believing that *p*.

[182] Though some (e.g. Walton 1990, ch.7) do come close to denying this.

> **M-BEL$_3$**
> One who make-believes that *p* will not typically be disposed to experience the emotions stereotypically associated with believing that *p*, or to engage in the behaviour stereotypically associated with believing that *p*.

Three quick caveats here. First, although talk of stereotypical emotions and behaviour seems sensible enough, we may sometimes want to relativise stereotypical responses to a particular individual. Second, and as I have suggested, it is consistent with M-BEL$_3$ that make-belief can prompt *somewhat similar* emotions and behaviour to belief. Finally, although I have used fear as a demonstrative example, it is not at all implausible that the same holds true for other kind of affective response. One who loves their neighbour will be motivated to seek him out and win his affections, but someone who feels for Emily Brontë's Heathcliff will make no attempts to get in touch. One who loses a pet one will mourn their death and experience grief. But no one holds a funeral for Lassie.

Let's take stock. We began with Olson's suspicion that an immersed moral make-believer is at risk of relapsing into moral belief. I made two attempts to put some flesh on the bones of this suspicion. The first attempt was to construe the claim as an empirical one—an unpromising move, I argued. The second was to understand Olson's suspicion as stemming from particular conceptual presuppositions regarding the conditions under which an agent believes that *p*. In order to address the latter concern, the moral fictionalist needed to make the case for a meaningful distinction between make-believers and believers. I have argued that she is up to the task; attitudes of make-belief are importantly different from beliefs. So the fictionalist evades Olson's first horn—she is no moral believer. However, and as I will now argue, these insights only serve to sharpen the second horn; make-belief may very well be *too* different from belief to sustain a useful kind of moral practice.

§5.4.2 *Horn 2: the shortcomings of moral make-belief*

My objection to the fictionalist in what follows is similar to Olson's in spirit, but differs in its letter. Olson, recall, thinks that a moral make-believer must constantly remind herself that morality is a fiction in order to avoid becoming a moral believer. According to him, however, this act of constantly reminding herself that morality is merely a fiction will undo the practical benefits that the moral pretence promises to supply. I have rejected Olson's claim that the fictionalist must regularly attend to her belief that morality is a fiction in order to avoid a slip into belief. The important differences between these attitudes are sufficient to absolve her of the charge that she is guilty of any moral believing. Thus, there is no danger of such frequent reminders impacting upon the practical effects of the pretence.

However, there is a more fundamental worry in the vicinity here. Even if make-belief is not liable to slip into belief, the nature of make-belief itself is at odds with a firm enough commitment to moral practice that would be required to secure the practical benefits of moral behaviour. The differences between make-belief and belief that I have emphasised suggest that moral practice would be far less *useful* if it were preserved as a fiction. The problem isn't merely that make-beliefs wouldn't play the *same* role as beliefs—that is to be expected, after all. The criticism that I will now proceed to develop is that moral make-beliefs are unfit to play a *suitably similar* role.

My first concern pertains to the stability of the pretence. Consider Joyce's portrait of the fictionalist; one who assents to moral claims in non-critical contexts, only to withhold assent from them in critical contexts. The potential for such epistemic see-sawing seems likely to undermine the practical efficacy of the moral fiction; what's to stop a moral fictionalist from entering into a critical context whenever doing so is useful for her—when faced with the choice of breaking an inconvenient promise, say? Joyce anticipates this concern. He is at great pains to emphasise that

> …the decision to adopt morality as a fiction is not an ongoing calculation that one makes over and over…the resolution to accept the moral point of view is something that occurred in the person's past…–it is what Jon Elster calls a "precommitment." Its role is that when entering a shop the possibility of stealing doesn't even enter one's mind. (pp.223-4)

Joyce likens the fictionalist's precommitment to that of Odysseus, who ties himself to the mast of his ship to avoid the temptation of the sirens (p.224). The idea here is, I take it, that we are often capable of doing things psychologically that have similar effects to objective constraints (e.g., Odysseus's mast). Someone who would be fine having one drink might commit herself to not drinking at all because she knows that one drink leads to more. Or, to borrow another example from Joyce, someone who might remain fit even if she took the odd day off exercise might commit herself to doing exactly fifty sit-ups every day in order to avoid a "slippery slope to inactivity" (p.215.) A would-be moral fictionalist could, the thought seems to be, commit herself to keeping up the pretence in much the same way.[183]

Yet this seems to weaken the justification for becoming a moral fictionalist to begin with. Joyce doesn't want to over-emphasise the extent to which we are *already* capable of doing things psychologically that have a similar effect to objective constraints. Otherwise it is no longer clear that we really need a moral fiction to secure co-operative behaviour. Joyce suggests that we can effectively commit ourselves to doing fifty sit-ups a day by thinking that we *must* (non-morally) do so no matter what if we are to become fit. Why, then, are we not similarly capable of carrying out a commitment to be co-operative by thinking that we *must* (non-morally) do so no matter what in order to secure the goods of co-operative behaviour? Nadeem Hussain touches upon this concern:

> Surely by Joyce's own lights, a thought that I ought to do fifty sit-ups plays a very similar role to the thought that I morally ought not to steal. Such thoughts are instrumentally effective in helping me pursue my long-term self-interest. (2004, p.168)

As does Lenman when he writes,

> …it's hard to see how [the pretence] will help us much when the stakes are at all high except insofar as we remind ourselves of the urgency of the reasons we had to adopt it. And then, as with the exercise case, it all starts to look rather unnecessary. (2013, pp.405-6)

If we are already capable of doing things psychologically that have similar effects to objective constraints, then it's not clear how much there is to be gained from enlisting a moral fiction. Why won't it do to commit myself to co-operative behaviour by thinking that I *must* (non-morally) co-operate come what may if I am to further my long-term interests? This problem may not be insurmountable. But it does alert us to some tensions in the motivations for going fictionalist.

A second worry pertains to Joyce's way of carving up the contexts in which we are to engage in moral make-belief and those in which we are not. When we find ourselves in critical contexts, the belief in moral error theory enters into the equation. In

[183] I thank Daniel Nolan for helpful discussions on this point.

everyday, humdrum contexts, moral make-beliefs make an appearance. But moral make-beliefs may very well be *needed* in critical contexts. Consider a meta-ethics seminar in which a speaker's talk is inexcusably sexist. We should want make-beliefs regarding the wrongness of sexism to be available here.[184] But we are to have no recourse to these make-beliefs in the meta-ethics seminar. This is an undesirable result. If we are going to adopt fictionalist attitudes, then we should want to be able to put them to use when we need them.

A third concern is that make-beliefs don't seem like a strong enough basis for securing the practical benefits that the fictionalist promises to deliver. By way of contrast, we can begin by reflecting upon the nature of moral beliefs. Consider someone in dire financial straits who is tempted to siphon some funds from an unsuspecting friend. She believes that doing so would be morally wrong, and so, she is ultimately motivated to refrain. Joyce, recall, thinks that it is precisely this feature of moral beliefs that explains their usefulness; they prompt us to behave in ways that are in our long-term interests (p.181). True, a cash injection might further an agent's immediate interests. But stealing from her friends is also likely to be contrary to her long-term interests, insofar as friendship and a trustworthy reputation are the sorts of things that benefit her in the long run.

Now, moral beliefs are beliefs, and so, they plausibly have the following feature which, I have argued, is characteristic of belief: they persist even when the stakes are suitably high so as to render them contrary to our practical interests. And this is, in large part, what accounts for their usefulness. Just imagine what would follow if moral beliefs were apt to *disappear* whenever convenient. Suppose that we were accustomed to cease believing in the cruelty of the animal fur industry whenever we eyed an attractive pair of leather boots, or to cease believing that stealing from a friend was morally wrong whenever we were short on funds.[185] It is doubtful that moral be-

[184] We should also presumably want some recourse to the sanctioning resources of moral language. It would be preferable to call the sexist an inconsiderate cad; declaring that we don't like him very much doesn't quite get at the itch that we want to scratch.

[185] Of course, people do sometimes act contrary to their moral beliefs, and they often have creative ways of justifying these transgressions to themselves. But moral beliefs don't typically *disappear* whenever convenient. In this respect, they seem to me to be more reliable than make-beliefs.

liefs would play the same role if they were apt to disappear when they were most needed to stave off temptation.

Yet *make-beliefs* have precisely this feature; they are highly overridable, and they typically *are* overridden when the practical stakes are high (Oddie & Demetriou 2007, p.487). And the practical stakes can be incredibly high in moral decision-making. As Lillehammer points out, morality "frequently prescribes costly sacrifices, such as the abandonment of basic personal projects…" (2004, p.103; see also Cuneo & Christy 2011, p.99). Whatever long-term gains the fiction affords us, then, it is doubtful that these will be sufficient to motivate persisting with the pretence when our important personal interests are at stake. The situations in which we would most want moral make-beliefs to work for us are precisely those situations in which they are apt to let us down.

A final concern for the preface fictionalist concerns the ways in which make-beliefs are likely to integrate with our psychology and behaviour. Most of us are emotionally invested in moral issues. We are often outraged when we believe that someone has committed a serious moral wrong, and feel guilty upon coming to believe that we ourselves have engaged in moral wrongdoing. Moreover, these affective experiences typically motivate distinct sorts of behaviour; moral transgressions are usually met with punishment (Boyd et. al 2003), and guilt usually prompts us to make amends (Wicker et. al 1983; Ferguson et. al 1991; Tangney et. al 2013). All in all, our moral beliefs tend to *integrate* with our behaviour (and the rest of our psychology) in fairly predictable and distinctive ways.

An important question for the moral fictionalist is whether make-belief is likely to integrate with our behaviour and our psychology in a suitably similar way; is make-believe capable of eliciting these powerful emotions and the behaviours that they usually motivate? This is far from obvious. If a spider the size of Tolkien's Shelob appeared at my window (and if I believed it was there), I would experience a very real kind of fear—something I plausibly don't experience when I watch Shelob pursuing Frodo Baggins. In the former case, I would surely be motivated to barricade the windows, and to seek out help. When watching Frodo flee, I merely avert my eyes in horror.

Yet moral fictionalists are confident that moral make-beliefs *would* be capable of eliciting the affective responses typically prompted by moral beliefs. Joyce proposes that fictive thoughts

> … can engage our emotions. If one sits vividly thinking about one's house burning down and all one's worldly belongings with it—not believing it, nor even believing it particularly likely—that may be sufficient to prompt anxiety or fear. This, I take it, is what happens when we engage emotionally with fiction–when we feel fear at horror movies or sadness at novels. (p.197; see also Joyce 2006, p.99)

As I understand Joyce, the basic idea here is that a lack of belief doesn't necessarily imply a lack of emotion. We very often *do* seem to be emotionally responsive to fiction. So perhaps beliefs have limited penetration into emotion; even when we don't believe that the objects of our fear are instantiated, we are still apt to feel (at least something meaningfully like) fear. And perhaps this lesson applies to moral fictions; these may engage us emotionally, even if we don't believe that there is any such thing as moral wrongness.

Yet the analogy with paradigmatic fictions here will only get the moral fictionalist so far. In order for the moral fictionalist to establish that a moral fiction is likely to engage us emotionally, it won't suffice to point out that fictions are generally capable of prompting emotion. She must make the case that *a moral fiction* will be so capable. This is difficult to establish—especially once we reflect upon why it is that fictions often elicit affective responses. There is overwhelming empirical evidence that our emotional engagement with fictions is very often the result of our having identified with *the characters* that are portrayed. People usually adopt the standpoint of the protagonist (Rinck & Bower 1995), and process the emotional implications of the fiction's events from their perspective (Gernsbacher et. al, 1992). For this reason, empathetic perspective-taking is increasingly thought to be "a standard part of… engagement with fictional narratives" (Coplan 2004, p.143; see also Harris, 2000, pp. 70-78; Bourg 1996, pp. 246-257; Rall & Harris 2000, pp. 206-207). It is plausibly because of this identification with fictional characters that we often feel on behalf of them; we share in Captain Kirk's sadness when Spock sacrifices himself for the needs of the many. (And we share in his apathy when those unfortunate enough to be wearing red shirts suffer an untimely demise.)

Yet in what way are such cases analogous to the case of a moral fiction? They seem to be to me rather disanalogous. The features in virtue of which we are emotionally 'taken in' by fictions would seem to be absent in *moral* fictions. There are no protagonists to be found in this moral story. There may be a few abstract entities floating around, of course (rights, duties, obligations, and what-have-you). But these presumably aren't the sorts of characters with whom we are likely to engage.

I am not arguing here that a moral fiction could not engage us emotionally. What I am rather arguing is that pointing towards our engagement with paradigmatic fictions does not seem to provide us with strong evidence that it could. The issue is that the fictionalist is using our emotional engagement with paradigmatic fictions as evidence for her claim that a moral fiction can be expected to engage us emotionally. Yet the features which explain our emotional engagement with paradigmatic fictions seem absent in moral fictions.

Even granting that the moral fiction would engage us emotionally, there is a further problem for the moral fictionalist. It is not merely that moral judgments prompt emotions; these emotions also tend to motivate distinct kinds of behaviour—punishment, reparations, and the like. Even if moral make-believe could elicit *bona fide* emotions, then, we would still be left with the following question: are these emotions likely to motivate useful sorts of behaviour?

It is not unlikely that the emotions prompted by make-belief can motivate *somewhat similar* sorts of behaviour as those prompted by belief—especially within the confines of the pretence. One plausible understanding of participation in make-belief takes it to be driven by a desire to behave in a *similar* way to how one would behave were the pretence a reality (Nichols & Stich 2000, p.128). But there are important limits, especially when behaving in *too similar* a way would carry real life costs. Someone pretending that there is a burglar in her house is careful not to actually call 9-1-1.

This suggests a worry for the moral fictionalist. Even if moral make-belief prompted certain emotions, it is debatable whether these would motivate the co-operative behaviours that we tend to associate with them. Agents who participate in moral make-believe may very well desire to behave in a *somewhat similar* manner to how they would were the moral fiction reality. Yet they are likely to stop short of behaving in the *same* manner when that behaviour is sufficiently costly—as punishment and making amends can often be.

Let's summarise. In §5.4.1, I argued that the preface moral fictionalist can plausibly make good on her epistemic promise. However, I have now suggested (in §5.4.2) that she is likely to have trouble keeping up the practical end of the bargain. Moral make-belief seems to be neither a stable enough nor a strong enough basis for securing the practical benefits that the fictionalist promises us.

Yet it might be thought that these problems do not apply to the *prefix fictionalist*, who shirks make-belief in favour of beliefs (beliefs about what is true according to some moral fiction). As I will now proceed to argue, however, a parallel challenge can be raised against the prefix fictionalist as well.

§5.4.3 *The shortcomings of fictional belief*

In order to properly understand why the prefix fictionalist cannot deliver the practical goods, I want to begin, once again, by comparing her to our moral believer. The prefix fictionalist differs from the moral believer not in virtue of holding a distinct sort of attitude towards moral propositions, but in what the content of her attitudes are. Whereas the moral believer believes some moral proposition *p* (e.g., 'gratuitous killing is wrong'), the prefix fictionalist believes that *p**: that 'according to the moral fiction, *p*'. (Hereafter, I shall use '*p*' to refer to a proposition lacking a fictional operator and '*p**' to refer to its fictional counterpart, 'according to the fiction, *p*'.)

This difference in the content of the fictionalist's beliefs can be expected to carry important implications for her behaviour. This should come as no surprise. The claim that beliefs with different contents generate distinct behavioural dispositions is so obvious as to be almost unworthy of mention.[186] It is for this reason that a belief that has the content *p** can be expected to have very different implications for behav-

[186] Of course, beliefs don't generate behavioural dispositions all by themselves; desires (and downstream, intentions) also have a role to play. The point is simply that two agents with the same desires but very different beliefs about the world are likely to have different behavioural dispositions. There are tricky cases here, of course—indexicals, and differences in wide content among them. I put these to the side for the purposes of the discussion.

iour compared to a belief with the content *p*. Indeed, it is my suspicion that a belief that *p** will have very similar properties to *a make-belief* that *p*.

Like a make-belief that *p*, a belief that *p** seems liable to being overridden when the practical stakes are high. I hasten to add, however, that this isn't overridability of the kind that spelled trouble for make-belief. The overridability of make-belief was explained by its being constitutively tied to our will—by its being an attitude that can typically be *dispensed with* at our choosing. A belief that *p** isn't overridable *in this sense.* (It is still a belief, after all.) Nonetheless, a belief that *p** does seem overridable in a slightly different sense.

Let me explain. Suppose that you are a fan of Arthur Conan Doyle, and believe that according to Doyle's fiction, Holmes lives on Baker St. You very much desire to see Holmes's place of residence, and so, you elect to pay Baker St a visit. Imagine that before setting off, you are informed (and comes to believe) that Baker St has recently become infested with pick-pocketers. You now believe that *d*: 'Baker St is an especially dangerous part of town'. But of course, you also believe that ¬*d**: 'according to Doyle's fiction, Baker St is not an especially dangerous part of town'. (Baker St's crime rate is no part of the stories, let's suppose.)

What would you to do in such circumstances? Would you simply disregard your belief that *d,* allowing your belief that ¬*d** to guide your behaviour instead? One should hope not. Under such circumstances, I should expect that you would cast to the side your belief that ¬*d**, and act on your belief that *d* instead, steering clear of Baker St as best you can. Given that your personal interests are at stake, you would seem to do better to attend to your beliefs concerning the real Baker St rather than the purely fictional one.

Here, then, is my conjecture: a belief that *p**, though not dispensable, is nonetheless something that is apt to be *cast to the side* when the practical stakes go up. It is something to which we will choose not to attend—something from we will shift our attention.[187] Not only that, but a belief that *p** is *especially liable* to being cast to the

[187] It is a separate question just what is going on in our minds when we cast a belief to the side. Perhaps it is put back into storage in one's 'belief box'. I have not much to say on this, and I suspect that having more to say would require a foray into the metaphysics of mind—something well beyond the scope of this thesis.

side when the practical stakes go up. Under such circumstances, a belief that *p** inevitability gives way to beliefs about reality.

There is a natural explanation as to why this is so. On the usual view of things, a belief is an attitude with a world-to-mind direction of fit. (See Anscombe 1957.) The job of beliefs is to represent the world—they provide us with a kind of 'map' with which to steer our way around (Ramsey 1931). Now, there is clearly a sense in which a belief that *p** is doing its job; it is representing the world to be a certain way—namely, as containing a fiction with a distinct sort of content. But care needs to be taken in *how* we allow that belief to navigate our actions. You (a Holmes fan, we are supposing) would be ill-advised to permit the belief that ¬*d** to guide your behaviour when you consider paying Baker St a visit. When the practical stakes go up, we turn to our beliefs about *reality* to successfully navigate our way around. Under such circumstances, affording our fictional beliefs an important role in deliberation is likely to be an impediment to achieving our ends.

What does all of this mean for the prefix moral fictionalist, though? Well, like a moral make-belief that *p*, a belief that *p** seems liable to be overridden (albeit in a distinct sense) when our important personal interests are at stake. When faced with the prospect of great sacrifice for the sake of the moral good, our beliefs regarding what is the case according to the moral fiction are likely to strike us as rather poor candidates for guiding action. 'Well, according to the moral fiction, I ought to φ', one might say, 'but the moral fiction is a bunch of baloney! I need not act as it prescribes when my important personal projects are thrown into jeopardy'. It is very easy (and often wise) not to care about one's fictional beliefs, and not to afford them any role in deliberation when the practical stakes are high. Again, a belief that *p** is not something that we are apt to *dispense with* at our choosing. But it is nonetheless something that we are apt to cast to the side when our interests would be better served by allowing our beliefs about reality to guide our behaviour.

There is a worry that needs to be addressed before proceeding. One might think that my arguments here contain the seeds of their own undermining. Earlier, I claimed that unlike a *make-belief* that *p*, a belief that *p* is *not* typically dispensed with when the practical stakes are high. In the preceding discussion, I have argued that a belief that *p* is* typically cast to the side when the practical stakes are high. This suggests that beliefs, though not dispensable, are nonetheless cast-to-the-side-able. But if beliefs are cast-to-the-side-able, then what practical advantage can our error-ridden moral beliefs possibly claim over make-beliefs? This worry is to be anticipated. However, it misconstrues the ambitions of my argument. My intention has been to argue that *a belief that p** will typically be cast to the side when the practical stakes go up. The claim is not that *beliefs* are typically cast to the side under such circumstances.

Yet what explains this difference? Does a belief that *p** constitute a distinct and special kind of belief? That seems unlikely. What is more likely is that a belief that *p** has a distinct and special kind of content—it is a belief about the content of a fiction. The reason that a belief that *p** will typically be cast to the side when the practical stakes go up, then, is not because it is a belief, and that beliefs generally have this property. The cast-to-the-side-ability of a belief that *p** is rather explained by the fact that it is a belief about the content of a fiction. Since acting upon fictional beliefs is unlikely to be in our interests when the practical stakes are high, we tend in such circumstances to act upon our beliefs about reality instead.

There is a second problem that I now want to raise for fictional beliefs. Like a make-belief that *p*, a belief that *p** also seems likely to integrate with our behaviour and the rest of our psychology in different ways than a belief that *p*. Compare the belief that the surrounding woods are inhabited by hungry grizzly bears (*b*) with the belief that *according to the fiction*, the surrounding woods are inhabited by hungry grizzly bears (*b**). These beliefs would no doubt prompt distinct emotional responses and behaviour. One who believes that *b* would fear the bears nearby, make a run for the hills, and call out for help. One who believes that *b** is unlikely to fear any nearby bears, and would be careful not to alarm any strangers in the vicinity.

The integration problem for the preface fictionalist would therefore seem to apply to the prefix fictionalist as well. The *belief* that an individual acted wrongly in breaking their promise is likely to inspire indignation and reproach. It is far from obvious that the same could be expected of a belief that *according to the moral fiction*, an individual acted wrongly in breaking their promise. This is not to deny that our beliefs about what is true according to a fiction can prompt *somewhat similar* emotional responses and behaviour as our beliefs about reality. After coming to believe that according to Shakespeare's *Hamlet*, Claudius betrays his brother, I may be angered by his callousness, and resolve to treat my own family better in future. But that is a far cry from beliefs that *p** prompting the *same*, or *very similar* kinds of behaviour as beliefs that *p*.

No one goes looking for Claudius to mete out punishment, or makes any efforts to tarnish his reputation.

NRW anticipate quite a few of these challenges. "Why", they ask, "might people change their preferences about how to act on the grounds of what is true in a story?" (2005, p.313). One solution that they offer is to recommend that the fictionalist

> … connect nonmoral preferences and what is true in the fiction via internalist bridge laws (though care must be taken in stating these). If the fiction is set up in such a way that it is guaranteed by the rules about the fiction's content that something is counted as good in the story only if, in fact, people have the appropriate non-cognitive attitudes towards it, then coming to realize that some course of action does have certain moral properties according to the story should prompt the realization that the action is one that the agent has certain attitudes towards. (2005, p.313)

These internalist bridge laws answer to the integration problem; for they construct a link between the moral fiction and motivation.[188] They also go some way towards addressing the overridability problem. If my fictional moral obligations track what I am *already* motivated to do, then it's easy to see how I might tend to aim for things that I believe are (fictionally) good and avoid things I believe are (fictionally) bad, even when the stakes are high; if I hate violence, I'm likely to avoid it even when violence pays.

But these internalist bridge laws also seem antithetical to the purpose of the moral fiction. According to NRW, the moral fiction earns its practical keep by providing a shared framework with which to reason our way through practical disputes, and coordinate our behaviour. But if the content of the moral fiction is determined by internalist bridge laws, then that content will surely differ from person to person, different people approving of different things. The moral fiction (or, perhaps we should say, moral *fictions*), so understood, becomes little more than a guise for personal interest—it is far from representing the more familiar, impartial moral framework that serves as a backdrop to reasoned discussion and compromise.

[188] The resultant fictionalist proposal therefore comes to resemble (to some degree) a revolutionary variety of Kalderon's (2005) *hermeneutic* moral fictionalism, which takes fictionalist attitudes to be more desire-like than belief-like.

There may also be conceptual obstacles here. The internalist bridge laws allow the content of the moral fiction to be determined by each individual's non-cognitive attitudes; something is morally right for *a* just in case *a* approves of it, wrong just in case she disapproves. Yet people can lend their seal of approval to all sorts of things, many of which we would be loath to label 'morally right'. The phenomenon of imaginative resistance suggests that we would find it difficult to comply with prescriptions to imagine such deviant moral judgments. (Perhaps it is something like this problem that NRW have in mind when they note that "care must be taken" in stating the relevant internalist bridge laws.) Unless more is said to constrain the sorts of connections that can be drawn between each individual's patterns of (dis)approval and what is morally wrong or right, a moral fiction constructed with internalist bridge laws may not be user-friendly.

It therefore seems to me that prefix fictionalists are not completely immune from the worries that spelled trouble for their preface counterparts. Both are likely to have trouble making good on their practical promise. Fictionalist attitudes do not seem like the right sorts of attitudes to underwrite a useful moral practice.

§5.5 MORAL FICTIONALISM: VERDICT AND LESSONS

On the face of it, moral fictionalism seems like a promising option. But I have suggested that the proposal has less going for it than first appearances suggest. At this stage, I would not go so far as to say that moral fictionalism is the *wrong* answer to the WNQ for moral discourse. That would be premature. I have shown that moral fictionalism has some shortcomings. Yet it may be that a fictional moral discourse is the best that we error theorists can hope for. Since abolitionism and revisionism don't inspire much confidence, perhaps fictionalism is our best bet, and we must resign ourselves to the fact that we can no longer enjoy many of the practical benefits of our error-ridden moral practices.

But we need not adopt this bleak perspective just yet. As I will argue in the following chapter, the conservationist option is very promising, and is capable of recouping many of the fictionalist's losses. Before proceeding to evaluate and defend conservationism, though, it will be useful to reflect upon the implications of what has been said this chapter.

Why should fictionalism strike us an unsatisfying response to the WNQ for moral discourse? A large part of the answer, I think, pertains to the important connection between moral judgment and action. Morality is not merely a theoretical enterprise; we take our moral judgments to have important implications for *what we do*, and this is, in large part, why moral practice is so useful to us. As was noted early on (§2.3.2), our subjective concerns and consequent motivations and behaviour co-vary fairly closely with our moral beliefs. In this chapter, I have elaborated upon this basic idea by suggesting that our moral beliefs integrate with our behaviour and the rest of our psychology in fairly reliable and characteristic ways. Since fictionalist attitudes have a very different psychological and behavioural profile to belief, they are likely to integrate *very differently* with our behaviour and the rest of our psychology. It was for this reason that substituting fictionalist attitudes for moral beliefs risked severing the important ties between moral judgments and behaviour.

This suggests that some fictionalist projects will fare better than others. None of what has been said should suggest to us that mathematical fictionalism is unpromising. Mathematical discourse is not tied to action (at least not to nearly the same extent) as moral discourse is.[189] However, my arguments here do suggest a rather grim prognosis for other fictionalist proposals. Some have suggested that an atheist can enjoy the benefits of religious belief by preserving religious discourse in the spirit of a useful fiction (see Eshleman 2005). But the considerations raised in this chapter should suggest to us that religious fictionalists will have trouble making good on this practical promise. This is because religious discourse *is* intimately tied to action; religious practices structure people's relationships to one another, and shape how they choose to lead their lives. Religious beliefs are also associated with a number of psychological and behavioural dispositions—the disposition to make costly sacrifices in the form of abstinence or chastity, for example. But these dispositions are likely to differ when an agent's attitudes towards the relevant propositions are fictional ones; calls to abstinence or chastity are apt to be overridden when they conflict with the

[189] There may be some exceptions, of course. Those who regard 13 as an unlucky number may avoid renting a house numbered 13. And the Pythagoreans acted in peculiar ways on account of their mathematical beliefs. But I feel that these exceptions are sufficiently rare so as not to render the general claim above implausible.

religious fictionalist's other interests. Given this, fictionalists attitudes may very well be unfitting attitudes to take towards religious belief systems as well.

Chapter Six

Conservationism: Offense

None of the proposals canvassed so far provide us with an entirely satisfying answer to our WNQ. But, worry not: I have saved the best for last. I shall argue that conservationism is our most promising option. Like fictionalists and revisionists, conservationists take our error-ridden moral discourse to be incredibly useful. But they part ways from their rivals in recommending that we *hold onto* our false moral beliefs.

The case for conservationism will be developed across two chapters. My aims in this chapter are largely offensive. Here, I will put the conservationist's best foot forward, developing what I take to be the most promising variant of the proposal. Responding to challenges and appeasing potential concerns with the view are defensive tasks that I defer to chapter 7.

I will begin by saying more about what conservationism involves and what motivates it (§6.1). I will then assess the most developed variety of conservationism—that of Olson (2014). Though promising, Olson's conservationism has a number of shortcomings. It is not obvious that we could believe both (first-order) moral propositions and the moral error theory, still less how we could keep this up for very long. Yet Olson does not do very much to address these problems. And to my mind, he does not do quite enough to quell the concerns that the conservationist's doxastic policy (that of intentionally preserving false beliefs) is apt to raise either.

In §6.2, I take these lessons on board and propose some desiderata for a workable conservationist proposal. The goal in the remainder of the chapter will then be to develop my own brand of conservationism, which preserves the benefits of Olson's approach while avoiding its associated problems. I first do some work to make sense of the idea that we could believe both moral propositions and the moral error theory (§6.3). Building upon some recent work in the philosophy of mind, I then explain why we should expect beliefs in moral propositions to be active in some contexts, and the belief in moral error theory to be active in others (§6.4). Finally, I argue that the intentional cultivation of false beliefs, though not generally advisable, can be justified in this particular case (§6.5).

§6.1 What is conservationism?

Mackie's *Ethics* makes for a rather peculiar narrative. The argument for moral error theory occupies only the first portion of the work. For the remainder of the text, Mackie can be found making quite a number of *first-order* moral claims. Understandably, this move strikes many as strange; "it is…", James Lenman remarks, "…as if someone were to write a book where, in Part I, she argued that astrology is all the rankest, most hopeless nonsense, only to go on, in Part II, to argue that you can never trust Librans" (2013, p.399). But perhaps Mackie just thought that he could have his cake and eat it too; that he could believe his moral error theory while continuing to use moral discourse in much the same way as before. Perhaps he was a *conservationist.*[190]

The conservationist recommends the preservation of ordinary (error-ridden) moral discourse. But she does not merely advise us to believe and assert moral propositions. The conservationist wishes for us to preserve *moral practice* in its entirety. She advises us to continue to make moral judgments about the behaviour and character of others, to promote certain social policies on moral grounds, to dole out praise and blame, and to invoke moral considerations when deliberating about what we ought to do. In effect, the conservationist recommends a 'business as usual' approach to life after moral error theory; we are to go about our days more or less as we did before.

I say *more or less* as we did before; for none of this involves conveniently forgetting or wilfully suppressing the reality of moral error theory. The conservationist advises us to believe the error theory as well. The relevant advice is not necessarily to believe the moral error theory and (first-order) moral propositions *simultaneously*. (Doing so

[190] This diagnosis is suggested by Oddie and Demetriou (2007, p.486), and West (2010, p.184). Others interpret Mackie as a fictionalist (Garner 1993, p.87; Joyce 2005, pp.296-8; Kroon 2011, p.790) or a revisionisist (Blackburn 1993, p.150; Schroeder 2007b, fn.24; Lenman 2013, p.399; Lutz 2014, p.354).

may very well be impossible for creatures like us.[191]) To borrow a phrase from West (2010, p.195) the conservationist might "distinguish between belief in theory and belief in practice", and argue that the belief in the moral error theory will be guide some aspects of life, but not others.

Note that conservationists need not assume that we have *direct* voluntary control over what we believe. They usually recommend *indirect* methods for cultivating moral beliefs. In this respect, conservationism is somewhat reminiscent of Pascal's (1660) approach to religious faith (Olson 2014, p.191). Though Pascal took belief in God to be justified on pragmatic grounds, he did not assume that people could simply inject that belief into their mental economy. Instead, he advised taking steps that were likely to lead to the acquisition of religious belief—prayer or attending church, for example.[192] (Another potential parallel is Kant (1781), who thought that we could have no knowledge of phenomena outside our experience, but took belief in an afterlife to be a demand of practical reason.)

If workable, conservationism seems likely to be far more effective than its rivals at securing the practical goods of moral practice. Moral beliefs plausibly have a better chance of preserving these practical goods than the feeble attitudes that the fictionalist recommends, or the revisionist's schmoral beliefs. Presumably, the best way to secure the benefits of moral belief is *to be* a moral believer. Holding onto our false moral beliefs might also be easier than keeping up a pretence or eliminating them altogether. If this is right, then conservationism might present us with a more feasible and stable option going forward than its rivals.

[191] Some caution is warranted here. It seems perfectly conceivable that an agent could simultaneously believe both that (i) *f* is an elliptical equation, and (ii) that *f* is not correlated with modular form, even though this would be inconsistent. (It took Andrew Wiles quite a bit of work to discover this inconsistency, after all.) And dialethiests think we are capable of believing contradictions (see Priest 1985-6). Such cases notwithstanding, it does seem that simultaneously believing that *p* and that ¬*p* will often be rather difficult; it is not obviously possible that an agent could simultaneously believe both that there is and that there is not a table in front of her, for example. (See Alston 1989, p.122; Pettit and Smith 1996, p.448.)

[192] There are disanalogies here as well, of course. Whereas Pascal sought to justify believing a proposition that he took to be insufficiently supported by our evidence, our conservationist seeks to justify believing propositions that we take to be false.

Admittedly, there is one respect in which conservationism loses out against all of its rivals: it recommends false belief! Not only that, but it recommends *inconsistent* beliefs. The conservationist certainly goes against the philosophical grain in advising us to hold onto beliefs that we know to be false. The advice to believe both that *p and* that $\neg p$ is apt to strike us as more bizarre still. As I have noted (§2.2.4), there is plausibly a presumption against having false beliefs. If the conservationist's proposal is to be attractive, then she must do some work to motivate overriding it.

In what follows, I consider Olson's (2014) variety of conservationism (§6.1.1), and draw attention to where I feel there is room for improvement (§6.1.2). As we shall see, these challenges are only challenges for Olson's implementation. They need not spell trouble for conservationism more generally. The task for the remainder of the chapter will be to develop a debugged conservationism—one that is designed to avoid these problems.

§6.1.1 *Olson's conservationism*

As we saw in chapter 5, Olson is not optimistic about the prospects of moral fictionalism. He regards conservationism as a more fitting solution to the WNQ for moral discourse. The distinctive benefits of moral discourse—enabling us to "coexist peacefully, to prevent conflicts, to regulate and co-ordinate behaviour, and to counteract limited sympathies" (2014, p.197)—are, he thinks, better secured by preserving false beliefs in moral propositions.

Indeed, Olson advises the error theorist to believe both the error theory *and* moral propositions. More specifically, he "…recommends moral belief in morally engaged and everyday contexts and reserves attendance to the belief that moral error theory is true to detached and critical contexts" (2014, p.192). In this respect, Olson's conservationism is structurally similar to Joyce's fictionalism, which also recommends attending to the belief that the moral error theory is true in critical contexts (p.191).[193]

[193] Olson's conservationist might therefore seem vulnerable to the very same instability worries that plagued the fictionalist; she could simply raise the epistemic stakes whenever doing so were to her immediate benefit. I will not pursue this objection here. But I will explain why *my* conservationist is less vulnerable to this problem (§7.5).

Olson does some work to make sense of this inconsistent believing, arguing that it is quite common for agents to occurrently believe that *p* in certain contexts even when they are generally disposed to believe that ¬*p*. This seems plausible. A doting parent will often declare that *their* child is the sweetest or cleverest of all children. And they're not necessary being deceitful when they do so. But when confronted in a serious tone with the question, 'do you *really* believe that Jimmy is the smartest kid in the school, the country, the world?', they will often disavow their earlier claims ('of course not!'). Still, it seems wrong to say that they don't believe that their child is the greatest in the former context. Such examples are by no means hard to come by. As Olson notes,

> … it is a psychologically familiar fact that we sometimes temporarily believe things we, in more reflective and detached contexts, are disposed to disbelieve. In such cases, the more reflective beliefs are suppressed or not attended to, due to e.g. emotional engagement, affection, peer pressure, or a combination of these factors. For instance, someone might say truly the following about a cunning politician: 'I knew she was lying, but hearing her speech and the audience's reaction last week, I really believed what she said'. (2014, p.192)

This is thought to suggest a promising prognosis for conservationism. Just as we can be seduced into (occurrently) believing the false claims of charismatic politicians, so too (perhaps) can we be sucked into believing that certain actions are right and wrong—especially when these actions engage our emotions, as they are wont to do. Olson predicts that these affective responses will do some work in helping to silence the belief that the moral error theory is true in non-critical contexts.

Olson also makes some effort to motivate committing the cardinal sin of intentionally preserving false beliefs. As he notes, there is experimental evidence that certain kinds of false beliefs are essential to everyday functioning. Psychologists have proposed that "overly positive self-evaluations, exaggerated perceptions of controls and mastery, and unrealistic optimism" are not only typical, but important for promoting mental health, feelings of contentment, and creativity (Taylor & Brown 1988, p.193 in Olson 2014, p.185). Given that there are many examples of false but instrumentally valuable non-moral beliefs, Olson thinks "… it is highly plausible that there are false, but instrumentally valuable, moral beliefs too" (2014, p.185).

§6.1.2 *Some gaps in Olson's proposal*

Though Olson's conservationism has much to recommend it, I feel that there is much more work to be done. We can do more, for example, to motivate the supposition that the conservationist's attitudes towards moral propositions would be *beliefs*. One could simply stipulate this away, of course. But that would be to secure the benefits of theft over honest toil. Though agents can no doubt have inconsistent beliefs, a principled case needs to be offered for thinking that a moral error theorist could believe moral propositions. Unless more is said, one may suspect that she is only *accepting* them, or *pretending* to believe them—in which case conservationism is in danger of representing a mere terminological variant of fictionalism (see Jaquet & Naar 2016).

Moreover, Olson doesn't do quite enough to motivate the supposition that the moral error theorist will attend to her moral beliefs in "…morally engaged and everyday contexts" and only attend "to the belief that moral error theory is true to detached and critical contexts" (2014, pp.192-3). To be sure, he does suggest that affective attitudes will "silence" the belief that the moral error theory is true in morally engaged contexts. Yet this seems a dangerous path to steer. Theoretical debates in applied ethics aren't always marked by heated emotion. But we shouldn't want the belief that the moral error theory is true to make an appearance in these contexts either. So it seems that more work needs to be done on this front; further reasons must be provided for expecting that we are likely to attend to our moral beliefs in some contexts, and to our belief that the moral error theory is true in others. (This isn't to deny that affective responses will be a large part of the story—I will have more to say about them in §7.4. My point is merely that they shouldn't be the *only* part of the story.)

A related worry is that it's not clear how the conservationist's inconsistent beliefs could play the roles that beliefs usually play in guiding our behaviour. We usually act in ways that would satisfy our desires if our beliefs were true. It seems that it would be rather difficult to navigate a happy path through life believing *both* that p and that ¬p. (How should you act to satisfy your desires if you believe both that cheating at cards is wrong and that it isn't?) Given this, one might expect that consistency will, as a matter of practical necessity, eventually be restored.

Finally, Olson doesn't do quite enough to justify preserving our false moral beliefs. Though he does gesture towards some cases in which the preservation of false beliefs can be defended, this does not seem quite enough to defend the preservation of false beliefs *in this case*. Without these details filled in, one may worry that the conservationist's recommendation emanates from an awfully fickle doxastic policy; one which permits us to cultivate whatever beliefs are to our fancy. Such a policy is unlikely to arouse much enthusiasm.

§6.2 A NEW CONSERVATIONISM

The preceding discussion has been instructive. We are now in a position to identify some important desiderata for a workable conservationist proposal. If her proposal is to be attractive, then the conservationist must:

> **D1.** Motivate the claim that the moral error theorist is properly characterised as *believing* both moral propositions and the moral error theory.
>
> **D2.** Motivate the expectation that the moral error theorist will attend to her beliefs in moral propositions in some contexts and to her belief that the moral error theory is true in others.
>
> **D3.** Explain how the moral error theorist's inconsistent beliefs could play the roles that beliefs typically play in guiding behaviour.
>
> **D4.** Offer a suitable justification for overriding the presumption against intentionally preserving false beliefs.

D1 requires that the conservationist offer a principled basis for taking the moral error theorist's attitudes towards both (first-order) moral propositions and the moral error theory to be beliefs. Doing so will be my first task in what follows. In §6.3, I will argue that the moral error theorist could plausibly be said to believe moral propositions if she reliably matched the *dispositional stereotype* for belief with respect to those propositions in certain contexts.

Even if the conservationist convinces us that a moral error theorist could have these inconsistent beliefs, she still owes as an explanation of the unusual doxastic back-and-forth that is to be expected of her. **D2** invites the conservationist to explain in virtue of *what* the moral error theorist could be expected to attend to her beliefs in moral propositions in some contexts, and to her belief that the moral error theory is true in others. Providing such an explanation will be my goal in §6.4. Here, I will begin by arguing that in general, only *some* of an agent's total set of beliefs are active (i.e., available for use in reasoning and causally efficacious *qua* guides of behaviour) at any moment. (This idea also helps us to satisfy **D3**; as I shall explain, it suggests that we should take agents to act in ways that would satisfy their *active* desires if their *active* beliefs were true.) I will then propose that belief activation is determined, in large part, by the sorts of information that is salient to an agent in a particular context. Given that (i) *very different* sorts of information is salient in more critical contexts, as opposed to contexts where we debate first-order moral issues, and (ii) the information that is salient in the latter context is unlikely to 'activate' the belief in moral error theory, we should not expect the conservationist's belief that the moral error theory is true to be active in morally engaged contexts.

Finally, the conservationist does better if her proposal appeals to our special interest in preserving our false moral beliefs, rather than the general truism that false beliefs can sometimes be instrumentally valuable (**D4**). The task for §6.5 will therefore be to develop a case for preserving false moral beliefs in particular. I shall argue that we have a special interest in holding onto these beliefs on account of their being of paramount importance to interpersonal co-operation.

The positive case to be sketched in the remainder of chapter 6 will leave some questions unanswered. I will provide a principled basis for taking the conservationist's attitudes towards both the moral error theory and first-order moral propositions to be beliefs in §6.3. But a full defence of that claim would also involve ruling out other contenders, and explaining how this picture is consistent with important properties commonly associated with belief. In §6.4, I will motivate the claim that the conservationist could be expected to attend to her belief in the moral error theory only in certain contexts. But developing a more convincing case for this claim would require saying more about the extent to which we have *control* over what we believe. A justification for preserving our false moral beliefs will be offered in §6.5. But a complete justification would consider the potential costs of enacting conservationism—not only the benefits. Rest assured, these gaps will be filled in chapter 7, in the course of defending conservationism against some important challenges. It will be helpful to have a positive view on the table before switching to the defensive.

§6.3 Believing that p and that ¬p

My goal in this section will be to make sense of the idea that a moral error theorist could believe moral propositions as well as the moral error theory. I will begin by sketching the conditions under which an agent can plausibly be taken to hold inconsistent beliefs. I will then argue that these conditions could conceivably obtain for a moral error theorist. Keep in mind that the task here is purely offensive; my goal is to provide a principled case for understanding the moral error theorist's attitudes towards both moral propositions and the moral error theory to be beliefs. Ruling out other contenders and appeasing some concerns with this claim are defensive tasks that I defer to §7.1.

If we are to make the case that the conservationist's recommended attitudes are beliefs, then there is a question that cannot be avoided: what is it to believe that *p*? Answering this question is a delicate matter. It would seem wise to be as non-committal as the argumentative strategy allows. (I do not wish to expose my conservationist to the charge that her recommended attitudes only count as beliefs given a *manifestly implausible* account of belief.) However, taking a stand will be essential if I am to make my case.

Let me begin by declaring my own allegiances: it strikes me as highly plausible that some form of functionalism about belief is correct. Whatever beliefs are, they can plausibly be characterised in terms of their functional roles. Unfortunately, this is not yet of much help; for what kind of functionalism are we to assume? I suspect that behaviourism can swiftly be ruled out without any tears. But this still leaves us with a number of contenders; psycho-functionalism, analytic functionalism, dispositionalism, interpretivism, and what have you.[194]

Much of what I have to say will be consistent with analytic functionalism, dispositionalism, and interpretivism; the arguments in what follows could assume any one of them. But given that dispositionalism is already well-domesticated within this thesis

[194] 'Interpretivism' is a label associated with a number of different views. I have in mind here the view according to which facts about belief are just facts about idealised interpretation; for an agent to believe that *p* just is for the best (i.e., most charitable and rationalising) interpretation of her to attribute the belief that *p* to her. Interpretivism (so construed) and analytic functionalism are not mutually exclusive. Lewis (1966, 1974) was arguably a supporter of both views.

(recall chapter 5), and strikes me as independently plausible, my dialectical choice will be to use a dispositionalist account of belief as a concrete approach with which to frame the discussion.

As we learned in §5.4.1, dispositionalists are functionalists in that they propose to characterise beliefs in terms of their functional roles. The relevant roles here, being dispositions, are forward-looking in that they tell us what kinds of states a belief that *p* typically brings about. Ryle, an early dispositionalist, offers the example of believing that ice is dangerously thin. To believe this is

> … to be unhesitant in telling oneself and others that it is thin, in acquiescing in other people's assertions to that effect, in objecting to statements to the contrary, in drawing consequences from the original proposition, and so forth. But it is also to be prone to skate warily, to shudder, to dwell in imagination on possible disasters and to warn other skaters. It is a propensity not only to make certain theoretical moves but also to make certain executive and imaginative moves as well as to have certain feelings. (1949, pp.134-5)

Following Eric Schwitzgebel (2001, p.81), we can say of an agent who has such dispositions that she matches the "dispositional stereotype" for the belief in question (i.e., the belief that ice is dangerously thin); the relevant dispositions have the unique "cluster of properties" that we associate with belief. One who believes that *p* will, for instance, typically be disposed to assent to *q* if shown that *p* implies *q*, to express surprise upon hearing that not *p*, and so on. An agent who has these and other relevant psychological and behavioural dispositions can be said to believe that *p*.

An important caveat is needed here. Psychological and behavioural dispositions by themselves will not suffice for a *complete* specification of the conditions under which an agent believes that *p*. Someone who believes that *p* won't be disposed to express surprise upon hearing that $\neg p$ if she strongly desires to avoid any outward expressions of emotion. (See Chisholm 1957.) Schwitzgebel (2002) suggests that the dispositionalist can circumvent such problems by (i) associating beliefs with particular cognitive and phenomenal dispositions as well as behavioural dispositions, and (ii) specifying that the dispositions alluded to in the stereotype come attached with a *ceteribus paribus* clause. He proposes to understand

> …the dispositional characterisations as loaded with tacit "if" clauses. Not literally *all* else must be equal—but certain conditions must hold. Joe [who believes that there is beer in his fridge] is disposed to assent to assent to ut-

> terances meaning that there is beer in his fridge *if* he hears the utterance, *if* he has decided not to lie about or evade the matter, *if* he understands the language in which the utterances take place, *if* he has the physical capacity to indicate assent, and so forth. (2002, p.253)

Given this, my discussion of the dispositional stereotypes associated with *moral* beliefs in what follows should be understood as containing tacit 'if' clauses. Someone who believes that the animal fur industry is wrong, for example, will be disposed to assert that the animal fur industry is wrong *if* she has not decided to deceive others on the matter, *if* she has the physical capacities needed to express herself vocally, and so on.

As is to be expected in philosophy, not all will be particularly taken with this account of belief. Some may think that these claims about the states that a belief that *p* typically brings about are true enough, but only contingently so—such claims are not claims upon which we ought to rest an analysis or definition. To the extent that a person who believes that *p* will typically be disposed to assent to a sentence that means that *p*, this is because people typically have total sets of beliefs and desires such that, in typical circumstances, it is optimal for them to assent to many of the things that they believe.[195]

I cannot hope to mount a sustained defence of dispositionalism here. (Picking one's battles is a necessity in any work of limited length.) But the reader is of course free to take my conclusions in what follows to be conditional upon its truth. (Or, better: upon the truth of dispositionalism *or* some other varieties of functionalism; as I have noted, my arguments in what follows could also proceed by assuming the truth of other functionalist views.) Alternatively, a reader may wish to take my claims about agents matching particular dispositional stereotypes as *evidential claims*. Even if someone's matching the dispositional stereotype for belief with respect to some proposition *p* does not make it analytically true that she believes that *p*, it may nonetheless provide us very good evidence that that she believes that *p*.

With those qualifications in mind, we can now proceed to make sense of a moral error theorist believing moral propositions as well as the moral error theory. Dispositionalism suggests a natural strategy for doing so: we can say that the moral error

[195] I am grateful to Edward Elliott for helpful discussions on this point.

theorist could match the dispositional stereotype for belief with respect to moral propositions in some contexts, but not others. Here's the general idea. In one set of contexts C_1, the error theorist will match the dispositional stereotype for belief with respect to some moral proposition *p*—say, 'cheating is wrong'. She will, for instance, be disposed to assent to the proposition 'attempted cheating is wrong', if shown that 'cheating is wrong' implies that 'attempted cheating is wrong', to express surprise upon hearing that it is not the case that cheating is wrong, to oppose cheating, and so on. In another set of contexts C_2, however, she will match the dispositional stereotype for belief with respect to the proposition that $\neg p$. She will, for example, be disposed to deny that cheating really is wrong, to express no surprise whatsoever upon hearing that it is not the case that cheating is wrong, and so on.[196]

Admittedly, there is room for dissent here. One might insist that only reflective and critical contexts matter for the purposes of determining what an agent *really* believes. Joyce proposes that if an agent has at some point "adopted a critical perspective and therein sincerely denied T, and remains disposed to deny T were he again to adopt that perspective", then "…he disbelieves T, regardless of how he may think, act, and speak in less critical perspectives" (p.193).

However, I see no strong justification for always privileging this particular disposition—the disposition to assert that *p* in critical contexts—over all others. There is no one disposition to rule them all. Beliefs plausibly supervene upon a wide range of dispositions—not just the disposition to affirm that such-and-such is the case in the philosophy seminar (Braddon-Mitchell 2006, p.850). If an agent reliably matches the dispositional stereotype for belief that *p* in C_1, and reliably matches the dispositional stereotype for belief that $\neg p$ in C_2, then it seems much more natural to say of her that she believes both that *p* and that $\neg p$.

[196] As I have suggested, this argumentative strategy may be compatible with other varieties of functionalism as well. An interpretivist could claim that the best interpretation of the error theorist's behaviour would attribute these inconsistent beliefs to her. Likewise, an analytic functionalist could argue that the attitudes in question play the belief-role. Admittedly, enlisting these views would require us to say something more about how such attitudes could conceivably play the *backward-looking* roles of belief; for example, how they could exhibit an appropriate responsiveness to one's history of evidence. I do so later on when defending the claim that an error theorist's beliefs in moral propositions would be responsive to a *selective* history of evidence (§7.1).

Indeed, such inconsistency does not seem at all uncommon. The attribution of inconsistent beliefs should sometimes strike us as appropriate. Schwitzgebel offers the example of an inconsistent atheist:

> In certain moods and in certain contexts, Antonio feels quite sure that the universe is guided by a benevolent deity. In other moods and contexts, he finds himself inclined to think of talk about God as 'a beautiful metaphor' or even, sometimes, 'a crock of hooey'. When his atheistic buddies at work mock religious belief, he does not join in, but neither does he feel an impulse to defend belief in God; at such moments, especially if it is mid-week, the whole God business seems rather silly. When Antonio goes to church with his wife, he is not inclined to believe everything the pastor says, but, particularly if the pastor waxes poetic about the magnificence of creation, he may feel that there must be a divine force guiding the world. At the birth of a child or the death of a friend, he feels certain God is involved… (2001, p.78)

It may be tempting to insist that Antonio doesn't *really* believe that God exists (or that he isn't *really* an atheist). But neither answer seems to do complete justice to Antonio's psychological and behavioural dispositions. Given that Antonio has the dispositions described above, it is very plausible to say that he believes *both* that there is a God and that there isn't.[197]

Accordingly, it seems to me that we sometimes have good grounds for attributing to an agent inconsistent beliefs; we have decent grounds to do so when she reliably matches the dispositional stereotype for belief with respect to p in some contexts, and reliably matches the dispositional stereotype for belief with respect to $\neg p$ in others. I am yet to specify what these contexts are (that is a task for the following section). Likewise, I have not yet ruled out other candidates for the psychological attitudes in question (that is a job for §7.1). The present task is purely offensive. My claim is that if a moral error theorist satisfied these conditions (as the conservationist

[197] This is not *quite* the conclusion that Schwitzgebel himself reaches. He prefers to think of these cases as instances in which an agent is *in-between* believing that p and that $\neg p$. But for my part, I don't think that this is the best way to make sense of the phenomenon to which he draws attention. The phenomenon seems to me to involve genuine beliefs—albeit ones that are only active in certain contexts. I will motivate this way of seeing things in §6.4. This is not to say that I don't think there *can* be cases of in-between believing. As Schwitzgebel notes elsewhere (2012), delusions may be one such case.

predicts she will—more on this to follow) then there is a principled case to be made for taking her to believe both moral propositions *and* the moral error theory.

§6.4 BELIEFS AND CONTEXTS

I have argued that an agent can plausibly be said to believe both that *p* and that ¬*p* if she reliably matches the dispositional stereotype for believing that *p* in context C_1, and reliably matches the dispositional stereotype for believing that ¬*p* in context C_2. But it still not clear *why* we should expect an agent to attend to these beliefs only in certain contexts, nor how she could be expected to keep up such inconsistency over time. How could a mind divided against itself continue to stand? Quite easily, I think. Indeed, I will now argue that we are *already* living with divided minds. My case is premised upon the following, plausible idea:

> **The Activated Belief Hypothesis (ABH)**
> If T = {B1, B2, B3, …} is a person's total set of beliefs at a time, then in many cases, only a subset of those beliefs will be "activated"; that is, available for reasoning and causally efficacious *qua* guide of behaviour.

My goal in the remainder of §6.4 will be to motivate the ABH, and to explain how it helps the conservationist to satisfy both **D2** and **D3**.

The ABH does run contrary to the idea that our total set of beliefs guides our actions all of the time. But this idea seems to conflict with what we observe in any case. As Andy Egan notes, "…actual people have inconsistent beliefs, display failures of closure, and often fail to bring to bear some of the things that they believe on particular decisions" (2008, p.48). It is precisely these phenomena that motivate the ABH.[198]

[198] Egan in fact appeals to these phenomena in service of the hypothesis that our beliefs are *fragmented*; it is his contention that "…we have a number of distinct, compartmentalized systems of belief, different ones of which drive different aspects of our behavior in different contexts" (2008, p.48). However, the term 'fragmentation' is traditionally associated with a response given by Lewis (1982) and Stalnaker (1984) to the problem of logical omniscience as it arises for their very specific on the nature of belief and mental content. My arguments here do not rest upon the truth of the Lewis-Stalnaker view of belief, nor upon the idea that individual fragments of belief systems are closed under logical implication. So I will frame things slightly differently to Egan.

Let me briefly attend to some of these motivations in more detail before proceeding to explain how the ABH accounts for inconsistent believing, which interests me most. First, people often exhibit *failures of closure*; they seem to (i) believe that *p*, (ii) believe that if *p* then *q*, but (iii) don't seem to believe that *q*. I may, for example, (i) believe that today is Tuesday, (ii) believe that if today is Tuesday, then I ought to take out the rubbish, but (iii) fail to believe that I ought to take out the rubbish today. This is nicely accounted for by the ABH. Sometimes (on a Tuesday, one would hope), my belief that 'today is Tuesday' is activated. Other-times, my belief that 'if today is Tuesday, then I ought to take out the rubbish' is activated. But both may not be activated at the same time. (Hence the humorous cliché of a frazzled person chasing the rubbish collection truck down the road in her dressing gown.)

Second, we already recognise a distinction between recognition and recall, and a distinction between ignorance and a failure to bring particular information to bear upon a particular task. Adam Elga and Agustín Rayo consider the familiar titbit of forgetting a neighbour's name:

> Jack has a neighbor he sees only infrequently. The neighbor's name is "Beatrice Ogden", and she lives in apartment 23-H. If asked "What is the name of the person in 23-H?" Jack is disposed to groan, scratch his head, mutter "I know this, don't tell me..." but be unable to answer. But if instead asked "How do you know Beatrice Ogden?", Jack is disposed to immediately reply, "She's the person in 23-H. (2015, p.3)

The ABH nicely explains this phenomenon as well. If different beliefs are activated and guide our behaviours in different contexts, then they will be available for use in some contexts but not others.

We can now proceed to consider how the ABH makes sense of *inconsistent believing*. Recall Schwitzgebel's (2001) inconsistent atheist. We should want to say of Antonio that his beliefs play the following, action-guiding role; he is disposed to act in ways that would bring about his desires if his beliefs were true. (Such a claim is surely platitudinous of belief). We should also want to say that insight into Antonio's beliefs (and desires), would enable us to make more or less reliable predictions about his behaviour. But it's difficult to say either of these things if we assume that Antonio is *always* guided by his *total* set of beliefs, which includes both the belief that God exists and the belief that God does not exist. A perplexed Egan asks,

> Which of the actions available to me are the ones that would (tend to) bring about the satisfaction of my desires if P and not-P?... This sort of question is unlikely to lead to answers that will be of any help to us in attributing beliefs to one another. (2008, p.50)

If we are to preserve the idea that beliefs play an important role in guiding action, along with the idea that they enable reliable predictions of behaviour, then it helps to assume that not *all* of an agent's beliefs are activated at a particular time.[199] It helps to take Antonio, for example, to have inconsistent beliefs (the belief that God exists, and the belief that God does not exist) that are activated in different contexts.

Whereas a more unified picture of belief takes agents to be disposed to act in ways that would satisfy their desires if their beliefs were true, the ABH takes agents to be disposed to act in ways that would satisfy their *active* desires if their *active* beliefs were true (Egan 2008, p.52). In certain contexts, Antonio is disposed to act in ways that would bring about his desires if God existed (e.g., he attends church in a bid to avoid eternal damnation), and we would have a better shot at predicting his behaviour if we took him to be a religious believer. In other contexts, he is disposed to act in ways that would bring about his desires if God did not exist (e.g., he declares that religion is a "crock of hooey"). In these latter cases, taking him to be an atheist would help us to predict his behaviour more effectively.

An interesting question here concerns belief-activation; in virtue of what are some beliefs activated in certain contexts, but not others? Following Elga and Rayo (2015), (at least a large part of) the answer seems to concern both the information that is salient to an agent and the task(s) to which she is attending. When in church, the "magnificence of creation" is salient to Antonio, and he wants to feel part of something bigger than himself. So his belief that God exists is likely to be activated. When in the company of his atheist colleagues, however, the silliness of religious faith becomes especially salient to him, and he wants to join in on the fun. So his belief that God does not exist is more likely to be activated.

Let's take stock. When an agent is disposed to act in some contexts in ways that would be advisable (assuming certain facts about her desires) if *p*, and to act in other

[199] Parallel claims would of course need to be made about desires as well, but I will restrict my focus to beliefs here.

contexts in ways that would be advisable (assuming certain facts about her desires) if ¬*p*, we seem justified in attributing to her inconsistent beliefs. But if we are to preserve the plausible idea that beliefs play an important role in guiding-action and enabling predictions of behaviour, then it is useful to take such beliefs to be activated in different contexts. In what follows, I will suggest that this way of seeing things is of considerable help to the conservationist.

Before I do though, let me make a few quick clarifications. First, the ABH is not a story about the semantics of belief-attribution. The proposal is therefore to be distinguished from contextualist views, according to which the truth of belief-attribution sentences depends upon contexts of utterance. Given the ABH, the context-dependence of belief is a dependence of which beliefs *are active* upon an agent's context of action.

Second, nothing I have said should suggest to us that this is our *only* option for making sense of inconsistent believing. I have chosen the ABH, in large part, because I take it to be independently well-motivated—it can account for other phenomena as well. Given this, one should not think that the conservationist's way of making sense of inconsistent believing is unmotivated or objectionably *ad hoc*.

Finally, one may worry that we are, in attributing inconsistent beliefs to an agent, taking her to be mistaken about what she believes, or guilty of an objectionable species of incoherency—which seems woefully uncharitable. However, sometimes the attribution of inconsistent beliefs is the most charitable option available. It would, for example, arguably be more charitable to take Antonio to believe both that God exists and that God does not exist than to take him to believe that 'God' sometimes refers to a divine, all-powerful being and other-times refers to a planet between Earth and Mars. Insisting that he does not *really* believe that God exists may also amount to taking him to be grossly mistaken about what he believes; for example, when he is in attending church and is quite adamant that he is a *bona fide* believer.

Importantly for the conservationist, the ABH should suggest to us that it is not implausible that an agent could go about life holding inconsistent beliefs. If the ABH is true, then these inconsistent beliefs could be activated in different circumstances. Accordingly, it does not seem so far-fetched to think that we could be disposed to act in some contexts in ways that would satisfy our active desires were our active belief that (say) cheating is wrong true, and disposed to act in ways that would satisfy our active desires were our active belief in the moral error theory true in others.

But of course, we cannot simply *assume* that the chips will fall where the conservationist wants them to. We must motivate the idea that our beliefs in moral propositions *would be* disconnected from our meta-ethical beliefs—that they would be active in different contexts. (And it will be preferable to do so without falling back upon our affective experiences.) To some extent, this claim is already endorsed by a growing number of meta-ethicists. Quite a few have proposed that meta-ethical claims have no bearing whatsoever upon questions that arise in first-order ethics (Blackburn 1998; Dworkin 1996, 2011; Scanlon 2014).[200]

Dworkin offers a particularly vivid illustration of this line of thought, boldly declaring that "Value judgments are true, when they are true, not in virtue of any matching [of some metaphysical reality] but in virtue of the substantive case that can be made for them" (2011, p.11). This claim is rooted in the broader idea that we cannot occupy an external standpoint or "Archimedean position" with respect to ethics. In Dworkin's view, one cannot comment upon the status of moral propositions—declaring that they are objectively true, say—without committing oneself to a *first-order* moral position in turn. Indeed, Dworkin thinks that we can plausibly relegate *all* putatively meta-ethical statements to the status of first-order moral statements (1996, p.97; cf. Blackburn 1998, Fantl 2006). In effect, then, Dworkin's contention isn't merely that meta-ethics doesn't matter—it's that it doesn't *exist* (Bloomfield 2009).

I certainly do not endorse all of these claims. For one thing, I am inclined to think that meta-ethics exists, and that meta-ethical questions (*qua* second-order questions) make sense.[201] It does not seem confused to inquire into the nature of moral properties, nor incoherent to ask whether moral judgments are intrinsically motivating.

[200] The motivations for this claim are admittedly diverse. What is common to all of these theorists is that they deny (in some way or other) the existence or intelligibility of second-order questions. Scanlon (2014) thinks that all that is needed in order for reasons (and other posits of first-order normative discourse) to exist is (roughly) for talk of them to be licensed by standards *internal* to the normative domain. (He thinks the same is true for other domains as well, such as mathematics.) There is no further meta-normative or metaphysical question to be answered. Blackburn (1998, p.295) argues that putatively meta-ethical claims are best understood as first-order ethical ones.

[201] The points I raise here have been developed in detail by Bloomfield (2009), Shafer-Landau (2010), and Enoch (2011, ch.5).

These questions are not plausibly construed as questions about substantive moral issues. Nonetheless, they are obviously intelligible. Adopting an Archimedean vantage point therefore seems to me to be eminently possible. When we're 'looking down' (so to speak) *from* the meta-ethics classroom at the applied ethics classroom, we error theorists are perfectly entitled to believe that those engaged in first-order disputes are exchanging systematic falsehoods.

But here is where I *agree* with Dworkin and other anti-Archimedeans (as they are sometimes called): we rarely if ever do (or need to) *look up from* the applied ethics classroom at the meta-ethics classroom. There is rarely if ever any need to attend to moral linguistics or metaphysics in order to address pressing moral issues.[202] Questions having to do with the pragmatics of moral judgments or the multiple realisability of moral properties do not seem at all pertinent to the question regarding whether Euthanasia ought to be legalised.

There is something importantly right about the thought that day to day moralising typically proceeds without raising any questions about underlying metaphysical truths. We do not need to get on the phone to the meta-ethicists every time we want to address a pressing moral issue. If I am concerned to determine whether it is wrong to break an inconvenient promise, then we would expect my beliefs about the permissibility of betraying a friend to come to the fore—not my ontological beliefs about the property of permissibility. As Dworkin puts it, "The moral realm is the realm of argument not brute, raw fact" (2011, p.11). Within the context of everyday moral theorising, we tend to focus upon considerations that are pertinent to first-order moral disputes.[203]

[202] Admittedly, meta-ethics may be more easily disconnected from *applied ethics* than normative ethics, since debates in normative ethics may sometimes trade upon meta-ethical assumptions—the assumption that wrongness implies blameworthiness, say. But not *all* meta-ethical information is likely to be relevant in the normative ethics classroom. Facts about blameworthiness, for example, seem more likely to be pertinent than facts about moral metaphysics or categorical reasons.

[203] In §2.3.1, I claimed that I would ultimately resist the idea that the moral error theory carries worrying implications for first-order moral theorising. We can now see why: if our meta-ethical beliefs are activated in different contexts from our first-order moral beliefs, then the latter enjoy some degree of protection. But this claim is far from obvious—we had to *earn the right* to say as much. We couldn't simply take the ABH for granted at the outset.

(Of course, all of this is just to say that we don't ordinarily *take* claims about moral metaphysics to be relevant to settling first-order moral issues. It is a separate question whether it is epistemically improper for us do so (see Lang, 2011, p.22). Later (§7.1), I will tentatively suggest that contrastivist conceptions of justification may help the conservationist to restore epistemic propriety. But unfortunately, there won't be space to develop a systematic conservationist epistemology. In any event, we should not expect conservationism to cohere with *all* of our epistemic values (it recommends false beliefs, after all). The real question is the extent to which it does, and to what extent we should be willing to trade them in.)

In any event, the considerations above suggest to me that the chips have already fallen more or less where the conservationist wants them to: our meta-ethical beliefs and our first-order moral beliefs tend to be active in different contexts. We should expect our beliefs in moral propositions to be active when we are debating the moral permissibility of a particular social policy. But we should expect our belief that the moral error theory is true to be active when we are comparing different accounts of moral metaphysics.

To summarise, I have in §6.4 enlisted the ABH as a way to explain how an agent could go about life holding inconsistent beliefs. These inconsistent beliefs could be expected to play the roles that beliefs typically play in guiding behaviour insofar as they are activated in different contexts. I have also proposed that belief activation is determined, in large part, by the sorts of *information* that is salient to us in a particular context, and have suggested that very different information is salient to us in meta-ethical contexts, as opposed to morally engaged ones. We therefore have good grounds for the expectation that beliefs in moral propositions and the belief that the moral error theory is true will be active in different contexts.

Before moving on, it's worth emphasising that my central claim here is that different beliefs can be *active* in different contexts. (This is, to my mind, the best way to expand upon Olson's talk of "attending" to different beliefs in different contexts.) It is important to note that this is distinct from the claim that whether or not an agent believes that *p* depends upon the context in which she finds herself. My conservationist *always* believes both the moral error theory and moral propositions; it's just that these beliefs are not (or are very unlikely to be) activated at the same time.

§6.5 A JUSTIFICATION FOR PRESERVING OUR FALSE MORAL BELIEFS

The intentional cultivation of false beliefs is arguably the cardinal sin of philosophical inquiry. It also seems unwise. Generally speaking, we seem to do better to have a stock of true beliefs, and we seem to do worse if we rack up too many false ones. The conservationist therefore owes us a decent justification for preserving our false moral beliefs. Moreover, that justification had better not be premised upon a doxastic policy which sanctions choosing beliefs at our fancy. If the conservationist is too epistemically cavalier, then she is unlikely to find many willing partners.

I will now proceed to develop such a justification. My arguments will draw heavily upon Crispin Wright's (2004) notion of an *entitlement to cognitive project*. So it will be helpful to familiarise ourselves with his work before proceeding. Wright's concern is with so-called "cornerstone" propositions, which are central to a domain of thought in that a lack of warrant for them can deprive us of any warrant for other beliefs in that domain (2004, pp.167-8). He offers the example of the cornerstone proposition targeted by the Cartesian sceptic—that we are not now in a dream. We should want some warrant for believing that we are not in the midst of a dream if we are to be warranted in believing a range of empirical propositions. But the sceptic claims that no such warrant can be found.

Wright's response to the sceptic involves positing a form of warrant for a proposition which is "beyond rational reproach even though [one] can point to no cognitive accomplishment ... whose upshot could reasonably be contended to be that [one] had come to know p, or had succeeded in getting evidence justifying p" (2004, pp. 174-5). According to Wright, agents can have an *entitlement of cognitive project* with regard to cornerstone propositions: when a cornerstone proposition constitutes the founding presuppositions for particular cognitive project, one can have a warrant for believing it—for to doubt it would be to threaten the project itself. Importantly, the entitlement in question does not apply to just any old cognitive project. The project in question must either be "indispensable", or else so significant that its "failure would at least be no worse than the costs of not executing it, and its success would be better" (2004, p.192).

Now, I am not so much concerned with how (or indeed, whether) this strategy helps Wright to respond to the sceptic. I only include these details so as to avoid taking his proposal out of context. The strategy that I will pursue in what follows differs in some important respects from Wright's (though the parallels should be clear). For one thing, it is not especially important to me that the moral error theorist have a distinctively *epistemic* warrant for believing particular propositions. I am quite happy for the warrant to be a practical one.[204] Indeed, I expect that Wright himself would balk at the suggestion that a moral error theorist could be epistemically entitled to believe moral propositions. (The epistemic entitlements with which he is concerned only kick in when we lack sufficient evidence for thinking that the relevant cornerstone propositions are false.) Moreover, (and as I will clarify further in §7.3), my justification is best thought of as a justification for *the general policy* of cultivating beliefs in moral propositions, rather than a justification for believing any particular moral proposition at any particular time. What I wish to borrow from Wright, then, is the following, basic idea: we may sometimes be entitled to believe propositions on account of their being indispensable or else immensely important to projects of great significance.

Following Wright, it is plausible to think that there are some projects that are sufficiently important to us that their failure would at least be no worse than the costs of not executing them, and their success would be far better. But such projects need not be cognitive; they may very well be practical in character. Among these practical projects is, I think, the distinctively *social* project of interpersonal co-operation. In order to successfully partake in this social project, humans must, generally speaking, be capable of co-ordinating their actions, behaving in a prosocial manner, and shaping one another's conduct so as to achieve these ends. It is not implausible to suppose that this project is of immense importance to us; for our social context is not one from which we can readily escape. Social creatures that we are, it is imperative that we co-operate with one another as best we can.

[204] Some (e.g. Jenkins 2007) have raised the worry that Wright doesn't actually succeed in supplying us with an epistemic (as opposed to pragmatic) warrant for believing cornerstone propositions. But doing so is his ambition.

Here is the suggestion that I want to develop: we may be said to have an entitlement of social project with regard to moral propositions. This is not to suggest that we have such an entitlement with regard to just any old moral proposition. The basic idea is that given this social project, we may be entitled to adopt a policy of believing that some things really are right or wrong. (We might therefore want to limit the entitlement to existential statements and moral platitudes from which more substantive moral claims could be derived—for example, 'some things are morally right or wrong', 'we have moral duties to other people', and 'we have moral reasons to treat others fairly'.) Moral propositions are immensely important presuppositions for the project of social co-operation in that to doubt them would be to threaten the project itself. I take this suggestion to be strongly supported by the arguments developed and surveyed in the previous chapters. Let me rehearse them briefly.

First, moral discourse plays a valuable role in co-ordinating our attitudes and regulating interpersonal conflict (NRW 2005). With moral considerations in hand, we have an established framework—one of rights, duties, obligations, and the like—with which to navigate our way through practical disputes. Second, moral judgments (i.e., beliefs) are especially effective deliberation-stoppers; they prevent competing considerations that would interfere with prosocial motivation from entering into the deliberative sphere (Joyce 2006, p.111). Third, public moral judgments function as conversation-stoppers; they block any further negotiations from taking place when making interpersonal decisions (Dennett 1986, p.123). Finally, deliverances of blame serve as important checks and balances upon one another's behaviour (Recall that I assumed in §2.2.3 an account of blame according to which it involves a judgment of wrongdoing.)

Importantly, moral beliefs plausibly play these valuable roles because we tend to conceive of our moral obligations as *practically authoritative*—as things that we ought to do, period. As was argued in chapter 4, the commitment to categorical reasons is arguably of great importance if moral discourse is to serve these practical purposes. This suggests that the social project of interpersonal co-operation would have a far lesser chance of success if we were to merely believe *schmoral* propositions. Moreover, moral beliefs—as opposed to fictionalist attitudes—integrate with our behaviour and the rest of our psychology in fairly reliable and characteristic ways. So it is also doubtful that we could succeed in this social project were we to merely *pretend* to believe moral propositions.

Given these distinctive social benefits that come with having (false) moral beliefs in particular, there is a case to be made for our having an entitlement of social project with regard to moral propositions. To doubt these propositions would be to threaten the social project of interpersonal co-operation. This is not to say that our erroneous moral beliefs are *indispensable to* this social project. It is only to suggest that that project has a *far better* chance of success if we are afforded access to those beliefs.[205]

Thus, provided that we take the social project of interpersonal co-operation to be of paramount importance, and provided that we take (false) moral beliefs to be very important for this project to succeed, I think that we have in hand a powerful justification for working to preserve our false moral beliefs. Importantly, this justification nowhere appeals to a promiscuous doxastic policy that would let a thousand beliefs bloom. My conservationist need not think that *in general*, agents are free to pick and choose beliefs at their fancy. Quite the contrary; she sets the bar rather high. The defence that she offers for these beliefs appeals to a fundamentally important human project—one that we should want to succeed.

Two important clarifications should be made before concluding chapter 6. First, my justification should not be mistaken for a transcendental argument intended to license ontological commitment. I have not proposed, for example, that the existence of moral properties is indispensable to a rationally non-optional project (cf. Enoch 2011, ch.3). That would steer us dangerously close to a *vindication* of moral discourse—something which, we are assuming, is no longer on the cards. Second, what I have offered is best thought of an *initial* justification—one that appeals to what we stand to gain. A complete justification would require considering what we stand to *lose* as well. An important task for chapter 7 will be to consider some potential costs of conservationism.

[205] Note that Wright does not take indispensability to be necessary either; he thinks the relevant project must be "indispensable, or anyway sufficiently valuable to us" (2004, p.192).

Chapter Seven

Conservationism: Defence

I have argued that conservationism is a promising option for the moral error theorist. But I do not anticipate that everyone will share in my optimism, at least not yet. There are some important challenges to which we must attend.

For one thing, some may want to put pressure upon the claim that the attitudes that conservationism recommends are *beliefs* (Suikkanen 2013). Since these attitudes are taken towards propositions that are taken to be false, they would seem to be insensitive to one's evidence. But a sensitivity to evidence is commonly thought to be an important (if not constitutive) property of belief. In response, I shall argue that the conservationist's attitudes satisfy this requirement inasmuch as they are sensitive to a restricted body of her evidence (§7.1). I will also explain why such attitudes are not better characterised as attitudes of acceptance.

Even if the attitudes that conservationism recommends are properly called beliefs, one might worry that there are *significant costs* associated with cultivating false beliefs (Garner 1993, Joyce, ch.7)—far too many costs for doing so to be worth our while. Though intentionally holding onto false beliefs may sometimes be unwise, I will suggest that we shouldn't expect the cultivation of false *moral* beliefs to pose a significant threat to our interests or epistemic well-being (§7.2).

Yet another concern attaches to the *kind of justification* that the conservationist offers for continuing to believe moral propositions. This justification is pragmatic, appealing to our strong interest in the social project of interpersonal co-operation. Yet some philosophers deny that there can be pragmatic reasons for belief (e.g., Clifford 1877, Kelly 2002, Shah 2006, Whiting 2014). Thankfully, the conservationist need not take a stand on this issue. To see why, it is important to distinguish the justification that the conservationist offers for *the policy* of having moral beliefs (rather than no moral beliefs at all) from the justification that she offers for holding any particular moral belief at a particular time—as I shall do in §7.3.

A further problem is that it is not obvious that we have sufficient *control* over our beliefs to get conservationism going; if we do not, then the proposal may very well be infeasible. In response to this worry, I will explain why conservationism need not rest upon any strong or controversial species of doxastic voluntarism (§7.4). Indeed, there is a sense in which the difficulty we have believing propositions at will is *congenial* to the conservationist. It might very well be far *easier* to continue to believe moral propositions that to disbelieve them—or so I will argue.

One might also wonder whether conservationism is likely to be any more *stable* than fictionalism. Perhaps the conservationist is also in danger of ceasing to believe that some things are really right or wrong when the practical stakes are high. I will propose that to the extent that the conservationist is vulnerable to instability worries, she is far *less* vulnerable to them than the fictionalist (§7.5). After summarising the case for conservationism (§7.6), I conclude by seeing how well the proposal stacks up against its rivals (§7.7)

§7.1 SOME PUZZLES ABOUT BELIEF

I have serviced both a dispositionalist account of belief and the ABH in motivating the claim that the conservationist could *believe* both moral propositions and the moral error theory. But some may think that these attitudes towards moral propositions are better classified as attitudes of *acceptance*. Others may claim that whatever these attitudes are, they couldn't possibly be beliefs, since they seem wholly insensitive to evidence. This is concerning; if the conservationist cannot in the end recommend moral *beliefs*, then she may be forced to recommend rather feeble attitudes in their place—ones that are just as undesirable as those recommended by the fictionalist.

I begin with the first concern, according to which a more plausible contender for the attitude that the conservationist advises us to hold towards moral propositions is *acceptance*. As is the case with belief, acceptance of a proposition principally consists in as acting as though that proposition is true. But acceptance is commonly thought to differ from belief along a number of dimensions (Cohen 1989, Bratman 1992). For example,

> **Control**
> Unlike believing that *p*, accepting that *p* is usually within an agent's direct control.

> **Pragmatic justification**
> Unlike beliefs, which are justified by our evidence, acceptance is usually justified by some practical purpose that we wish to serve.
>
> **Context-dependence**
> Usually an agent either believes that *p* or she does not, regardless of the context in which she finds herself. Whether or not one accepts that *p,* by contrast, will be heavily dependent upon one's context.

On first blush, the conservationist's recommended attitude towards moral propositions would seem to bear all the hallmarks of acceptance. The justification for adopting this attitude seems to be pragmatic rather than epistemic, doing so is (it seems) assumed to be within an agent's control, and the attitude is only active in certain contexts.

It is debatable just how much of a problem this poses for the conservationist; for it is debatable just how much (if at all) acceptance really differs from belief. Horwich (1991), recall, thinks that acceptance *just is* belief. I am inclined to agree with Horwich that the distinction is overblown. But I hesitate to declare, as he does, that it is a "distinction without a difference" (1991, p.3). There do seem to be intuitive cases of believing a proposition without accepting it. One may sincerely believe that a ladder is stable, but refuse to accept that this is so before checking it more carefully (Bratman 1992, p.7). And if van Fraassen (1980) is right, then there can be cases of acceptance without belief. A scientist may not *believe* that her theory is the literal truth, because the evidence in its favour is not sufficiently decisive. But she may still choose to *accept* the theory—that is, to cease to inquire into its truth, and assume it as a basis for further research.

So I think we ought to concede that there are differences between belief and acceptance, even if they are not in the end very great differences. Accepting that *p* seems to involve the in principle possibility of acting as if *p* is true while being disposed to deny that one truly believes that *p* (Daly 2008). And unlike belief, acceptance is an attitude that is typically within an agent's direct (as opposed to merely indirect) control.[206] Let us assume, then, that there *is* a meaningful distinction to be drawn between acceptance and belief.

To begin with, I want to note that I don't think it would be utterly devastating for the conservationist if she were in the end forced to recommend attitudes of acceptance towards moral propositions. At the very least, acceptance would seem to be an improvement upon make-beliefs, which seemed unable to secure sufficiently similar behaviour and psychological responses to belief. One would expect accepting that *p* to have *remarkably similar* psychological and behavioural implications as believing that *p* does.[207]

Nonetheless, I want to resist the claim that the conservationist's recommended attitude towards moral propositions is more plausibly classified as acceptance. Presumably, we would expect an agent who accepted but did not believe that *p* to act as if *p* while being disposed to declare that she does not *really believe that p*. And this is certainly not what the conservationist recommends. She does *not* advise us to act in certain contexts as if moral propositions are true while being disposed to deny in *those* contexts that we believe them. (She does, of course, advise us to act in certain contexts as if moral propositions are true while being disposed to deny in *other* contexts that we believe them. But that is a different matter altogether.) If we take the conservationist's advice, then, in certain contexts, we will match the dispositional stereotype for belief with respect to moral propositions—something which plausibly involves a disposition to declare that we believe them.

[206] Cohen suggests that a belief that *p* is also distinguished by the disposition to *feel* that *p* is true (1989, p.368). But worries about the phenomenology of belief aside, this is far from clear. Does van Fraassen's scientist never feel, in the course of her daily experiments, that her theory is true? It's not clear to me that this is so. On these and related issues, see Church (2002).

[207] This issue is complicated by some murky boundaries. It is not uncommon to interpret van Fraassen's (1980) epistemic instrumentalism (which recommends attitudes of acceptance) as a form of fictionalism. So it is possible to understand acceptance as the kind of attitude that a moral fictionalist could recommend. But I suspect that a moral fictionalist who took this route would be especially vulnerable to the first horn of Olson's dilemma. Whereas a make-believer seems sufficiently different from a believer, acceptance may very well become suspiciously belief-y over time. (See Cohen 1992.) This is of course no worry for the conservationist. She has no qualms about slipping into false moral beliefs.

Moreover, and as I have noted (and will elaborate in §7.4) conservationists traditionally recommend *indirect* methods for cultivating the attitude in question. This would be odd advice coming from someone who was recommending attitudes of *acceptance*, which seem perfectly amenable to direct control. The conservationist's moral attitudes are also to some extent justified by her evidence—a claim that I will substantiate shortly (as well as in §7.3).

Let me now move on to address our second challenge, according to which the conservationist's attitude towards moral propositions would lack an important (if not constitutive) property of belief. It is common to think that beliefs are characteristically judgment-sensitive attitudes; that they are constitutively sensitive to an agent's thoughts about her evidence (e.g., Scanlon 1998, p.19, McDowell 1994, p.60). Yet the conservationist claims to believe moral propositions in spite of taking herself to have sufficient evidence for their falsehood. Given this, Jussi Suikkanen (2013) has argued that whatever attitudes a moral error theorist has towards moral propositions, they could not be beliefs.

There are a number of ways that we could go about responding to this challenge. One option would be to resist the claim that beliefs are constitutively responsive to evidence. Perhaps this claim is only true of *good* beliefs, rather than beliefs more generally (Huddleston 2012, p.214). Indeed, Andrew Huddleston (2012) has argued for the possibility of *naughty beliefs*, which are so-called because they persist *in spite of* the balance of one's evidence. A potential response for the conservationist would therefore be to characterise our beliefs in moral propositions as recalcitrant ones.

But I think we can do better than recommend naughty moral beliefs. Instead, I think the conservationist should claim that there is a sense in which the moral error theorist's attitudes towards moral propositions *are* responsive to her evidence. These beliefs do not, after all, float free of any considerations regarding the information that comes her way, or license outlandish inferences. We can expect them to play many of the roles that beliefs typically play in inferences and reasoning. To demonstrate, recall Schwitzgebel's inconsistent atheist. In certain contexts, Antonio is disposed to infer from the "magnificence of creation" that there is a "divine guiding force in the world", and to reason from significant life events that God has some role to play. Though Antonio's inferences are far from fantastic, they are certainly not *outlandish*. (He does not infer from the magnificence of creation that everything supervenes upon a turtle.) These attitudes certainly seem responsive to his evidence. It's just that in these contexts, only a restricted body of evidence is salient to him.

Parallel lessons apply to the conservationist. As I am imagining them, the conservationist's moral beliefs are not an epistemic free for all. They still exhibit good doxastic behaviour in those contexts where they make an appearance. Insofar as these attitudes match the dispositional stereotype of belief, they can be expected to play (at least many of) the relevant inferential roles. The conservationist who believes that murder is wrong, for example, is disposed to infer from the fact that 'murder is wrong' entails 'attempted murder is wrong' that attempted murder is wrong. Likewise, should she (within this context) come to discover that she also believes that 'murder is right', she will be disposed to dispense with one of these beliefs.

The conservationist's moral beliefs therefore seem to be sensitive to *a restricted body of evidence* that is *pertinent in certain contexts*. More specifically, they are responsive to evidence that we take to bear directly upon matters of first-order or applied ethics—her moral intuitions, and the cogency of first-order moral arguments, for example. The evidence to which these beliefs are not responsive (e.g., the evidence in favour of moral error theory) is not at all salient in these contexts. When deliberating about what to do when Philippa Foot's (1967) trolley is barrelling down the tracks, we rarely find ourselves asking questions about moral metaphysics.[208]

One might object that in order to be a belief, an attitude must be sensitive to *all* of an agent's evidence. But this seems false. The inconsistent atheist's beliefs aren't plausibly responsive to *all* of his evidence—but there are still good grounds for taking these attitudes to be beliefs.[209]

[208] Suikkanen (2013) worries that if one's beliefs are sensitive to only a restricted body of evidence, then there is nothing to distinguish them from make-beliefs. This strikes me as false, given that there are quite a number of features that distinguish beliefs from make-beliefs (recall §5.4.1).

[209] What of the related idea (mentioned in §5.4.1) that beliefs characteristically aim at truth? There are many ways to unpack this claim, and I cannot hope to devote space to all of them here. One plausible view takes this aim to be realised in truth-conducive processes (e.g., processes that are responsive to evidence) and requires that an attitude be formed and regulated by such processes if it is to count as a belief. (See Steglich-Petersen 2006, p.502.) If I am right that the conservationist's moral beliefs are responsive to (a restricted body of) her evidence, then they may at least be said to aim at truth in this sense.

It is a separate question whether or not the moral error theorist's moral beliefs could be *justified* by her evidence. (Even assuming that justification is fallible; that one can be justified in believing false propositions.) Establishing as much would require engaging in some heavy-duty epistemology—which would take us to far afield. But let me make a brief suggestion. Walter Sinnott-Armstrong (2006) has persuasively argued that our moral beliefs can be justified *even if* we cannot rule out the truth of moral error theory. His case is premised upon *contrastivism* about justified belief, according to which beliefs are only ever justified *relative to* a *relevant contrast class*. He proposes that

> Moral beliefs are modestly justified if they are justified with respect to a contrast class that by definition does not include moral nihilism or any other extreme alternative that would not be taken seriously in everyday moral deliberation. Hence, there is no need to rule out any extreme alternative like moral nihilism in order for a moral belief to be modestly justified. (Sinnott-Armstrong 2006, p.131)

The basic idea here is that a moral belief—say, the belief that one is morally required to give 10% of one's income to charity (*p*)—can be *modestly justified*; justified relative to a modest contrast class that includes 'one is morally required to give 99% of one's income to charity' (*q*) and 'one is morally required to give 0% of one's income to charity' (*r*). (Roughly, one can be justified in believing that *p rather than* that *q* or that *r*.) This is so even if the relevant moral belief is not *extremely justified*; justified relative to an extreme contrast class that includes 'one is not really morally required to do anything'.[210] More work would no doubt need to be done to explain how the conservationist could enlist this framework to make sense of justified moral beliefs. But it seems to me that a contrastivist epistemology (or perhaps even a contextualist one) would be the place to turn.

[210] Sinnott-Armstrong (2006, ch.6) does not think that any moral beliefs are justified *without qualification*; that is, justified out of *the* relevant contrast class—for he argues that we should suspend judgment upon which contrast class is *the* relevant one. The resultant position is what he terms 'Pyrrhonean Moral Skepticism'.

§7.2 The perils of false belief

According to the conservationist, the error theorist would do best to hold onto her false moral beliefs. However, it must be admitted that there are significant worries with holding onto false beliefs. As Joyce notes, true beliefs tend to be instrumentally valuable, and false beliefs, instrumentally disvaluable. Agents typically act in ways that would satisfy their desires if their beliefs were true, and we would expect that

> In the vast majority of cases having a true belief to act upon is more likely to bring satisfaction of desire than having a false belief on the matter… given that we don't know in advance how and when we are going to employ a particular belief, the safest bet is to have the true one over the false one. (Joyce, p.179)

There is something importantly right about this thought. We (tend to) act in such a way as to best satisfy our desires given the way we take the world to be. If all else is held equal, having a more accurate picture of the world will lead to actions more likely to satisfy our desires. But it does not follow from this that the best strategy for improving the likely outcomes of our choices will *always* be to rack up more true beliefs. After all, the world may be set up in such a way as to *punish* having more true beliefs. An agent with a more accurate picture of the world will make better decisions given the options available to her and her assessment of their potential outcomes. But having a more accurate picture of the world might also change her circumstances such that the options available to her are, on the whole, *worse*.

To take an extreme case, imagine an evil demon who hates know-it-alls, and consequently limits the options of the person with more knowledge to only the most undesirable ones. In this kind of case, the person who knows all of the truths will make the best decisions available to them, but all of their options will be worse than if they'd known nothing at all! Or, as a less fanciful example that's closer to home in the present debate, it may be the case that learning the truth of the world causes one to evaluate every outcome as worse than they considered it before they knew the truth. Where before learning the truth of the moral error theory one might have gotten a lot of pleasure out of doing 'the right thing', post-error theory one could see all of their options as bleak and meaningless.

It is the conservationist's contention that the world *is* set up in such a way so as to systematically punish (always having active) true beliefs about the status of moral propositions (i.e., that they are systematically false). Lacking false moral beliefs might result in a more unified and accurate picture of the world. But it would also foreclose valuable certain opportunities; in surrendering our false moral beliefs, we are likely to lose many desirable practical goods.[211]

Joyce also takes issue with a *general doxastic policy* that would permit us to hold onto false beliefs whenever it suited our fancy. "The policy of aiming for truth", is, he maintains, "the best doxastic policy around" (p.179). However, and as I have emphasised, conservationism need not be premised upon any sort of objectionable doxastic policy. My own conservationist insists that we must clear a high bar before we can be justified in making an exception to our truth-seeking policies. This isn't to say that one couldn't take issue with the proposed justification. But it would be unfair to charge her with unbounded epistemic promiscuity.

Nonetheless, Joyce is reluctant to permit such exceptions. Citing C.S. Peirce, he warns that doing so may lead to "a rapid deterioration of intellectual vigor" (p.179). Once we admit too many false beliefs, we could very well be led down a slippery slope into epistemic chaos. But I do not think that the slope here is quite so slippery; for we are perfectly capable of being *discriminatory* in our cognitive policies. To preserve false beliefs in one domain on pragmatic grounds is not necessarily to threaten stringent truth-seeking policies in other domains. Following Olson, it seems implausible to expect that one who "adopts a pragmatic policy regarding beliefs about libertarian free will to be less committed to a truth-seeking policy in mathematics" (2014, p.186).

Still, there is a concern in the vicinity of Joyce's worry. Garner (1993) has forcefully argued that the propagation of false beliefs can be psychologically harmful.[212] Plato's so-called "noble lies" are a case in point. The citizens of Plato's envisioned society are to be told that they were created by a God who made rulers from gold, auxiliaries from silver, and artisans from iron and brass. Garner predicts that

[211] I am grateful to Edward Elliott for helpful discussions on this point.

[212] Joyce briefly touches upon this worry as well when he claims that "A seemingly useful false belief… will require all manner of compensating false beliefs to make it fit with what else one knows…" (p.179). But Garner does more to elaborate these concerns, and so, I focus upon his developments here.

> … a massive fabrication like the "myth of the metals" would do serious damage if injected into a person's or a society's cognitive system and accepted as true. We never believe single, isolated facts; rather we subscribe to complex, interrelated networks of mutually supporting beliefs. In order to sustain the belief [in Plato's myth], we would have to find some way to neutralize the many facts that don't fit with that fanciful claim. … the result would be a very confused group of people, unsure of what to believe, and unable to trust their normal belief-producing mechanisms. (1993, p.96)

Thus, injecting false moral beliefs into a society may pose a certain kind of risk. In doing so, we run the risk of reducing its members to "epistemological wrecks" who struggle to integrate these false beliefs with everything else that they know.

Though what Garner says about Plato's noble lies may very well be true, it is not clear to me that believing even preposterous things is generally likely to a render us epistemological wrecks. Many *do* go about life believing what others regard as absurd. Rumpologists believe that the shape of our derrières provides insight into our futures. Members of flat earth societies believe that the world is flat—and that the sun is a mere five thousand kilometres away! Quite a number of us are inclined to regard these claims as no less fanciful than the myth of the metals. But whatever one has to say about these people, they don't see to go about life suffering, confused, and unable to trust their belief-forming mechanisms.

Thus, believing falsehoods—even fanciful falsehoods—does not always lead to epistemological turmoil. Even if it sometimes can, though I don't think that this makes for a very plausible prognosis for our false *moral* beliefs. Given the absurdity of the myth of the metals, a belief that that myth is true will be difficult to square with a lot of the information that comes our way. We very often come across evidence that we were not fashioned from scraps of metal (e.g., we see pregnant women, and check our weight on the scales). Holding onto this myth is likely to require a radical reinterpretation of such evidence. Our false moral beliefs, by contrast, have peacefully co-existed with our true non-moral beliefs for quite some time. And the evidence for moral error theory is not something that many people are likely to come across in day to day life.

§7.3 NO PRAGMATIC REASONS FOR BELIEF

I have argued that we are justified in holding onto our false moral beliefs on account of their pragmatic benefits—specifically, on account of their being of paramount importance to the social project of interpersonal co-operation. But it is controversial whether beliefs are amenable to pragmatic justification. A distinguished line of philosophers has answered in the negative (Clifford 1877, Kelly 2002, Shah 2006, Whiting 2014). One may have thought that this was demonstrably false. Presumably, I could hold a gun to your head and threaten to pull the trigger unless you believe that your mother is a vampire. Hey presto! You have a pragmatic reason for belief!

Quod erat demonstrandum? Not quite. Opponents of pragmatically justified belief would insist that I have merely supplied you with an ulterior motive for believing that your mother is a vampire—I have not given you any real reason for believing that this is so.[213] The specific opponent I have in mind here is the *evidentialist*. In her view, only facts about one's evidence can constitute reasons for belief. Within evidentialist circles, believing a proposition for pragmatic reasons is regarded as a form of intellectual dishonesty, self-deception, or "wishful thinking" (Bratman 1992, p.3).

One might suspect that evidentialism has already been ruled out on account of our having assumed epistemic instrumentalism (in §2.1.1); the view according to which an agent has an epistemic reason to believe some proposition *p* when (and because) doing so would be instrumentally rational given her goals. But as Cowie (2014b) notes, the instrumentalism–intrinsicalism debate is somewhat orthogonal to the evidentialism–pragmatism debate. Instrumentalists are committed to the claim that "the practical utility of evidentially supported belief is the *explanatory grounds* of the normativity of evidence for belief" (Cowie 2014b, p.4005, emphasis added). (In this respect, they differ from intrinsicalists, who instead appeal to a brute, normative truth.) But instrumentalists are not thereby committed to the claim that pragmatic

[213] I borrow the phrase from van Fraassen (2002, p.89), who distinguishes between an epistemic reason for a belief (something which makes it more likely that the belief is true) from an ulterior motive for a belief (which does not make it more likely that the belief is true).

considerations *are themselves* reasons for belief.[214] Even if it is the practical utility of evidentially supported belief that *grounds* our reasons to believe in accordance with our evidence, it is still an open question whether or not only evidential considerations *constitute* reasons for belief.

Now, I am inclined to think that pragmatic considerations *can* constitute reasons for belief. Indeed, I suspect that they may sometimes give us reasons to believe contrary to our evidence.[215] But this is not the line that I am going to pursue here. Instead, I want to show why, despite first appearances, much of what the conservationist has to say is perfectly consistent with evidentialism. If I am right, then conservationism need not be rejected by those who deny that pragmatic considerations can be reasons for belief.

To show why, it will be helpful to build upon a suggestion made by Berislav Marušić (2011, p.36). As Marušić notes, there are two sorts of questions that we could be asking when we ask what an agent has reason to believe. We may be asking (i) the *synchronic* question as to what she should believe at any particular moment, or (ii) the *diachronic* question as to what she should get herself to believe over time. Marušić suggests that evidentialists are chiefly concerned with the former question. He cites Richard Feldman, who is explicit about the restricted scope of the evidentialist's ambitions:

> Evidentialism is best seen as a theory about synchronic rationality. It holds that the epistemically rational thing to do at any moment is to follow the evidence you have at that moment. It doesn't address questions of how to conduct inquiry over periods of time. Thus, it does not address questions about how to gather evidence, when one ought to seek additional evidence, and so on. In my view, these diachronic questions are moral or prudential questions rather than epistemic questions. (2000, p.689)

[214] This is not to deny that many instrumentalists *do* think that pragmatic considerations can constitute reasons for belief. Following Cowie, this may even be seen as "a natural extension of the instrumentalist view" (2014b, p.4005). But it is not a necessary one.

[215] Pace cites the example of the addict who "…chooses the road to recovery—believing that this time he can change—because he recognizes that not to believe would almost certainly make recovery impossible, causing harm to himself and those he loves" (2011, pp.240-1; see also Marušić 2011, Leary 2016).

What is true of Feldman may not be true of evidentialists more generally. But Marušić is surely correct that the diachronic and synchronic questions are dissociable. Distinguishing the two can, I think, help us to appreciate the sense in which conservationism is consistent with evidentialism.

To this end, let us distinguish a particular sort of diachronic project from a particular sort of synchronic one. In particular, let us distinguish (a) the project of justifying the policy of holding beliefs in some domain from (b) the project of justifying holding any one belief in that domain at any particular moment. The conservationist's venture is one of the former kind; her goal is to offer a diachronic justification for *the policy of holding onto moral beliefs*. This policy involves taking steps to make moral belief-formation more likely; exposing ourselves to particular sorts of evidence, and placing ourselves in certain kinds of situations, for example. (More on this in §7.4.) The justification that the conservationist offers for this policy is pragmatic, appealing to our strong interest in the social project of interpersonal co-operation.

However—and importantly—the conservationist's justification for holding any particular moral belief at any particular time is *not* pragmatic.[216] The considerations that justify her believing some moral proposition *p* at some time *t* are evidentialist in character. She will, for example, take herself to be justified in believing that *p* because her moral intuitions afford her strong evidence for believing that *p*, or because *p* follows from what she takes to be the best moral theory. Thus, whereas the conservationist's justification for *having moral beliefs* (rather than no moral beliefs at all) is pragmatic, her justification for *having any particular moral belief* (e.g., for believing that *p* rather than that ¬*p*) is an epistemic one.

As I have repeatedly insisted, the conservationist's moral beliefs do not float free of her evidence. In contexts where these beliefs make an appearance, there is a selective body of evidence to which they remain responsive—that which we take to bear upon first-order moral issues. When considering the merits of utilitarianism, the conservationist will draw upon the sorts of considerations that an evidentialist would en-

[216] Given epistemic instrumentalism, there will always be a sense in which any justification 'bottoms out' in pragmatic factors; an agent's goals will still serve as the explanatory grounds for her reasons for belief. But again, they need not *constitute* her reasons for belief.

dorse; that it is a unified and simple theory, that it coheres (or conflicts) with many of her intuitions, and the like.

§7.4 A FEASIBILITY WORRY

The conservationist advises us to believe both the error theory and moral propositions. But one may worry that we often have very little (if any) voluntary control over what we believe. Alston asks,

> Can you, at this moment, start to believe that the United States is still a colony of Great Britain, just by deciding to do so? … suppose that someone offers you $500,000,000 to believe it, and you are much more interested in the money that in believing the truth. Could you do what it takes to get that reward? [. . .] Can you switch propositional attitudes toward that proposition just by deciding to do so? It seems clear to me that I have no such power (1989, p.122).

Heil similarly notes that the phenomenology of forming a belief is not that of having reached a decision:

> When we attend to the matter, beliefs seem most often to come to us, unsought and unbidden, on the heels of thought and investigation. The notion that one might come to believe something simply by willing it has, if not exactly an air of contradiction, at least a strong whiff of implausibility. (1984, p.59)

These remarks are but a few samples. But in my experience, they are representative ones. The idea that agents typically have voluntary control over their beliefs is often regarded as a non-starter. Or, at least, it is generally viewed with suspicion.

Before proceeding to examine such suspicions, it will be helpful to clarify what is meant by 'voluntary control'. For my purposes here, I shall follow Alston in supposing that to have voluntary control over a belief that *p* is to have the "power to carry out an intention to" believe that *p* (1989, p.136). As Chuard and Southwood (2009, p.603) elaborate, Alston's notion of "carrying out an intention" to believe that *p* plausibly requires that the intention to believe that *p* be causally efficacious (in the right way—no deviant causal chains permitted) in bringing about the belief that *p*. Further, it seems that the control in question must be *direct*. Alston and others do not deny that we can have *indirect control* or *voluntary influence* over what we believe. The whiff of implausibility attaches to the idea that we typically have direct voluntary control over beliefs themselves; that we can effectively choose whether to believe some proposition *p* right here and now.

Corresponding to this latter distinction between indirect and direct control is a distinction between two species of doxastic voluntarism (Vitz 2009). Proponents of *direct doxastic voluntarism* take agents to have direct voluntary control over (at least some of) their beliefs. In their view, an agent can choose to believe that *p* right here and now, much like she can choose to think about her favourite colour. Sponsors of indirect doxastic voluntarism, by contrast, take agents to have only indirect voluntary control over (at least some) of their beliefs; an agent can, for instance, control what she believes by way of gathering further evidence, or changing the world in some way.

As should be clear, Alston's and Heil's remarks are targeted at direct doxastic voluntarism. Very few (if any) deny that agents can have *indirect* control over what they believe. An agent can plausibly have indirect voluntary control over whether she believes that *p* when she has control over whether *p* is true. One can, for example, have indirect voluntary control over whether they will believe the proposition 'the light in this room is off' by having control of the light switch (Feldman 2000, pp.671-2). Alternatively, an agent may have indirect voluntary control over whether she believes that *p* by having control over whether she will seek out the evidence in its favour or disfavour (Vitz 2009).

It is therefore relatively uncontroversial that agents can have indirect control over what they believe. A natural strategy for the conservationist, then, is to recommend *indirect* methods for cultivating moral beliefs. To this end, she may follow what Gale (1999, p.87) terms the "acting-as-if-you-believe recipe for self-inducing belief". Much like Pascal advised acting like a sincere religious believer in order to acquire belief in God, the conservationist can recommend that we go about our days acting just as moral believers do: assessing others in moral terms, promoting certain social policies on moral grounds, invoking moral considerations when deliberating about what to do, and so on.

More specifically, the conservationist can recommend continuing to place ourselves in contexts where moral beliefs are likely to be activated: the first-order ethics classroom, political protests, and the like. Doing so would make certain kinds of information especially salient to us. When debating the merits of Donald Trump's latest executive order, the sort of evidence that will be relevant will presumably be how that order is likely to impact upon people's lives, the ulterior motives of Trump's advisors, and the like. Upon being faced with this evidence, it will be difficult not to *believe* that the executive order is morally reprehensible, or indeed, downright wrong. Of course, this sort of information won't be salient in *all* contexts. When discussing meta-ethics over a coffee, the evidence in favour of moral error theory is likely to be salient to us. In the latter context, we may continue to find the case for moral error theory incredibly convincing, and so, it may be quite easy to for us to believe that *nothing* is right or wrong.

It seems to me very plausible that we can in some contexts take the arguments in favour of a proposition *p* to be overwhelmingly convincing—so convincing that we find ourselves believing that *p*—and in other contexts, take believing that ¬*p* to be irresistible. Indeed, I think this is an experience with which many philosophers should be familiar. At was noted at the very beginning, not everyone has taken Lewis's maxim of honesty to heart ("never put forward a…theory that you yourself cannot believe in your least philosophical and most commonsensical moments"). Peter Unger (1979) believes there are no people. And Caspar Hare (2009) believes that there is just one (himself). Indeed, I suspect that Hare may find it rather easy to believe as much when caught in the throes of solipsistic reasoning.[217] But I strongly suspect that he finds it rather difficult *not* to believe that other people exist when he is buying groceries or out to dinner with friends. This is precisely my prognosis for our moral beliefs. Though believing moral propositions is likely to be rather difficult in the meta-ethics seminar, it is likely to be rather easy when discussing the behaviour of tyrannical world leaders or ruthless businessmen.

Some may be inclined deny that the solipsist every *really* believes that she is the only person who exists. Yet this diagnosis is far from obvious, and I think we should be reluctant to offer it in all such cases. But perhaps more needs to be said to ac-

[217] Hare defends solipsism in his aptly titled *On Myself, and Other, Less Important Subjects* (2009). I should note that he in fact defends a rather sophisticated form of solipsism, according to which other people exist, but their experiences are *not present*. I think what I say above applies with appropriate transformations to this variety of solipsism as well, but will simplify Hare's view for ease of illustration.

count for this phenomenon of being a solipsist who sometimes believes there are other people, and being a moral error theorist who sometimes believes that cheating is wrong. Doing so will help to support my contention that conservationism is likely to make for a feasible option going forward.

To begin with, I think this phenomenon should be relatively unsurprising when some of the beliefs in question are perceptual beliefs. Our perceptual belief-forming mechanisms are, for the most part, automatic. Experimental studies suggest that evaluation (and rejection) of what is perceived requires additional processing (e.g., Gilbert et. al 1993). Given this, we should expect that it will typically be very easy to believe that one's perceptual-forming mechanisms are reliable *while* in the throes of perception—that is to say, it will often be easy to believe our eyes.[218] But there could very well be other contexts where one is not attending to what one perceives, and is overwhelmed by the evidence in favour of the unreliability of one's perceptual capacities. In the latter contexts, it may be very difficult *to believe* any of the past deliverances of one's perceptual system.

Though it would be implausible to claim that everything that is true of perceptual cognition is also true of moral cognition, there are some interesting analogies (Sterelny 2010). Vision presents us with a particular picture of the world; one which we usually believe, but can overrule—especially at a later time. Similarly, it seems that we often make quick intuitive judgments about the moral status of particular actions that we may later overrule following further reflection. As Sterelny notes, many of our moral assessments are "…fast and automatic. We do not have to decide whether to evaluate a situation normatively when we read of the abduction of a three-year old girl" (2010, p.287).

A complementary idea here is Haidt's (2001) hypothesis that moral judgments are typically the products of "intuitions"—roughly, emotionally-laden, gut responses. He proposes that moral reasoning has very little causal influence upon our moral judgments. In Haidt's view, much of moral reasoning merely amounts to a *post hoc* rationalisation for these deeply held moral intuitions. Indeed, some of Haidt's studies

[218] Granted, there may be exceptions. If I know that the Müller-Lyer illusion is an *illusion*, then I might be able to resist believing my eyes. My suggestion here is merely that belief is plausibly the default.

suggest that these affect-laden moral assessments can persist even when *no* reasons can readily be offered in support of them—a phenomenon that has been referred to as "moral dumbfounding". (See for example, Haidt 2001, Cushman et. al 2006, Hauser et. al 2007.) I think the automaticity with which we often make moral judgments is an important point to appreciate in the context of the motivating the feasibility of error theorists continuing to believe moral propositions. But some clarifications are needed before we can draw conclusions from it.

To begin with, we need to appreciate that there are disanalogies here as well.[219] Our visual systems process information quickly and automatically, which makes sense given that the sort of informational currency in which they trade is often going to be most useful in the here and now. As Sterelny observes, moral cognition doesn't always have this element of urgency. Much of moral thinking takes place *offline*; we can and often do assess moral issues slowly and carefully (2010, pp.286-7).

Indeed, Haidt's view seems to have difficulty capturing the ways in which a careful consideration of moral issues can plausibly shape our moral judgments.[220] Some theorists have attempted to do so by developing "dual process" models of moral cognition, according to which moral judgments can issue from two different sorts of processes.[221] One sort of process is automatic, associative, emotionally-laden and typically takes place with little to no conscious effort. These are the kinds of processes that seem to be at work when we read of the abduction of a three-year old girl, and arrive at a moral assessment of the situation quickly and effortlessly. The other sort

[219] Indeed, part of Sterelny's (2010) project is to argue that moral cognition, unlike vision, is not plausibly modular. He explains the appearance of moral modularity by appealing to pattern recognition, which is often automatic and effortless.

[220] For developments of this criticism, see Fine (2006) and Paxton and Greene (2010). In fairness to Haidt, he does attribute *some* causal power to moral reasoning—he suggests that the "sheer force of logic" can give rise to moral judgments (2001, p.819). But not nearly enough, it seems; for such cases are hypothesised to be rare.

[221] It's worth noting that dual-process theorists are not committed to associating these processes with different brain regions, even though some (e.g. Greene et. al 2004) seem to do so. For reasons given by Klein (2011), great caution should be exercised here.

of process is slow, rule-based, and deliberative, and is at work when we're faced with moral decisions that require careful consideration and difficult trade-offs.[222]

Importantly, each of these processes can shape the deliverances of the other. On the one hand, our considered moral judgments can (at least over time) shape the kinds of quick, associative moral judgments we are disposed to make (Fine 2006, Paxton & Greene 2010). For example, studies suggest that our explicit moral values have the potential to counteract implicit moral attitudes, such as racist biases (Monteith et. al 1998, Moskowitz et al. 1999). On the flipside, our moral intuitions are important *inputs* into moral reasoning. When I sit down and think long and hard about the prospects of utilitarianism, I take it to count against the position (at least somewhat) if it goes against my intuitive judgments—the strong aversion I have to causing harm, say. Moral intuitions can therefore play an important role in guiding the acquisition and rejection of moral beliefs even in more engaged contexts when there is adequate time for deliberation.

Having made these clarifications, I am finally in a position to offer a conjecture. Just as it would be difficult for a solipsist not to believe there are other people while he is perceiving his dinner guests, we should expect that it will be difficult for a conservationist not to believe moral propositions when she is presented with the sorts of stimuli that prompt quick moral assessments. Moral intuitions and implicit moral attitudes play an important role in shaping our moral judgments, and they would plausibly help to reinforce and sustain the conservationist's moral beliefs. When we observe some children setting fire to a cat, it will be hard not to *believe* that they are doing something wrong. Importantly, these sorts of assessments can also be expected to make an appearance in contexts where we attend to moral issues slowly and carefully—as they plausibly do when we engage our imaginative capacities and think of the harmful consequences of utilitarianism.

[222] Uriah Kriegel (2012) has suggested that we should take the products of the automatic, associative processes to be moral *aliefs* (of the kind described by Gendler (2008)) and those of the slower, rule-based system to be moral beliefs. But I don't that we should be so quick to do so. Dual-process theorists more generally tend to count the outputs of associative processes as beliefs, and Kriegel doesn't provide any reason for thinking that matters ought to be any different in the moral case. The alief–belief distinction does not seem to me to map neatly onto the dual-process distinction (Kriegel may of course be right in thinking that aliefs are the product of associative processes. My claim is that we shouldn't think that it is *only* aliefs that can result from these processes.)

Accordingly, I think there is a sense in which the implausibility of direct doxastic voluntarism is *congenial* to conservationism. At first, the conservationist seemed to face the problem of explaining how we could conceivably get ourselves to believe moral propositions. But given the considerations above, we should expect that it will be rather *easy* to believe moral propositions in certain contexts where particular information is salient to us.[223] Of course, such information won't be salient in *all* contexts. When we are in a detached context doing moral metaphysics, the evidence for moral error theory is what will be salient to us. In the latter situation, it is may very well be easy to believe that nothing is really right or wrong.

§7.5 A STABILITY WORRY

I want to consider now some worries one may have regarding the stability of conservationism. Consider the following scenario:[224]

> Jane is a conservationist. She works for a company doing something seriously morally wrong (as we would ordinarily say, anyway), and she is considering whether to become a whistle-blower. She knows that by doing so she will incur a great risk to her career, perhaps endanger friendships she has in the company, and it has an uncertain chance of success. Still, she believes that what the company is doing is very wrong, and believes she has a moral obligation to try to stop it. Jane begins to deliberate whether to become a whistle-blower. While deliberating, she considers a wide range of her beliefs: beliefs about what it is that the company is doing, whether publicity is likely to change its behaviour, how risky exposing them would be to her current job, whether she can get another job in the industry with a reputation as a trouble-maker, whether the company will try any dirty tricks against her, and so on.

One might worry that it would be hard for Jane to avoid attending to the fact that what the company is doing is not *really* wrong (since nothing is wrong). This is, after all, something that she believes, and it seems relevant to whether she should do

[223] In this respect, a moral error theorist may be akin to Hume's Pyrrhonian sceptic, who cannot take on her scepticism in everyday life because "nature [is] too strong for it" (1740/1978, p.657).

[224] I am grateful to Daniel Nolan for suggesting this case to me.

something that is supposed to have the benefit of being the right thing to do even at the cost of wrecking a lot of other things in her life that she cares about.

Now, it's not obvious to me that the truth of moral error theory really *would* occur to Jane in the course of her deliberations, even if the stakes are high and there is adequate time to think things over. Getting the numbers right on one's tax return is important, and there is often plenty of time to get things sorted. But I don't think that nominalists are likely to attend to their belief that there are no numbers when filling out their tax returns. Having registered these doubts, let me grant for the sake of argument that Jane might consider the truth of moral error theory while deliberating. This suggests two problems. Firstly, it seems that Jane would be less likely to follow through on her moral beliefs than a moral realist analogue of Jane might, since that analogue wouldn't bring into deliberation the belief that there is no moral value to be gained by stopping the company. This invites the question as to why a conservationist shouldn't favour self-expunging error theory.[225] Secondly, it seems that conservationism, like fictionalism, has the potential for instability—at least when the stakes are high, and there is adequate time for deliberation.

Regarding the first problem, I think the conservationist should be willing to concede that Jane might not be as reliable as her realist analogue. But I also think that the conservationist should hold onto the moral error theory for the reasons that I cautioned against propagandism (§2.3.5). Even if she did manage to conveniently suppress the reality of moral error theory, there is no guarantee that the evidence in its favour won't come her way some time in the future. So there would always be a risk of relapse. She could, of course, try to introduce some safeguards, perhaps avoiding philosophical discussions at all costs or leading a movement to remove all error-theoretic texts (as well as commentaries and any associated discussions) from circulation. But this seems highly undesirable. The outright suppression of philosophical discussion and arguments (for fear of their dangerous consequences) is not something that we typically want to encourage.

[225] Put differently, one might ask why a conservationist does not favour what Cuneo and Christy (2011, p.93) call *intransigentism*, which involves a refusal to ever "…entertain seriously any evidence that contradicts the claims made in moral discourse".

As far as the second problem is concerned, I think the conservationist is at least far *less* vulnerable to instability worries than the fictionalist. Fictionalist attitudes are highly overridable. But moral beliefs are not so easily overridden—especially in contexts where particular information is salient to us. It is difficult not to believe that the child pornography industry is morally heinous when hearing about the harm that it is doing to young children.

Moreover, even if the truth of moral error theory did occur to Jane in the course of her deliberations, I think it is at the very least debatable how *deeply* she could believe it in a context where other sorts of evidence is salient to her. If her company is testing on animals, for example, and she is strongly averse to causing animals pain, then I expect it will be difficult for the belief that testing on animals is not *really* wrong to approach anything close to a deep conviction. (Notice that the fictionalist can't enlist this reply; she doesn't want to rely upon the ease with which we can *believe* that some things really are right or wrong.)

Talk of depth (as opposed to strength) of beliefs might seem mysterious at first. But as Jennifer Church (2002) has suggested, we seem to be tracking something close to this idea when we speak of *taking something to heart*. Church distinguishes a jury member who confidently arrives at a guilty verdict but remains on some deeper level unconvinced from one for whom this guilty verdict has really sunk in. For the latter person, the guilty verdict is not merely a "phrase on a page or in someone's mouth", but something that integrates with her thoughts, feelings, and actions; she finds herself making negative assessments of the defendant's character, and feeling anger towards him, for example (2002, pp.366-7).[226]

At the very least, then, I think it would be difficult for the truth of moral error theory to *sink in* when Jane is in the course of deliberating about how much these animals are suffering—for that belief to integrate smoothly with her other thoughts, her feelings, and her actions. It is difficult to imagine that Jane will cease to think that the CEO of her company is a bad person, or to feel morally outraged at the company's actions because there is no such thing as badness, and no moral outrage to be had.

[226] See also Buckwalter, Rose and Turri (2015), who make a similar distinction between 'thin' and 'thick' belief, and draw attention to its explanatory serviceability.

§7.6 The case for conservationism: a summary

Chapters 6 and 7 have quite a number of moving parts. It will be helpful to take stock. I began by pointing towards some gaps in Olson's conservationism (§6.1), and used these to develop four desiderata for a workable conservationist proposal (§6.2). I claimed that if the conservationist's proposal is to be attractive, then she must motivate taking the moral error theorist's attitudes towards moral propositions to be beliefs (D1), motivate the supposition that the moral error theorist would attend to her beliefs in moral propositions in some contexts and to her belief that the moral error theory is true in others (D2), explain how these inconsistent beliefs could be expected to play the roles that beliefs typically play in guiding behaviour (D3), and justify overriding the presumption against intentionally cultivating false beliefs (D4).

The remainder of chapter 6 was devoted to satisfying these desiderata. In §6.3, I argued that the conservationist would be properly characterised as believing both the moral error theory (*e*) and moral propositions (*m*) insofar as she reliably matched the dispositional stereotype for belief with respect to *e* in certain contexts, and the dispositional stereotype for belief with respect to *m* in others. The goal for §6.4 was to motivate the idea that this doxastic behaviour could plausibly be expected of her. Here, I argued for the ABH, according to which not all of an agent's beliefs are activated at a particular time. I also proposed that the information that is salient to an agent plays an important role in determining which of her beliefs are activated. In addition, I suggested that very different information is likely to activate moral beliefs, as opposed to the belief in the moral error theory. In §6.5, I offered a justification for preserving our false moral beliefs—one that did not simply appeal to the general truism that false beliefs can sometimes be instrumentally valuable. I argued that we have good reason to hold onto false *moral* beliefs in particular on account of their centrality to the social project of interpersonal co-operation.

The task for chapter 7 was to address some niggling worries. The first of these had to do with my characterising the conservationist's attitudes towards moral propositions as beliefs (§7.1). Here, I explained that these attitudes would have a number of important properties that we commonly associate with belief; they would at least be responsive to a *select body* of evidence, and they could be expected to play many of the roles that beliefs typically play in inferences and reasoning. I further argued that such attitudes were not better characterised as attitudes of acceptance.

A second worry was that conservationism might give rise to a kind of epistemic turmoil. Though some false beliefs may have the potential to do so, I argued that we shouldn't expect the same of false *moral* beliefs (§7.2). Another apparent problem for the conservationist's doxastic practices concerned her having offered a *pragmatic* justification for believing moral propositions (§7.3). Given this, it seemed that the conservationist was committed to the controversial claim that pragmatic considerations can constitute reasons for belief. In response, I clarified that the conservationist's pragmatic justification is restricted to the *general policy* of believing moral propositions.

Yet another worry was that the feasibility of conservationism may rest upon an overly strong and implausible variety of doxastic voluntarism (§7.4). As we have seen, the conservationist need only assume (plausibly) that we have some measure of *indirect* control over what we believe. Indeed, the implausibility of *direct* doxastic voluntarism is somewhat congenial to conservationism; it is likely to be incredibly difficult for us *not to* believe moral propositions—especially if we continue to place ourselves in contexts that are likely to 'activate' moral beliefs. In addition, I have argued that we should not expect conservationism to be unstable in the long run (§7.5). At the very least, it is likely to be a more stable option than fictionalism.

I hope my arguments have at the very least convinced the reader that conservationism is a proposal to be taken seriously. Whereas other proposed answers to our WNQ have notable shortcomings, conservationism promises to deliver (just about) everything we could ask for. I will now proceed to emphasise its advantages over rival proposals.

§7.7 A COMPARATIVE ASSESSMENT

My final task in this chapter will be to examine how well conservationism stacks up against its rivals. We can begin by comparing the proposal with the abolitionist's advice for life after moral error theory. *Pace* the abolitionist, I argued that we would seem to do better to hold onto our moral practices (in some form). Conservationism thus has a straightforward advantage over abolitionism; it is better placed to preserve the benefits of engaging in moral practice. Conservationism would also seem to fare better than abolitionism in terms of its feasibility. I have suggested that holding onto our false moral beliefs would not be nearly as difficult as we might expect. Certainly, it would be far *less* difficult than a wholesale purge of moral language.

The conservationist would also seem to do better than the revisionist in recommending that we hold onto moral discourse in its current, error-ridden form. The revisionist advises against taking moral requirements to have categorical authority. Given this, her schmoral discourse seemed unlikely to preserve many of the benefits of moral discourse. The conservationist is not vulnerable to this worry; for she recommends the preservation of moral discourse as it stands (categorical reasons, warts and all).

To my mind, the conservationist's strongest rival is the fictionalist, who is also well-placed to preserve talk of categorical reasons. However, the conservationist would seem to do better than the fictionalist in recommending that we continue to *believe* moral propositions. In contrast to the full-blooded moral beliefs that conservationism recommends, fictionalist attitudes didn't seem to be a stable enough or a strong enough basis for securing the benefits of moral practice. Our moral beliefs tend to integrate in fairly reliable and characteristic ways with our behaviour and the rest of our psychology—and these, I argued, are connections that we should want to preserve. Given that fictionalist attitudes can often have very *different* implications for our motivations and our behaviour, they seemed unlikely to secure the right sorts of ties between moral judgment and action.

Conservationism also seems likely to be a more stable option than fictionalism. Sustaining a pretence requires cognitive effort, especially when the suspension of make-belief would carry an immediate advantage. Holding onto deeply ingrained, and pervasive beliefs is not quite so mentally taxing—indeed, I have argued that it may very well to be rather easy in this case.

Admittedly, conservationism has one significant shortcoming: it recommends false beliefs. Not only that, but it recommends *inconsistent* beliefs. There is a strong presumption against believing that p when p is known to be false, and there is perhaps a stronger presumption still against believing *both* that p and that $\neg p$. However, I have argued that there is a powerful case in favour of overriding these presumptions in the event of a moral error theory—especially once we appreciate that the alternatives available leave much to be desired.

That said, it may be difficult to develop a powerful case for conservationism in other error-ridden discursive domains. When fictionalist attitudes do not introduce significant practical obstacles, it will often be preferable to engage in pretence rather than commit the cardinal sin of intentionally believing falsehoods. Generally speaking, it will always be wise to exercise some degree of epistemic caution; the relevant pragmatic justification for preserving false beliefs must be suitably strong, lest we risk becoming too epistemically promiscuous.

Moreover, maintaining inconsistent beliefs may sometimes be harmful. While I have suggested that holding onto our false moral beliefs is unlikely to render us epistemological wrecks, the same may not be true of Plato's noble lies. Preposterous false beliefs will sometimes be unstable in the long-run, and more likely to impact upon our epistemic well-being. Conservationism will not be a very promising option under such circumstances.

Conclusion

Life after moral error theory

My aim in this thesis was to determine what we ought to do with our moral discourse (and our moral practices more generally) in light of the moral error theory. I began in the introduction by familiarising the reader with error theories and the solutions that are often proposed in response to the WNQs that accompany them. The task for Chapter 1 was to identify the background meta-ethical assumptions that underwrite the moral error theory, to explain the arguments available for it, and to specify which variety of the position would form my background assumption in the remainder of the thesis. Here, I aligned myself with Joyce's (2001) development of Mackie's error theory, which derives from concerns having to do with the special kind of reasons that morality purports to supply.

The goal of Chapter 2 was to (i) clarify and (ii) motivate the project of addressing the WNQ for moral discourse. Regarding (i), I specified that the WNQ is a collective, normative question; it is the question regarding what we, as a linguistic community (or, at least most of us with similar concerns), ought to do with our moral discourse if we believe that the moral error theory is true. It was also necessary here to address an important challenge, according to which moral error theorists are committed to error theories about other sectors of normative discourse as well. Doing so was needed to earn the right to normative language for the remainder of the work. Regarding (ii), I argued that our believing the moral error theory gives rise to a tension between various kinds of values and interests that most of us share. In order to address this tension, we needed to answer the WNQ.

The remainder of the thesis was devoted to finding an appropriate solution to the WNQ for moral discourse. An additional goal was to identify the features in virtue of which those proposals that came up short were unfitting solutions to our WNQ, and to explore what implications (if any) this may have for analogous WNQs in other discursive domains.

In chapter 3, I argued that moral abolitionism was the wrong response to our WNQ; for it involved surrendering many of the desirable practical goods that our moral practices provide. *Pace* the abolitionist, the harms of engaging in moral practice do not plausibly outweigh the benefits. But even if they did, these harms are certainly not unavoidable. Though morality can no doubt be put to bad use, we can work to circumvent this by improving ourselves as critical thinkers, interlocutors, and empathisers. Abolitionism also seemed likely to be an infeasible option going forward; ridding ourselves of our moral practices may very well be something that we cannot do. Given these shortcomings of moral abolitionism, I proposed that generally speaking, abolitionism is likely to make for a fitting solution to a WNQ when it is feasible, and preserving the discourse is likely to be on-balance costly. These conditions are plausibly met in the case of phlogiston discourse and talk of witches.

In chapter 4, I argued that revisionism was also an unfitting solution to our WNQ. The revisionist's schmoralities seemed to make for rather poor stand-ins for morality. This is because the action-guiding function of morality seems intimately tied to its problematic conceptual commitments; in purging moral discourse of a commitment to categorical reasons, we rob it of its distinctive practical force. In addition, I suggested that there was a deeper explanation as to why moral revisionism was an unfitting solution to our WNQ: morality (unlike science) is not plausibly a domain in which concepts can be substantially modified and continue to be put to good use. Scientific concepts are far more amenable to (fruitful) modification than moral ones, and this, I argued, is owing to the distinctive functions of scientific discourse. In light of this, I suggested that in general, revisionist proposals are more likely to be fitting in the scientific domain (e.g., species discourse) than in domains where problematic conceptual commitments are needed for a discourse to serve its core (or 'proper') functions (e.g., Santa discourse).

In chapter 5, I argued that moral fictionalism, though more promising than abolitionism and revisionism, was not an entirely satisfactory solution either. The central issue here concerned the attitudes that the fictionalist intended to substitute for our moral beliefs. Beliefs integrate with our behaviour and the rest of our psychology in distinctive ways; unlike make-beliefs, they are not highly overridable, and they tend to give rise to particular sorts of emotions and behaviour. It is in virtue of this that *moral* beliefs play a valuable role in guiding our actions. The central problem for the fictionalist's proposal was that it carried the risk of surrendering the distinctive benefits associated with moral beliefs. I concluded chapter 5 by suggesting that fictionalist proposals are likely to be more appropriate for discursive domains that are not intimately tied to guiding action and shaping behaviour. This prognosis did not cast doubt upon the prospects of mathematical fictionalism, but it did suggest that religious fictionalism may not be viable.

In chapters 6 and 7, I argued that conservationism is the most fitting solution to the WNQ for moral discourse. Though this proposal involves intentionally cultivating false beliefs, it is my contention that the presumption against doing so can be overridden in the event of a moral error theory—especially given that the alternatives available to us don't carry the same practical promise. Conservationism is not vulnerable to the most pressing problems for its rivals. Since the conservationist proposes to retain our moral practices, she does not risk surrendering the many desirable practical goods that they provide. She also resists purging moral discourse of its problematic conceptual commitments. Thus, there is no risk of losing the distinctive benefits associated with conceiving of moral requirements as categorically authoritative. And since she recommends believing moral propositions, there is no risk of losing the distinctive benefits associated with moral beliefs either.

If we are conservationists, then life after moral error theory will not be all that bad. Indeed, life will go on much the same as before. Life will not go on *exactly* the same as before, perhaps. But did we ever really expect it to?

Bibliography

Adams, D. (2014). *Dirk Gently's Holistic Detective Agency*. Simon and Schuster.

Allport, G. W. and Kramer, B. (1946). 'Some roots of prejudice.' *Journal of Psychology* **22**: 9-39.

Allport, G. W. and Ross, J. M. (1967). 'Personal Religious Orientation and Prejudice.' *Journal of Personality and Social Psychology* **5** (4): 432-443.

Alston, W. (1989). *Epistemic Justification*. Ithaca, NY: Cornell University Press.

Altemeyer, B. and Hunsberger, B. (1992). 'Authoritarianism, religious fundamentalism, quest, and prejudice.' *The International Journal for the Psychology of Religion* **2** (2): 113-133.

Alvarez, M. (2010). *Kinds of Reasons*. Oxford: Oxford University Press.

Anscombe, G. E. M. (1957). *Intention*. Oxford: Basil Blackwell.

———— (1958). 'Modern Moral Philosophy.' *Philosophy* **33** (124): 1-19.

Armstrong, D. M. (2006). 'The Scope and Limits of Human Knowledge.' *Australasian Journal of Philosophy* **84** (2): 159-66.

Balaguer, M. (2009). 'Fictionalism, Theft, and the Story of Mathematics.' *Philosophia Mathematica* **17**: 131–62.

Bales, K. (1999). *Disposable People: New Slavery in the Global Economy*. Berkeley: University of California Press.

Batson, C. D. and Ventis, W. L. (1982). *The religious experience: A social-psychological perspective*. New York: Oxford University Press.

Bealer, G. (1998). 'Intuition and the Autonomy of Philosophy.' In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. M. DePaul and W. Ramsey, Eds. Lanham, MD: Rowman & Littlefield, 201–239.

Bedke, M. S. (2010). 'Might all normativity be queer?' *Australasian Journal of Philosophy* **88** (1): 41-58.

Bentham, J. (1792). 'Anarchical Fallacies; Being an Examination of the Declaration of Rights Issued during the French Revolution.' Republished in *The Works of Jeremy Bentham, Volume 2*. J. Bowring, Ed. (1843). Edinburgh: William Tait.

Blackburn, S. (1984). *Spreading the Word*. Oxford: Oxford University Press.

———— (1993). 'Errors and the Phenomenology of Value.' In his *Essays in Quasi-Realism*. NY: Oxford University Press, 149-165.

———— (1998). *Ruling Passions*. Oxford: Clarendon Press.

Bloomfield, P. (2009). 'Archimedeanism and Why Metaethics Matters.' In *Oxford Studies in Metaethics, Volume 4*. R. Shafer-Landau, Ed. Oxford: Oxford University Press, 283-302.

Boghossian, P. A. (2003). 'The normativity of content.' *Philosophical Issues* **13** (1): 31–45.

Boghossian, P. and Velleman, D. (1989). 'Colour as a Secondary Quality.' *Mind*, **98** (389): 81-103.

Bourg, T. (1996). 'The Role of Emotion, Empathy, and Text Structure in Children's and Adults' Narrative Text Comprehension.' In *Empirical Approaches to Literature and Aesthetics*. R.J. Kreuz and M.S MacNealy, Eds. Norwood, NJ: Ablex, 241-60.

Bourget, D. and Chalmers, D. (2014). 'What do philosophers believe?' *Philosophical Studies* **170** (3): 465-500.

Boyd, R. N. (1983). 'On the Current Status of the Issue of Scientific Realism.' *Erkenntnis* **19**: 45–90.

———— (1988). 'How to be a Moral Realist.' In *Essays on Moral Realism*. G. Sayre-McCord, Ed. Ithaca: Cornell University Press, 181-228.

Boyd, R., Gintis, H., Bowles, S. and Richerson, P. J. (2003). 'The Evolution of Altruistic Punishment.' *Proceedings of the National Academy of Sciences* **100**: 3531-3535.

Braddon-Mitchell, D. (2003). 'Qualia and Analytical Conditionals.' *The Journal of Philosophy* **100**: 111-135.

———— (2005). 'The Subsumption of Reference.' *British Journal of Philosophy of Science* **56**: 157-178.

———— (2006). 'Believing Falsely Makes It So.' *Mind* **115** (460): 833-865.

Brandt, R. (1979). *A Theory of the Good and the Right*. Oxford: Clarendon Press.

Bratman, M. E. (1992). 'Practical Reasoning and Acceptance in a Context.' *Mind* **101** (401): 1-15.

Brennan, G. and Southwood, N. (2007). 'Feasibility in Action and Attitude.' In Hommage á Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz. Rønnow-Rasmussen, T., Petersson, B., Josefsson, J. and Egonsson, D. Eds. Available online at www.fil.lu.se/hommageawlodek.

Brigandt, I. (2013). 'A critique of Dave Chalmers' and Frank Jackson's account of concepts.' *Protosociology* **30**: 63-88.

Brink, D. (1984). 'Moral Realism and the Sceptical Arguments from Disagreement and Queerness.' *Australasian Journal of Philosophy* **62** (2): 111-125.

———— (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Brocke, S. (2002). 'Fictionalism about Fictional Characters.' *Noûs* **36** (1): 1-21.

Brogaard, B. (2006). 'Two Modal –Isms: Fictionalism and Ersatzism.' *Philosophical Perspectives* **20**: 77-94.

Broome, J. (1999). 'Normative Requirements.' Ratio, 12: 398–419.

———— (2004). 'Reasons.' In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. R. J. Wallace, P. Pettit, S. Scheffler and M. Smith, Eds. NY: Oxford University Press, 28–55.

———— (2005). 'Does rationality give us reasons?' *Philosophical Issues* **15** (1): 321-337.

———— (2007). 'Is rationality normative?' *Disputatio* **2** (23): 161-178.

———— (2013). *Rationality Through Reasoning*. Malden, MA: Wiley Blackwell

Brown, P. (2013). 'The Possibility of Morality.' *Philosophical Studies* **163** (3): 627-636.

Buckwalter, W., Rose, D. and Turri, J. (2015). 'Belief through Thick and Thin.' *Noûs* **49** (4): 748–775.

Bukoski, M. (2016). 'A Critique of Smith's Constitutivism.' *Ethics* **127**: 116–146.

Burge, T. (1979). 'Individualism and the Mental.' *Midwest Studies in Philosophy* **4**: 73–121.

Burgess, J. (1983). 'Why I Am Not a Nominalist.' *Notre Dame Journal of Formal Logic* **24** (1): 93-105.

Burgess, J. P. (2007). 'Against Ethics.' *Ethical Theory and Moral Practice* **10**: 427–439.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. NY: Oxford University Press.

———— (2002a). 'On Sense and Intension.' *Noûs* **36** (16): 135–182.

———— (2002b). 'The components of content.' In *Philosophy of Mind: Classical and Contemporary Readings*. D. Chalmers, Ed. Oxford: Oxford University Press, 608–633.

———— (2008). 'Two-Dimensional Semantics.' In *Oxford Handbook of the Philosophy of Language*. E. Lepore and B. Smith, Eds. Oxford: Oxford University Press, 574-606.

———— (2012). *Constructing the World*. Oxford: Oxford University Press.

Chalmers, D. J. and Jackson, F. (2001). 'Conceptual analysis and reductive explanation.' *Philosophical Review* **110** (3): 315-61.

Chisholm, R.M. (1957). *Perceiving: A Philosophical Study*. NY: Cornell University Press.

Chuard, P. and Southwood, N. (2009). 'Epistemic Norms without Voluntary Control.' *Noûs* **43** (4): 599–632.

Church, J. (2002). 'Taking it to Heart: What Choice Do We Have?' *Monist* **85** (3): 361-380.

Churchland, P. (1981). 'Eliminative Materialism and the Propositional Attitudes.' *Journal of Philosophy* **78**: 67–90.

Clifford, W. K. (1877/1999). 'The Ethics of Belief.' In *The Ethics of Belief and Other Essays*. T. J. Madigan, Ed. Amherst, MA: Prometheus, 70–96.

Cohen, A. J. (2004). 'What Toleration Is.' *Ethics* **115** (1): 68-95.

Cohen, G. A. (2000). *If You're an Egalitarian, How Come You're So Rich?* Cambridge, MA: Harvard University Press.

Cohen, L. J. (1989). 'Belief and Acceptance.' *Mind* **98**: 367-89.

———— (1992). *An Essay on Belief and Acceptance*. NY: Clarendon Press.

Cole, P. (2006). *The Myth of Evil: Demonizing the Enemy*. Westport, Connecticut: Praeger.

Coons, C. (2011). 'How to prove that some acts are wrong (without using substantive moral premises).' *Philosophical Studies* **155** (1): 83-98.

Coplan, A. (2004). 'Empathic Engagement with Narrative Fictions.' *The Journal of Aesthetics and Art Criticism* **62** (2): 141-152.

Copp, D. (1990). 'Explanation and Justification in Ethics.' *Ethics* **100** (2): 237-258.

———— (2006). 'Introduction: Metaethics and Normative Ethics.' In *The Oxford Handbook of Ethical Theory*. D. Copp, Ed. NY: Oxford University Press, 3-35.

———— (2009). 'Toward A Pluralist and Teleological Theory on Normativity.' *Philosophical Issues* **19**: 21-37.

Cowie, C. (2014a). 'Good News for Moral Error Theorists: A Master Argument Against Companions in Guilt Strategies.' *Australasian Journal of Philosophy* **94** (1): 115-30.

————. (2014b). 'In defence of instrumentalism about epistemic normativity.' *Synthese* **191**: 4003–4017.

Cuneo, T. (2007). *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.

Cuneo, T. and Christy, S. (2011). 'The Myth of Moral Fictionalism'. In *New Waves in Metaethics*. M. Brady, Ed. Basingstoke: Palgrave MacMillan, 85-102.

Currie, G. (2000). 'Imagination, Delusion, and Hallucinations.' *Mind & Language* **15** (1): 168–183.

Currie, G. and Sterelny, S. (2000). 'How to think about the modularity of mind-reading.' *Philosophical Quarterly* **50** (199): 145-160.

Cushman, F. A., Young, L., and Hauser, M. D. (2006). 'The role of conscious reasoning and intuitions in moral judgment: testing three principles of harm.' *Psychological Science* **17** (12): 1082-1089.

Daly, C. J. (2008). 'Fictionalism and the Attitudes.' *Philosophical Studies* **139** (3): 423-440.

Daly, C. and Liggins, D. (2010). 'In Defence of Error Theory.' *Philosophical Studies* **149**: 209-230.

Dancy, J. (2000). *Practical Reality*. Oxford: Clarendon Press.

———— (2004). *Ethics Without Principles*. Oxford: Clarendon University Press.

Darwall, S. (1983). *Impartial Reason*. Ithaca: Cornell University Press.

———— (1992). 'Internalism and Agency.' *Philosophical Perspectives* **6**: 155-174.

———— (2006). 'How Should Ethics Relate to (the Rest of) Philosophy? Moore's Legacy.' In *Metaethics after Moore*. T. Horgan and M. Timmons, Eds. Oxford: Clarendon Press, 17-38.

Davidson, D. (1963). 'Actions, Reasons, and Causes.' Reprinted in *Essays on Actions and Events* (1980). Oxford: Clarendon Press.

Dennett, D. C. (1986). 'The Moral First Aid Manual.' *The Tanner Lectures on Human Values*, University of Michigan.

———— (2006). 'Higher-order truths about chmess.' *Topoi* 39–41.

Doris, J. and Plakias, A. (2008). 'How to Argue About Disagreement: Evaluative Diversity and Moral Realism.' In *Moral Psychology Volume 2, The Cognitive Science of Morality: Intuition and Diversity*, W. Sinnott-Armstrong, Ed. Cambridge MA: MIT Press, 303-331.

Dorr, C. (2002). 'Non-cognitivism and Wishful Thinking.' *Noûs* **36** (1): 97–103.

Dorr, C. and Rosen, G. (2002). 'Composition as a Fiction.' In *The Blackwell Guide to Metaphysics*. R. Gale, Ed. Oxford: Basil Blackwell, 151–74.

Dreier, J. (1997). 'Humean Doubts about the Practical Justification of Morality.' In *Ethics and Practical Reason*. G. Cullity and B. Gaut, Eds. Oxford: Oxford University Press, 81-100.

———— (2002). 'Meta-Ethics and Normative Commitment.' *Philosophical Issues*, **12**: 241-263.

———— (2010). 'Mackie's Realism: Queer Pigs and the Web of Belief.' In *A World Without Values: Essays on John Mackie's Moral Error Theory*. R. Joyce and S. Kirchin, Eds. Dordrecht: Springer, 71-86.

Dummett, M. (1973). *Frege: Philosophy of Language*. London: Duckworth.

Dworkin, R. (1996). 'Objectivity and Truth: You'd Better Believe It.' *Philosophy and Public Affairs*, **25**: 87-139.

———— (2011). *Justice for Hedgehogs*. Cambridge, MA: Harvard University Press.

Egan, A. (2008). 'Seeing and believing: perception, belief formation and the divided mind.' *Philosophical Studies* **140**: 47–63.

Eklund, M. (2015). 'Fictionalism.' In *The Stanford Encyclopedia of Philosophy*. (Winter 2015 Edition.) E.N. Zalta, Ed. URL = <http://plato.stanford.edu/archives/win2015/entries/fictionalism/>.

Elga, A. and Rayo, A. (2015). 'Fragmentation and information access.' *Unpublished Manuscript*.

Enoch, D. (2011). *Taking Morality Seriously*. Oxford: Oxford University Press.

Eshleman, A. (2005). 'Can an atheist believe in God?' *Religious Studies* **41** (2): 183-199.

Fantl, J. (2006). 'Is Metaethics Morally Neutral?' *Pacific Philosophical Quarterly* **87**: 24–44.

Fahrbach, L. (2011). 'How the Growth of Science Ends Theory Change.' *Synthese* **180**: 139–155.

Feldman, R. (2000). 'The Ethics of Belief.' *Philosophy and Phenomenological Research* **60** (3): 667-95.

Feltz, A. & Cokely, E.T. (2008). 'The fragmented folk: More evidence of stable individual differences in moral judgments and folk intuitions.' In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. B.C. Love, K. McRae and V.M. Sloutsky, Eds. Austin, TX: Cognitive Science Society, 1771-76.

Ferguson, T. J., Stegge, H., and Damhuis, I. (1991). 'Children's understanding of guilt and shame.' *Child Development* **62** (4): 827-839.

Field, H. (1973). 'Theory change and the indeterminacy of reference.' *Journal of Philosophy* **70** (14): 462-481.

———— (1980). *Science Without Numbers*. Princeton, NJ: Princeton University Press.

———— (1989). *Realism, Mathematics, and Morality*. NY: Blackwell.

———— (2000). 'A prioricity as an evaluative notion.' In *New essays on the a priori*. P. Boghossian and C. Peacocke, Eds. Oxford: Oxford University Press, 117-149.

Fine, C. (2006). 'Is the emotional dog wagging its rational tail, or chasing it?' *Philosophical Explorations* **9** (1): 83-98.

Finlay, S. (2008). 'The Error in the Error Theory.' *Australasian Journal of Philosophy* **86** (3): 347–369.

Finlay, S. and Schroeder, M. (2015). 'Reasons for Action: Internal vs. External.' In *The Stanford Encyclopedia of Philosophy*. (Winter 2015 Edition). E.N. Zalta, Ed. URL = <http://plato.stanford.edu/archives/win2015/entries/reasons-internal-external/>.

Fogle, T. (2000). 'The Dissolution of Protein Coding Genes in Molecular Biology.' In *The Concept of the Gene in Development and Evolution*. P. Beurton, R. Falk and H. J. Rheinberger, Eds. Cambridge: Cambridge University Press.

Foley, R. (1987). *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press.

Foot, P. (1967). 'The problem of abortion and the doctrine of the double effect.' *Oxford Review* **5**: 5–15.

———— (1972). 'Morality as a System of Hypothetical Imperatives.' *The Philosophical Review* **81**: 305-16.

Fraser, B. (2014). 'Moral error theories and folk metaethics.' *Philosophical Psychology* **27** (6): 789–806.

———— (2017). 'Moral Mismatch and Abolition.' In *The Cambridge Handbook of Evolutionary Ethics*. M. Ruse and R. J. Richards, Eds. Cambridge, UK: Cambridge University Press, 158-173.

Fricker, M. (2014). 'What's the Point of Blame? A Paradigm Based Explanation.' *Noûs* **50** (1): 165–183.

Friend, S. (2008). 'Hermeneutic moral fictionalism as an anti-realist strategy.' *Philosophical Books*, 49(1): 14–22.

Gale, R. M. (1999). 'William James and the Willfulness of Belief.' *Philosophy and Phenomenological Research* **59** (1): 71-91.

Garner, R. (1990). 'On the Genuine Queerness of Moral Properties and Facts.' *Australasian Journal of Philosophy* **68** (2): 137-46.

———— (1993). 'Are Convenient Fictions Harmful to Your Health?' *Philosophy East and West* **43** (1): 87-106.

———— (1994). *Beyond Morality*. Philadelphia, PA: Temple University Press.

———— (2007). 'Abolishing Morality.' *Ethical Theory and Moral Practice* **10** (5): 499-513.

Gauthier, D. (1967). 'Morality and Advantage.' *Philosophical Review* **76** (4): 460-475.

———— (1986). *Morals by Agreement*. Oxford: Clarendon Press.

Geach, P. (1965). 'Assertion.' *Philosophical Review* **74**: 449-465.

Gendler, T. S. (2000). 'The Puzzle of Imaginative Resistance.' *Journal of Philosophy* **97** (2): 55-81.

———— (2008). 'Alief and Belief.' *Journal of Philosophy* **105** (10): 634-663.

Gernsbacher, M. A., Goldsmith, H. H. and Robertson, R. R. W. (1992). 'Do Readers Mentally Represent Characters' Emotional States?' *Cognition and Emotion* **6**: 89-111.

Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.

———— (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Gilabert, P. and Lawford-Smith, H. (2012). 'Political Feasibility: A Conceptual Exploration.' *Political Studies* **60** (4): 809-25.

Gilbert, D., Tafarodi, R., and Malone, P. (1993). 'You can't not believe everything you read.' *Journal of Personality and Social Psychology* **65**: 221–233.

Godfrey-Smith, P. (1998). *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.

Goodwin, G. and Darley, J. (2008). 'The psychology of metaethics: Exploring objectivism.' *Cognition* **106** (3): 1339-1366.

———— (2010). 'The perceived objectivity of ethical beliefs: Psychological findings and implications for public policy.' *Review of Philosophy and Psychology* **1** (2): 161-188.

———— (2012). 'Why are some moral beliefs seen as more objective than others?' *Journal of Experimental Social Psychology* **48** (1): 250-256.

Greene, J. (2002). *The Terrible, Horrible, No Good, Very Bad Truth about Morality and What to Do About it*. Dissertation, Princeton University.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. and Cohen, J. D. (2004). 'The neural bases of cognitive conflict and control in moral judgment.' *Neuron* **44** (2): 389–400.

Griffiths, P. E. (1997). *What Emotions Really Are.* University of Chicago Press.

Griffiths, P. E. and Stotz, K. (2007). 'Gene.' In *The Cambridge Companion to the Philosophy of Biology*. D.L. Hull and M. Ruse, Eds. Cambridge: Cambridge University Press, 85–102.

Haidt, J. (2001). 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment.' *Psychological Review* **108** (4): 814-834.

———— (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. NY: Pantheon.

Hale, B. (1986). 'The Compleat Projectivist.' *Philosophical Quarterly* **36** (142): 65-84.

Hall, D. L., Matz, D. C. and Wood, W. (2010). 'Why don't we practice what we preach? A meta-analytic review of religious racism.' *Personality and Social Psychology Review* **14** (1) 126–139.

Hampton, J. (1995). 'Does Hume Have an Instrumental Conception of Practical Reason? *Hume Studies* **21** (1): 57-74.

———— (1998). *The Authority of Reason*. Cambridge: Cambridge University Press.

Hanlon, C. (2003). 'O.S. Fowler and Hereditary Descent.' In *A House Divided: The Antebellum Slavery Debates in America, 1776-1865.* M.I. Lowance, Jr., Ed. Princeton, NJ: Princeton University Press.

Hare, C. (2009). *On Myself, and Other, Less Important Subjects*. Princeton: Princeton University Press.

Hare, R. M. (1952). *The Language of Morals*. Oxford: Clarendon Press.

Harman, G. (1977). *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press.

Harman, G. and Thomson, J. J. (1996). *Moral Relativism and Moral Objectivity*. Cambridge, MA: Blackwell.

Harris, P. (2000). *The Work of the Imagination*. Oxford: Blackwell.

Hauser, M. D., Cushman, F. A., Young, L., Jin, R. and Mikhail, J. M. (2007). 'A dissociation between moral judgment and justification.' *Mind and Language* **22** (1): 1-21.

Heathwood, C. (2009). 'Moral and Epistemic Open-Question Arguments.' *Philosophical Books* **50**: 83-98.

Heil, J. (1984). 'Doxastic Incontinence.' *Mind* **93** (369): 56-70.

Held, V. (2001). 'The Language of Evil.' In *Controversies in Feminism*. J. Sterba, Ed. Oxford: Rowman & Littlefield, 107–110.

Hills, A. (2009). 'Moral Testimony and Moral Epistemology.' *Ethics* **120**: 94-127.

Hinckfuss, I. (1987). *The Moral Society: Its Structure and Effects*, Discussion Papers in Environmental Philosophy. Canberra: Australian National University.

———— (1993). 'Suppositions, Presuppositions and Ontology.' *Canadian Journal of Philosophy* **23** (4): 595-617.

Hoffman, M. L. (2015). 'Empathy, justice, and social change.' In *Empathy and Morality*. H. L. Maibom, Ed. NY: Oxford University Press, 71-96.

Hopkins, R. (2007). 'What Is Wrong with Moral Testimony?' *Philosophy and Phenomenological Research* **74**: 611–34.

Horwich, P. (1998). *Truth* (2nd ed.) Oxford: Oxford University Press.

———— (1991). 'On the Nature and Norms of Theoretical Commitment.' *Philosophy of Science* **58** (1): 1-14.

Hubin, D. (1999). 'What's Special about Humeanism?' *Noûs* **33** (1): 30-45.

———— (2001). 'The groundless normativity of instrumental rationality.' *Journal of Philosophy* **98** (9): 445-468.

Huddleston, A. (2012). 'Naughty beliefs.' *Philosophical Studies* **160**: 209–222.

Huemer, M. (2005). *Ethical Intuitionism*. NY: PalgraveMacmillan.

Hull, D. (1965). 'The Effect of Essentialism on Taxonomy: Two Thousand Years of Stasis.' *British Journal for the Philosophy of Science* **15**: 314–326, **16**: 1–18.

———— (1978). 'A Matter of Individuality.' *Philosophy of Science* **45**: 335–360.

Hume, D. (1740/1978). *A Treatise of Human Nature*. L. A. Selby-Bigge, Ed. Oxford: Clarendon Press.

Hunsberger, B. (1996). 'Religious fundamentalism, right-wing authoritarianism, and hostility toward homosexuals in non-Christian religious groups.' *The International Journal for the Psychology of Religion* **6** (1): 39-49.

Hussain, N. (2004). 'The Return of Moral Fictionalism.' *Philosophical Perspectives* **18** (1):149–188.

Ingram, S. (2015). 'After Moral Error Theory, After Moral Realism.' *The Southern Journal of Philosophy* **53** (2): 227-248.

———— (2017). 'Epistemology shmepistemology: moral error theory and epistemic expressivism.' *Inquiry* **doi:** 10.1080/0020174X.2017.1291362.

Jackson, F. (1994). 'Armchair metaphysics.' In *Philosophy in Mind: The Place of Philosophy in the Study of Mind*. M. Michaelis and J. O'Leary-Hawthorne, Eds. Dordrecht: Kluwer, 22–42.

———— (1998a). *From Metaphysics to Ethics*. Oxford: Oxford University Press.

———— (1998b). 'Reference and Description Revisited.' *Philosophical Perspectives* **12**: 201- 218.

———— (2004). 'Why we need A-intensions.' *Philosophical Studies* **118**: 257–277.

Jackson, F. & Pettit, P. (1990). 'In Defence of Folk Psychology.' *Philosophical Studies* **59**: 31-54.

Jackson, F., Stich, S., and Mason, K. (2009). 'Folk Psychology and Tacit Theories: A Correspondence between Frank Jackson, and Steve Stich and Kelby Mason.' In *Conceptual Analysis and Philosophical* Naturalism. D.Braddon-Mitchell and R. Nola, Eds. Cambridge, MA: MIT Press, 45-97.

Jaquet, F. and Naar, H. (2016). 'Moral Beliefs for the Error Theorist?' *Ethical Theory and Moral Practice* **19**: 193–207.

Jenkins, C.S. (2007). 'Entitlement and rationality.' *Synthese* **157**: 25–45.

Joyce, R. (2001). *The Myth of Morality*. Cambridge, MA: Cambridge University Press.

———— (2005). 'Moral Fictionalism.' In *Fictionalism in Metaphysics*. M. Kalderon, Ed. Oxford: Oxford University Press, 287-313.

———— (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.

———— (2008). 'Morality, schmorality.' In *Morality and Self-Interest*. P. Bloomfield, Ed. Oxford: Oxford University Press, 51–75.

———— (2011). 'The Error In 'The Error In The Error Theory'.' *Australasian Journal of Philosophy* **89** (3): 519-534.

———— (2012). 'Metaethical pluralism: How both moral naturalism and moral skepticism may be permissible positions.' In *Ethical Naturalism: Current debates*. S. Nuccetelli and G. Seay, Eds. Cambridge: Cambridge University Press, 89–109.

———— (2014). 'Taking moral skepticism seriously.' *Philosophical Studies* **168**: 843-851.

———— (2016). 'Reply to 'On the Validity of the Simple Argument for Moral Error Theory'.' *International Journal of Philosophical Studies* **24** (4): 518-522.

Joyce, R. and Kirchin, S. (2010). 'Introduction.' In *A World Without Values: Essays on John Mackie's Moral Error Theory*. R. Joyce and S. Kirchin, Eds. Springer: Dordrecht, ix-xxiv.

Kahane, G. (2016). 'If Nothing Matters.' *Noûs* **50** (2): 327-353.

Kalderon, M. (2005). *Moral Fictionalism*. NY: Oxford University Press.

Kalf, W. F. (2013). 'Moral Error Theory, Entailment and Presupposition.' *Ethical Theory and Moral Practice* **16** (5): 923-937.

Kant, I. (1781). *Critique of Pure Reason*. P. Guyer and A. Wood, Eds. Trans. (1998). Cambridge: Cambridge University Press.

Kaupinnen, A. (2007). 'The Rise and Fall of Experimental Philosophy.' *Philosophical Explorations* **10** (2): 95–118.

Kearns, S. and Star, D. (2009). 'Reasons as Evidence.' In *Oxford Studies in Metaethics, Volume 4*. R. Shafer-Landau, Ed. Oxford: Oxford University Press, 215-42.

Kelly, T. (2002). 'The Rationality of Belief and Some Other Propositional Attitudes.' *Philosophical Studies* **110** (2): 163-96.

———— (2003). 'Epistemic rationality as instrumental rationality: A critique.' *Philosophy and Phenomenological Research* **66** (3): 612-640.

Kirchin, S. (2010). 'A Tension in the Moral Error Theory.' In *A World Without Values: Essays on John Mackie's Moral Error Theory*. R. Joyce and S. Kirchin, Eds. Dordrecht: Springer, 167-182.

Kirkpatrick, L. A. (1993). 'Fundamentalism, Christian orthodoxy and intrinsic religious orientation as predictors of discriminatory attitudes.' *Journal for the Scientific Study of Religion* **32** (3): 256-268.

Kitcher, P. (2011). *The Ethical Project*. Cambridge, MA: Harvard University Press.

Klein, C. (2011). 'The Dual Track Theory of Moral Decision-Making: a Critique of the Neuroimaging Evidence.' *Neuroethics*, **4** (2): 143-162.

Knobe, J. and Nichols, S. (2007). 'An experimental philosophy manifesto.' In *Experimental philosophy*. J. Knobe, and S. Nichols, Eds. Oxford: Oxford University Press, 3-14.

Köhler, S. and Ridge, M. (2013). 'Revolutionary Expressivism.' *Ratio* **26** (4): 428–44.

Kolodny, N. (2005). 'Why be rational?' *Mind* **114** (455): 509-563.

Korsgaard, C. (1996a). 'Skepticism about practical reason.' In her *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

———— (1996b). *The Sources of Normativity*. Cambridge: Cambridge University Press.

———— (1997). 'The Normativity of Instrumental Reason.' In *Ethics and Practical Reason*. G. Cullity and B. Gaut, Eds. Oxford: Clarendon Press, 215-254.

Kriegel, U. (2012). 'Moral Motivation, Moral Phenomenology, And The Alief/Belief Distinction.' *Australasian Journal of Philosophy* **90** (3): 469–486.

Kripke, S. (1972). 'Naming and Necessity.' In *Semantics of Natural Language*. D. Davidson and G. Harman, Eds. Dordrecht: Reidel.

Kroon F. (2011). 'Fictionalism in Metaphysics.' *Philosophy Compass* **6** (11): 786-803.

Kornblith, H. (1993). 'Epistemic normativity.' *Synthese* **94**: 357–376.

———— (2001). *Knowledge and Its Place in Nature*. Oxford: Oxford University Press.

Lahti, D. C. and Weinstein, B. S. (2005). 'The better angels of our nature: group stability and the evolution of moral tension.' *Evolution and Human Behavior* **26**: 47–63.

Lalumera, E. (2014). 'On the Explanatory Value of the Concept-Conception Distinction.' *Rivista Italiana di Filosofia del linguaggio* **8** (3): 73-81.

Lang, G. (2011). 'How Far Can You Go With Quietism?' *Problema*, **4**: 3-37.

Laudan, L. (1981). 'A Confutation of Convergent Realism.' *Philosophy of Science* **48**: 19–49.

———— (1990). 'Normative naturalism.' *Philosophy of Science* **57**: 44-59.

———— (1991). 'Aimless epistemology?' *Studies in History and Philosophy of Science* **21**: 315-22.

Laurence, S. and Margolis, E. (2003). 'Concepts and Conceptual Analysis.' *Philosophy and Phenomenological Research* **67** (2): 253–282.

Leary, S. (2016). 'In Defense of Practical Reasons for Belief.' *Australasian Journal of Philosophy* **doi:** 10.1080/00048402.2016.1237532.

Leite, A. (2007). 'Epistemic Instrumentalism and Reasons for Belief: A Reply to Tom Kelly's "Epistemic Rationality as Instrumental Rationality: A Critique".' *Philosophy and Phenomenological Research* **75** (2): 456-464.

Leng, M. (2010). *Mathematics and Reality*. NY: Oxford University Press.

Lenman, J. (2008). 'Review of Terence Cuneo, The Normative Web: An Argument for Moral Realism.' *Notre Dame Philosophical Reviews*. URL = < http://ndpr.nd.edu/news/the-normative-web-an-argument-for-moral-realism/>.

———— (2013). 'Ethics without Errors.' *Ratio* **26**: 391-409.

Leslie, A. (1987). 'Pretense and Representation: The Origins of "Theory of Mind".' *Psychological Review* **94** (4): 412-426.

Lewis, D. (1966). 'An Argument for the Identity Theory.' *Journal of Philosophy* **63**: 17–25.

———— (1970). 'How to Define Theoretical Terms.' *Journal of Philosophy* **67** (13): 427-446.

———— (1974). 'Radical Interpretation.' *Synthese* **23**: 331–344.

———— (1978). 'Truth in Fiction.' *American Philosophical Quarterly* **15** (1): 37-46.

———— (1981). 'Index, Context, and Content.' In *Philosophy and Grammar*. S. Kanger and S. Ohlman, Eds. Dordrecth: Reidel, 79–100.

———— (1982). 'Logic for Equivocators.' *Noûs* **16** (3): 431-441.

———— (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

———— (1989). 'Dispositional Theories of Value.' *The Proceedings of the Aristotelian Society* **63**: 113-137.

———— (1994). 'Reduction of Mind.' In *A Companion to the Philosophy of Mind*. S. Gutteplan, Ed. Oxford: Blackwell, 412-431.

———— (2005). 'Quasi-Realism is Fictionalism.' In *Fictionalism in Metaphysics*. M. Kalderon, Ed. Oxford: Oxford University Press, 314-321.

Lillehammer, H. (1997). 'Smith on Moral Fetishism.' *Analysis* **57** (3): 187-195.

———— (2002). 'Moral Realism, Normative Reasons, and Rational Intelligibility.' *Erkenntnis* **57** (1): 4-69.

———— (2004). 'Moral Error Theory.' *Proceedings of the Aristotelian Society* **104**: 93-109.

———— (2011). 'Constructivism and the Error Theory.' In *The Continuum Companion to Ethics*. C. Miller, Ed. NY: Continuum, 55-76.

Lockard, M. (2013). 'Epistemic instrumentalism.' *Synthese* **190** (9): 1701-1718.

Loeb, D. (1998). 'Moral Realism and the Argument from Disagreement.' *Philosophical Studies* **90**: 281–303.

———— (2008). 'Moral incoherentism: How to pull a metaphysical rabbit out of a semantic hat.' In *Moral psychology. The cognitive science of morality: Intuition and diversity*. W. Sinnott-Armstrong, Ed. Cambridge, MA: MIT Press.

Luco, A. (2016). 'Non-Negotiable: Why Moral Naturalism Cannot Do Away with Categorical Reasons.' *Philosophical Studies* **173** (9): 2511-28.

Lutz, M. (2014). 'The 'Now What' Problem for error theory.' *Philosophical Studies* **171**: 351-371.

Lycan, W. (1988). *Judgment and Justification*. Cambridge: Cambridge University Press.

MacFarlane, J. (2007). 'Relativism and Disagreement.' *Philosophical Studies* **132**: 17-31.

Mag Uidhir, C. (2013). 'What's So Bad about Blackface?' In *Race, Philosophy, and Film*. D. Flory and M. Bloodsworth-Lugo, Eds. NY: Routledge.

Marušić, B. (2011). 'The Ethics of Belief.' *Philosophy Compass* **6** (1): 33–43.

McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.

———— (2013). 'Acting in the Light of a Fact.' In *Thinking About Reasons: Themes from the Philosophy of Jonathan Dancy*. D. Bakhurst, B. Hooker, & M. O. Little, Eds. Oxford: Oxford University Press, 13–28.

McFarland, S. G. (1989). 'Religious orientations and the targets of discrimination.' *Journal for the Scientific Study of Religion* **28** (3): 324-336.

Machery, E. (2009). *Doing Without Concepts*. Oxford: Oxford University Press.

Mackie, J. L. (1946). A Refutation of Morals.' *Australasian Journal of Psychology and Philosophy*, **24**: 77-90.

———— (1977). *Ethics: Inventing Right and Wrong*. NY: Penguin Books.

Maffie, J. (1990). 'Naturalism and the normativity of epistemology.' *Philosophical Studies* **59**: 333-49.

Mandeville, B. (1714). *The Fable of the Bees: or, Private Vices, Publick Benefits*. London: Printed for J. Roberts.

Margolis, E. and Laurence, S. (2007). 'The Ontology of Concepts —Abstract Objects or Mental Representations?' *Noûs* **41** (4): 561-593.

———— (2014). 'Concepts.' In *The Stanford Encyclopedia of Philosophy*. E.N. Zalta, Ed. URL = <https://plato.stanford.edu/archives/spr2014/entries/concepts/>.

Marks, J. (2013a). *Ethics without Morals: In Defence of Amorality*. NY: Routledge.

———— (2013b). 'Animal Abolitionism Meets Moral Abolitionism: Cutting the Gordian Knot of Applied Ethics.' *Bioethical Inquiry* **10**: 445-455.

Marx, K. (1977). *Karl Marx: Selected writings*. D. McLellan, Ed. Oxford: Oxford University Press.

McGeer, V. (2013). 'Civilising Blame.' In *Blame: Its Nature and Norms*. D.J. Coates and N. Tognazzini, Eds. NY: Oxford University Press, 189-206.

Miller, K. (2010). 'On Contingently Error-Theoretic Concepts.' *American Philosophical Quarterly* **47** (2): 181-190.

Millikan, R. (1989). 'In Defense of Proper Functions.' *Philosophy of Science* **56** (2): 288–302.

Moeller, H. (2009). *The Moral Fool: A Case for Amorality*. NY: Columbia University Press.

Möller, N. and Lillehammer, H. (2015), 'We Can Believe the Error Theory.' *Ethical Theory and Moral Practice* **18** (3): 453-459.

Monteith, M. J., Sherman, J. W. and Devine, P. G. (1998). 'Suppression as a stereotype control strategy.' *Personality and Social Psychology Review* **2**: 63–82.

Moran, R. 'The Expression of Feeling in Imagination.' *Philosophical Review* **103**: 75-106.

Morton, J. and Sampson, E. (2014). 'Parsimony and the Argument from Queerness.' *Res Philosophica* **91** (4): 609-627.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W. and Schaal, B. (1999). 'Preconscious control of stereotype activation through chronic egalitarian goals.' *Journal of Personality and Social Psychology* **77**: 167–84.

Nichols, S. (2006). 'Just the Imagination: Why Imagining Doesn't Behave Like Believing.' *Mind & Language* **21** (4): 459–474.

Nichols, S. and Stich, S. (2000). 'A cognitive theory of pretense.' *Cognition* **74**: 115-147.

Nietzche, F. (1994). *The Genealogy of Morals*. K. Ansell-Pearson, Ed. C. Diethe, Trans. Cambridge: Cambridge University Press.

Nisbett, R. and Cohen, D. (1996). *Culture of Honor: The Psychology of Violence in the South*. Boulder, CO: Westview Press.

Nolan, D. (1997). 'Impossible Worlds: A Modest Approach.' *Notre Dame Journal of Formal Logic* **38** (4): 535-572.

———— (2005). 'Fictionalist Attitudes About Fictional Matters.' In *Fictionalism in Metaphysics*. M. Kalderon, Ed. Oxford: Oxford University Press, 204-233.

———— (2009). 'Platitudes and Metaphysics.' In *Conceptual Analysis and Philosophical Naturalism*. D. Braddon-Mitchell and R. Nola, Eds. Cambridge, MA: MIT Press, 267-300.

———— (2014). 'The Question of Moral Ontology.' *Philosophical Perspectives* **28**: 201-221.

Nolan, D., Restall, G., & West, C. (2002). 'Moral Fictionalism.' *Unpublished manuscript.*

———— (2005). 'Moral Fictionalism versus the Rest.' *Australasian Journal of Philosophy* **83** (3): 307-330.

Oddie, G. and Demetriou, D. (2007). 'The Fictionalist's Attitude Problem.' *Ethical Theory and Moral Practice* **10**: 485–498.

O'Leary-Hawthorne, J. (1994). 'What does van Fraassen's critique of scientific realism show?' *Monist* **77** (1): 128-145.

Olson, J. (2011a). 'In Defence of Moral Error Theory.' In *New Waves in Metaethics*. M. Brady, Ed. London: Palgrave Macmillan, 62–84.

———— (2011b). 'Projectivism and Error in Hume's Ethics.' *Hume Studies* **37** (1): 19-42.

———— (2011c). 'Getting Real about Moral Fictionalism.' In *Oxford Studies in Metaethics*, *Volume 6*. R. Shafer-Landau, Ed. Oxford: Oxford University Press, 181-204.

———— (2011d). 'Error theory and reasons for belief.' In *Reasons for Belief.* A. Reisner and A. Steglich-Petersen, Eds. NY: Cambridge University Press, 75-93.

———— (2014). *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Papineau, D. (1999). 'Normativity and judgement.' *Proceedings of the Aristotelian Society* **73**: 17-41.

Parfit, D. (2011). *On What Matters, Volume II*. Oxford: Oxford University Press

Park, S. (2011). 'A Confutation of the Pessimistic Induction.' *Journal for General Philosophy of Science* **42**: 75–84.

Pascal, B. (1660). *Pensées*. W. F. Trotter, Ed. & Trans. Michigan: Christian Classics Ethereal Library.

Paxton, J. M. and Greene, J. D. (2010). 'Moral Reasoning: Hints and Allegations.' *Topics in Cognitive Science* **2**: 511–527.

Pauer-Studer, H. (2009). 'Humean Sources of Normativity.' In *Hume on Motivation and Virtue*. C. Pigden, Ed. Palgrave Macmillan, 186-207.

Pereboom, D. (2001). *Living Without Free Will.* NY: Cambridge University Press.

———— (2013). 'Free Will Skepticism, Blame, and Obligation.' In *Blame: Its Nature and Norms*. D. J. Coates and N. Tognazzini, Eds. NY: Oxford University Press, 189-206.

Pettit, P. (2003). *Rules, reasons, and norms: Selected essays*. Oxford: Oxford University Press.

Pettit, P. and Smith, M. (1996). 'Freedom in belief and desire.' *Journal of Philosophy* **93** (9): 429–449.

Pigden, C. (1991). *Non-Naturalism versus Nihilism*. Coursebook, University of Otago.

———— (2007). 'Nihilism, Nietzche and the Doppelganger Problem.' *Ethical Theory and Moral Practice* **10**: 441-456.

Pinker, S. (2011). *The Better Angels of our Nature*. NY: Penguin Books.

Plunkett, D. (2011). 'Expressivism, representation, and the nature of conceptual analysis.' *Philosophical Studies* **156**: 15–31.

Priest, G. (1985-6) 'Contradiction, Belief and Rationality.' *Proceedings of the Aristotelian Society* **86**: 99-116.

———— (1997). 'Sexual Perversion.' *Australasian Journal of Philosophy* **75** (3): 360-372.

Putnam, H. (1962). 'It Ain't Necessarily So.' *Journal of Philosophy* **59** (22): 658-671.

———— (1967) 'The thesis that mathematics is logic.' In *Bertrand Russell, Philosopher of the Century*. R. Schoenman, Ed. London: Allen and Unwin.

———— (1975a). 'The Meaning of Meaning.' In his *Mind, Language, and Reality*. Cambridge: Cambridge University Press.

———— (1975b). *Mathematics, Matter and Method.* Cambridge: Cambridge University Press.

———— (1979). 'Philosophy of Logic.' Reprinted in *Mathematics Matter and Method: Philosophical Papers, Volume 1* (2nd edition). Cambridge: Cambridge University Press, 323–357.

Quine, W. V. O. (1951). 'Two Dogmas of Empiricism.' *Philosophical Review* **60**: 20-43.

———— (1953). 'Reference and Modality.' In his *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

———— (1976). 'Carnap and Logical Truth.' Reprinted in *The Ways of Paradox and Other Essays* (revised edition). Cambridge, MA: Harvard University Press, 107–132.

Rachels, J. (2012). 'The Challenge of Cultural Relativism.' In *Ethics: Essential Readings in Moral Theory*. G. Sher, Ed. London: Routledge, 151-58.

Railton, P. (1986). 'Moral Realism.' *Philosophical Review* **95**: 163-207.

———— (1989). 'Naturalism and Prescriptivity.' *Social Philosophy and Policy* **7** (1): 151-174.

———— (2003). 'On the Hypothetical and Non-Hypothetical in Reasoning About Belief and Action.' In his *Facts, Values, and Norms: Essays Towards a Morality of Consequence*. Cambridge: Cambridge University Press, 293-321.

Rall, J. and Harris, P. L. (2000). 'In Cinderella's Slippers? Story Comprehension from the Protagonist's Point of View.' *Developmental Psychology* **36**: 202-208.

Ramsey, F. P. (1931). *The Foundations of Mathematics, and Other Logical Essays*. London: Routledge & Kegan Paul.

Rawls, J. (1974-1975). 'The Independence of Moral Theory.' *Proceedings and Addresses of the American Philosophical Association* **48**: 5-22.

———— (1999). *A Theory of Justice* (revised edition). Cambridge, MA: Belknap Press.

Raz, J. (2005). 'The Myth of Instrumental Rationality.' *Journal of Ethics & Social Philosophy* **1** (1): 1-28.

Ridge, M. (2006). 'Ecumenical Expressivism: Finessing Frege.' *Ethics* **116** (2): 302-336.

Rinck, M and Bower, G. H. (1995). 'Anaphora Resolution and the Focus of Attention in Situation Models.' *Journal of Memory and Language* **34**: 110-131.

Rosen, G. (1990). 'Modal Fictionalism.' *Mind* **99** (395): 327–354.

———— (2003). 'Culpability and Ignorance.' *Proceedings of the Aristotelian Society* **103**: 61-84.

Rosen, G. and Burgess, J. P. (2004). 'Nominalism reconsidered.' In *The Oxford Handbook of Philosophy of Mathematics and Logic*. S. Shapiro, Ed. Oxford: Oxford University Press, 515-535.

Rowland, R. (2013). 'Moral Error Theory and the Argument from Epistemic Reasons.' *Journal of Ethics and Social Philosophy* **7** (1): 1-24.

Ruse, M. (1986). *Taking Darwin Seriously*. London: Blackwell.

Ryan, M. (1980). 'Fiction, non-factuals, and the principle of minimal departure.' *Poetics* **9** (4): 403-422.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.

Sarkissian, H., Park, J., Tien, D., Wright, J. C. and Knobe, J. (2011). 'Folk Moral Relativism.' *Mind & Language* **26** (4): 482–505.

Sayre-McCord, G. (1986). 'The Many Moral Realisms.' *The Southern Journal of Philosophy* **24**: 1-22.

———— (1989). 'Deception and Reasons to be Moral.' *American Philosophical Quarterly* **26** (2): 113-122.

Searle, J. (1975). 'The Logical Status of Fictional Discourse.' *New Literary History* **6** (2): 319-332.

Scanlon, T. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

———— (2014). *Being Realistic about Reasons*. NY: Oxford University Press.

Schroeder, M. (2007a). 'Reasons and Agent-neutrality.' *Philosophical Studies* **135** (2): 279-306.

———— (2007b). *Slaves of the Passions*. NY: Oxford University Press.

———— (2010). *Noncognitivism in Ethics*. London: Routledge.

Schroeter, L. 'Two-Dimensional Semantics.' In *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), E. N. Zalta, Ed. URL = <https://plato.stanford.edu/archives/win2012/entries/two-dimensional-semantics/>.

Schueler, G .F. (1988). 'Modus Ponens and Moral Realism.' *Ethics* **98** (3): 492-500.

Schwitzgebel, E. (2001). 'In-Between Believing.' *The Philosophical Quarterly* **51** (202): 76-82.

———— (2002). 'A phenomenal, dispositional account of belief.' *Noûs* **36**: 249–275.

———— (2012). 'Mad Belief?' *Neuroethics* **5** :13–17.

———— (2013). 'A dispositional approach to attitudes: Thinking outside the belief box.' In *New essays on belief.* N. Nottelmann, Ed. NY: Palgrave Macmillan, 75–99.

Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford and NY: Oxford University Press.

———— (2005). 'Error Theory and the Possibility of Normative Ethics.' *Philosophical Issues* **15**: 107-120.

———— (2010). 'The Possibility of Metaethics', *Boston University Law Review* **90**: 479-96.

Shah, N. (2003). 'How truth governs belief.' *Philosophical Review* **112** (4): 447–482.

———— (2006). 'A New Argument for Evidentialism.' *The Philosophical Quarterly* **56** (225): 481-98.

Shepski, L. (2008). 'The Vanishing Argument from Queerness.' *Australasian Journal of Philosophy* **86** (3): 371–387.

Sher, G. (2006). *In Praise of Blame*. Oxford: Oxford University Press.

Singer, P. (2011). *The Expanding Circle: Ethics, Evolution, and Moral Progress*. Princeton, NJ: Princeton University Press

Sinnott-Armstrong, W. (2006). *Moral Skepticisms*. NY: Oxford University Press.

———— (2008). 'A Contrastivist Manifesto.*' Social Epistemology* **22** (3): 257–270.

———— (2010). 'Mackie's Internalisms.' In *A World Without Values: Essays on John Mackie's Moral Error Theory*. R. Joyce and S. Kirchin, Eds. Dordrecht: Springer, 55-70.

Sinnott-Armstrong, W. and Wheatley, T. (2012). 'The Disunity of Morality and Why it Matters to Philosophy.' *Monist* **95**: 355-377.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

———— (2004). *Ethics and the a priori: Selected essays on moral psychology and meta-ethics*. Cambridge: Cambridge University Press.

———— (2007). 'Is there a nexus between reasons and rationality?' *Poznan Studies in the Philosophy of the Sciences and the Humanities* **94** (1): 279-298.

———— (2012). 'Naturalism, Absolutism, Relativism.' In *Ethical Naturalism: Current Debates*. S. Nuccetelli and G. Seay, Eds. Cambridge: Cambridge University Press, 226-244.

Sobel, D. (1999). 'Do the Desires of Rational Agents Converge?' *Analysis* **59**: 137–47.

Southwood, N. (2008). 'Vindicating the Normativity of Rationality.' *Ethics* **119**: 9-30.

———— (2011). 'The Moral/Conventional Distinction.' '*Mind* **120** (479): 761-802.

———— (2016). 'The Thing To Do" Implies "Can".' *Noûs* **50** (1): 61–72.

Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.

Steglich-Petersen, A. (2006). 'No norm needed: On the aim of belief.' *Philosophical Quarterly* **56** (225): 499–516.

Stephens, C. (2001). 'When is it selectively advantageous to have true beliefs? Sandwiching the better safe than sorry argument.' *Philosophical Studies* **105** (2): 161–189.

Sterelny, K. (2010). 'Moral Nativism: A Sceptical Response.' *Mind & Language*, **25** (3): 279–297.

———— (2012). 'Morality's Dark Past.' *Analyse & Kritik* **1**: 95-115.

Sterelny, K. and Fraser, B. (2013). 'Evolution and Moral Realism.' *British Journal for the Philosophy of Science* **doi**: https://doi.org/10.1093/bjps/axv060.

Stevenson, C. L. (1937). 'The emotive meaning of ethical terms.' *Mind* **46**: 14-31.

Stevenson, H. N. C. (1954). 'Status Evaluation in the Hindu Caste System.' *The Journal of the Royal Anthropological Institute of Great Britain and Ireland* **84** (1/2): 45-65.

Stotz, K. and Griffiths, P. (2004). 'Genes: Philosophical Analyses Put to the Test.' *History and Philosophy of the Life Sciences* **26** (1): 5-28.

Stotz, K., Griffiths, P. E. and Knight, R. (2004). 'How Biologists Conceptualize Genes: An Empirical Study.' *Studies in History and Philosophy of Science* **35** (4): 647-673.

Stouffer, S. A. (1955). *Communism, Conformity, and Civil Liberties*. NY: Doubleday.

Street, S. (2006). 'A Darwinian Dilemma for Realist Theories of Value.' *Philosophical Studies* **127**: 109–166.

Streumer, B. (2011). 'Are normative properties descriptive properties?' *Philosophical Studies* **154** (3): 325–348.

———— (2013). 'Can We Believe the Error Theory?' *Journal of Philosophy* **110**: 194-212.

Sturgeon, N. (1982). 'Brandt's Moral Empiricism.' *The Philosophical Review* **91** (3): 389-422.

———— (1994). 'Moral Disagreement and Moral Relativism.' *Social Philosophy and Policy* **11** (1): 80-115.

———— (2006). 'Ethical Naturalism.' In *The Oxford Handbook of Ethical Theory*. D. Copp, Ed. Oxford: Oxford University Press, 91-121.

Suikkanen, J. (2013). 'Moral Error Theory and the Belief Problem.' In *Oxford Studies in Metaethics, Volume 8*. R. Shafer-Landau, Ed. Oxford: Oxford University Press, 168-194.

Svavarsdóttir, S. (2001). 'Objective Values: Does Metaethics Rest on a Mistake?' In *Objectivity in Law and Morals*. B. Leiter, Ed. Cambridge: Cambridge University Press, 144-193**.**

Svoboda, T. (2017). 'Why Moral Error Theorists Should Become Revisionary Moral Expressivists.' *Journal of Moral Philosophy* **14** (1): 48-72.

Tangney, J. P., Stuewig, J., Malouf, E. T. and Youman, K. (2013). 'Communicative Functions of Shame and Guilt.' In *Cooperation and its Evolution*. Sterelny, K., Joyce, R., Calcott, B. and Fraser, B. Eds. Cambridge: MIT Press, 485-502.

Taylor, S. E. and Brown, J. D. (1988). 'Illusion and Wellbeing: A Social Psychological Perspective on Mental Health'. *Psychological Bulletin* **103**: 193-210.

Ugazio, G., Majdanžić, J. and Lamm, K. (2015). 'Are empathy and morality linked? Evidence from moral psychology, social and decision neuroscience, and philosophy.' In *Empathy and Morality*. H. L. Maibom, Ed. NY: Oxford University Press, 155-171.

Unger, P.K. (1979). 'Why There are No People.' *Midwest Studies in Philosophy* **4** (1): 177-222.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Oxford University Press.

van Inwagen, P. (1990). *Material Beings*. Ithaca: Cornell University Press.

van Roojen, M. (2005). 'Expressivism, Supervenience and Logic.' *Ratio* **18**: 190-205.

Velleman, D. (1988). 'Brandt's Definition of "Good".' *The Philosophical Review* **97** (3): 353-71.

———— (2000). 'On the aim of belief.' In his *The possibility of practical reason*. NY: Oxford University Press, 244–281.

Vitz, R. (2009). 'Doxastic Voluntarism.' In *The Internet Encyclopedia of Philosophy*. J. Fieser and B. Dowden, Eds. URL = < http://www.iep.utm.edu/doxa-vol/>.

Vogelstein, E. (2013). 'Moral normativity.' *Philosophical Studies* **165**: 1083–1095.

Wallace, R. J. (2003). 'Review of Richard Joyce, The Myth of Morality.' *Notre Dame Philosophical Reviews*. URL = <http://ndpr.nd.edu/news/the-myth-of-morality/>.

Walton, K. (1990). *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.

———— (1994). 'Morals in Fiction and Fictional Morality.' *Proceedings of the Aristotelian Society* **68**: 27-50.

Weatherson, B. (2004). 'Morality, Fiction, and Possibility.' *Philosophers' Imprint* **4** (3): 1-27.

Wedgwood, R. (2002). 'The aim of belief.' *Philosophical Perspectives* **16**: 267–97.

———— (2003). 'Choosing Rationally and Choosing Correctly.' In *Weakness of Will and Practical Irrationality*. S. Stroud and C. Tappolet, Eds. Oxford: Oxford University Press, 201-29.

Weiskopf, D. (2010). 'The theoretical indispensability of concepts.' *Behavioral and Brain Sciences* **33**: 228–229.

West, C. (2008). 'Personal Identity: Practical or Metaphysical?' In *Practical Identity and Narrative Agency*. K. Atkins and C. Mackenzie, Eds. NY: Routledge, 56–77.

———— (2010). 'Business as Usual? The Error Theory, Internalism, and the Function of Morality.' In A World Without Values: Essays on John Mackie's Error Theory. R. Joyce and S. Kirchin, Eds. Dordrecht: Springer, 183-198.

White, R. 'You just believe that because...' *Philosophical Perspectives* **24** (1): 573–615.

Whiting, D. (2014). 'Reasons for Belief, Reasons for Action, the Aim of Belief, and the Aim of Action.' In *Epistemic norms: New essays on action, belief, and assertion*. C. Littlejohn and J. Turri, Eds. Oxford: Oxford University Press, 219–38.

Wicker, F. W., Payne, G. C. and Morgan, R. D. (1983). 'Measurement of social-evaluative anxiety.' *Journal of Consulting and Clinical Psychology* **33**: 448-457.

Wilkins, J. S. (2013). 'Essentialism in Biology.' In *The Philosophy of Biology: A Companion for Educators*. K. Kampourakis, Ed. Dordrecht: Springer, 395-419.

Williams, B. (1973). 'Deciding to believe.' In his *Problems of the Self*. Cambridge, MA: Cambridge University Press, 136–51.

———— (1981). 'Internal and External Reasons.' In his *Moral Luck*. Cambridge: Cambridge University Press, 101–13.

———— (1995). *Making Sense of Humanity*. Cambridge: Cambridge University Press.

Williamson, T. (2001). 'Ethics, Supervenience, and Ramsey Sentences.' *Philosophy and Phenomenological Research*, **62** (3): 625-30.

Wright, C. (1996). 'Truth in ethics.' In *Truth in Ethics*. B. Hooker, Ed. Oxford: Blackwell, 1-18.

———— (2004). 'Warrant for Nothing (and Foundations for Free)?' *Aristotelian Society Supplementary Volume* **78** (1): 167–212.

Wright, J., McWhite, C. and Grandjean, P. (2014). 'The cognitive mechanisms of intolerance: Do our meta-ethical commitments matter?' In *Oxford Studies in Experimental Philosophy, Volume 1*. T. Lombrozo, S. Nichols and J. Knobe, Eds. Oxford: Oxford University Press, 28-61.

Yablo, S. (2001). 'Go Figure: A Path Through Fictionalism.' *Midwest Studies in Philosophy* **25**: 72–102.

———— (2002). 'Red, Bitter, Best.' *Philosophical Books* **41** (1): 13–23.

Zimmerman, M., (1988). *An Essay on Moral Responsibility*, Totowa, NJ: Rowman & Littlefield.