Metaethics and Psychology: an Investigation into the Links between Autism, OCD, and Depression and Metaethics

Riana Popat

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

Department of Philosophy

August 2024

Abstract

My thesis examines how psychological conditions impact metaethical theories. More precisely, I look at three separate mental health conditions and how they impact three different metaethical theories. My thesis argues that certain mental health perspectives should inform certain metaethical theories. These mental health conditions are Autism, Obsessive compulsive disorder (OCD), and Major Depressive Disorder, and I will look at how these conditions link to three separate metaethical theories, which are Moral Hermeneutic Fictionalism, Weak Motivational Internalism, and the Humean Theory of Motivation. As I will argue in this thesis that mental health conditions should inform metaethical theories, I will use the mental health conditions specified as case studies to support my overall argument. My findings also yield the conclusion that mental health conditions have different impacts on metaethical perspectives, but nonetheless, these conditions do have an impact on metaethics.

Acknowledgements

First, I would like to greatly thank my primary supervisor, Chris Woodard for all his guidance, mentorship, support, academic insight and his valuable time in both helping me to write this PhD thesis and also helping tremendously to develop me into a young confident professional. I also really value and appreciate all the time and effort he has put into guiding me on my first professional career. I would also like to greatly thank my second supervisor, Neil Sinclair for all his support, expertise, feedback, guidance and academic conversations on the PhD. I am also extremely grateful to Stefano Predelli for all his encouragement, support and guidance throughout my PhD. My gratitude also extends to Benedict Rumbold for his insightful advice and feedback on my work during the annual review processes. My extension of gratitude also extends to Joe Cunningham and Craig French for all their insightful conversations on metaethics and psychology. I would also like to greatly acknowledge Adina Covaci for all her helpful and encouraging conversations throughout the last two years of my PhD. I would also like to acknowledge Kasim Khorasanee and James Chamberlain for their helpful feedback on my work. I would also like to thank Jon Robson for his guidance on disability matters and his useful insights into Fictionalism (and Philosophy in general). I would also like to thank Andy Fisher for introducing me to metaethics and for his helpful conversations and comments on my work. Finally, I would like to extend my thanks to my examiners, Simon Kirchin and Benedict Rumbold for their time in reading and examining my thesis.

I made many friends and colleagues during my PhD, and I hope we will keep in touch as we all progress into our careers. I would particularly like to thank my friends and colleagues Alice Monypenny, Kash Sunghuttee, and Sean Shields for their helpful comments and conversations on my PhD. I would also like to thank my good friends for all their support throughout my PhD, including Kez Nightingale, Ellie Sherwood, Megan and Nathaneal Davison, and Tanya Kundu.I would also like to thank the whole of the Nottingham Philosophy Department for all their academic insights into my work. Moreover, I would also like to thank all my support workers (and the support worker service) including Deepak Assudani, Rachael Walker,

Mafeng Bitrus, Diret Tang'an, Isabella Wangler, and Gordon Stoner, and all my personal assistants (including Antonia Hristova, Elizabeth Gould, Elyn Jia Xuan Li, Dulcie Fisher Aakriti Bhardwaj, Gayathri Saisri Saranya) for all their help and support throughout my PhD.

I would like to thank the immense generosity of everybody who funded me. These organisations have supported me and my project greatly. They are the Snowdon Trust, the Disabled Leaders Network, the Humanitarian Trust, and the Funds for Women Graduates. I would also like to thank my places of work: The Brilliant Club, and Academic Oxford Summer Courses for the great teaching experiences that they have provided me with during my PhD.

I would like to thank my family who have supported and encouraged me and have always been there for me. I would first like to thank my Mum and Dad, Sheetal Popat, and Nilesh Popat for supporting me, guiding me, and always believing that I could get to the end of a PhD! Next, I want to thank my siblings, Ranai Popat, Niara Popat, and Aarin Popat who supported, guided, and kept me company during the PhD!

Finally, I would like to acknowledge and dedicate this PhD to my grandparents. All three of my grandparents, Prabhudas J Popat, Rama P Popat and Gunvanti M Morjaria, who have instilled high value in education in me and have supported my studies greatly. I dedicate therefore this PhD thesis to them!

Note to Examiners

Please note that I work with scribes as part of my support plan. I have tried my very best to format this PhD to the highest level, but formatting a document can be harder for me as I dictate to a scribe (as I cannot physically write and type myself).

Table of Contents

Introduction

Metaethics is often thought to be the subsection of ethics which discusses what we mean by concepts associated with morality. It is different to normative ethics (as normative ethics examines whether we *should* do an action) because metaethics is thought to be further removed from our everyday lives. For example, a metaethical paper might look at what individuals mean by the terms 'wrong' or 'right' when individuals say that 'murder is wrong' or 'giving to charity is right'. However, the way that people think and talk about ethics should inform metaethics and ultimately should result in a clearer picture of moral discourse. In other words, the psychology of an individual should influence how we think about concepts of right and wrong. Morality should be informed and led by how individuals think. Moreover, metaethics as a discipline should be inclusive and representative of the entirety of society's psyche. As mental health becomes more of a global issue in our society, metaethics needs to incorporate and consider mental health when it investigates how individuals understand and perform morality. This is the basic aim of my PhD chapters; I apply psychological conditions to metaethical discourse.

Moreover, in many areas of philosophy (including metaethics), philosophers have applied psychology to inform philosophical debates. For example, the Autism Objection has been applied to Hermeneutic Fictionalism (see Brock, 2014, pp. 588–89; Liggins, 2010, pp. 768–70; Stanley, 2001, p. 48; Tallant, 2013, pp. 819–20) and Weak Motivational Internalism has been criticised in connection with various forms of practical irrationality (Mele, 1996, pp. 733–34; Stocker, 1979, p. 744; Svavarsdottir, 1999, pp. 163–4). Furthermore, the Humean Theory has been criticised by many philosophers using different patterns of motivation incorporating different mental states. (See Gendler, 2008, p. 642; Park, 2013; Zangwill, 2008, p. 5). In recent years, metaethical debates concerning these theories have found themselves in a deadlock and others have sought to make some progress on psychology debates by introducing psychological evidence. Furthermore, I plan to build on the work of other philosophers to help make progress on the debate further by using psychology to influence metaethics. However, the psychological literature used in metaethics is often outdated or lacks detail, and thus metaethics as a

discipline requires updating with current psychological literature. Moreover, mental health conditions have often been mischaracterised in metaethics (perhaps due to a wider societal problem and a lack of evidence) and thus this motivates my research further to reform the metaethical discipline with psychological knowledge.

There are others that argue that metaethics needs to be updated. For example, Cholbi (2011, 2007) discusses Motivational Internalism and how specific psychological details of clinical depression and motivation are linked. Moreover, Stanley (2001, p. 48) argues for the Autism Objection (and uses psychological literature) but draws some conclusions which cannot be applied to Moral Hermeneutic Fictionalism (more explanation will be given in Chapter 2). Similarly, although the Humean Theory of Motivation has not yet considered how the motivation for an Obsessive compulsive disorder compulsion can be accounted for in their explanation, they have considered and refuted other mental states that seem to motivate without the Humean belief-desire pair pattern (for example, see Gendler, 2008; Park, 2013). I continue the conversation by applying psychological evidence to metaethical discourse, but I will also add a novel perspective on the debates, as I will bring current literature of psychological conditions to metaethical discourse.

In this thesis, I defend the claim that psychological conditions should impact metaethical theories. I must put a caveat here which is that, in this thesis, I do not discuss all psychological conditions or metaethical theories, but I discuss case studies of psychological conditions that provide evidence for my hypothesis.

In my thesis, I divide my chapters into three parts to reflect the different debates and conditions that I will be discussing. Broadly, my sections will be structured as follows; the first chapter will be focussed on the metaethical theory, and the second chapter will be focussed on how the psychological condition is related to the metaethical theory. The first section will be based on Autism Spectrum Disorder (ASD) and Moral Hermeneutic Fictionalism. Following the broader theme of the thesis, the first chapter of this section will be based on Moral Hermeneutic Fictionalism where I will argue that Moral Hermeneutic Fictionalism is a theory worth taking seriously. The second chapter will apply a specific objection to Moral Hermeneutic Fictionalism, where it will be argued that the Autism Objection may

not be applicable to Moral Hermeneutic Fictionalism. Although the Autism Objection looks as though it may be powerful against Moral Hermeneutic Fictionalism, I claim that it is not as powerful as it first appears, in light of psychological evidence. This is because there is psychological evidence to suggest that some autistic people may not be able to understand morality, which is a key premise of the Autism Objection.

The second section will focus on Weak Motivational Internalism and Depression. The first chapter of the section (Chapter 3) will be focussed on Weak Motivational Internalism, and I will discuss and argue for the plausibility of the theory. The second chapter of the section will offer a new counterexample, my Depression Counterexample, which is that depressed people with certain conditions may pose a problem for the Weak Motivational Internalism theory. This is because some depressed people, I argue, may meet the standards of rationality but they may still make a genuine moral judgement but fail to be motivated by that judgement.

The third section will be based on the Humean Theory of Motivation and how it links to Obsessive compulsive disorder. The first chapter in this section (Chapter 5) will argue that the Humean Theory of Motivation is a credible theory in metaethical discourse. The second chapter will argue that there is a counterexample to the Humean Theory which is that individuals with Obsessive compulsive disorder may be motivated to perform a compulsion without an appropriately related belief being present (but with a relevant desire) and this may serve as a counterexample to the Humean Theory.

I have chosen these particular areas of metaethics because these are the three areas, where looking at psychological conditions yield promising results, and I also think that in these three areas of metaethics, progress can be made by looking at psychological conditions. This is why, I focus on the particular metaethical theories and psychological conditions that I do, in this thesis. This is not, however, to claim that other psychological conditions and other metaethical theories cannot be linked together in this way, but my thesis could pave the way for further research (for other metaethical theories and psychological conditions to be explored), if my thesis' conclusions are indeed useful for the discipline of metaethics.

There are two larger questions that the reader should reflect on when they are considering my thesis. The first is the question of what the understanding of the psychological condition or the evidence of the psychological disorder contributes to our metaethical discourse. I make a case that these psychological conditions enhance metaethical understanding. The second question that the reader should reflect on is the nature of the evidence that I give and whether one should apply more psychological evidence to wider metaethics as well as to the specific metaethical theories that I discuss.

Part 1

Chapter 1: An Overview of Moral Hermeneutic Fictionalism

1. Introduction

In this chapter, I argue that Moral Hermeneutic Fictionalism is a theory worth taking seriously in metaethical discourse. I do this by first laying out the relevant landscape for Hermeneutic Fictionalism and Moral Hermeneutic Fictionalism. I do this in order to provide context for my next chapter, in which I defend Moral Hermeneutic Fictionalism against one particular objection, the Autism Objection, which is most prominently discussed by Stanley (2001, pp. 47–50). I will do this in the following way: first I will provide some context for the debate surrounding Hermeneutic Fictionalism, then I will discuss what Fictionalism and Hermeneutic Fictionalism amount to. In the next sections, I will give some reasons why Moral Hermeneutic Fictionalism is attractive and will give some responses to the worries presented regarding Moral Hermeneutic Fictionalism. Finally, I conclude that Moral Hermeneutic Fictionalism is a theory worth taking seriously in metaethical discourse.

2. Background to Fictionalism

Fictionalism is a family of theories about a discourse, according to which the discourse can be described as a pretence. In metaethics, it mainly refers to moral discourse; although, the concept of Fictionalism is widespread over all philosophical disciplines, for example, there is Fictionalism about morality, ontology and maths. Various authors have said that Fictionalism can be used for different discourses, for example Stanley (2001) investigates Hermeneutic Fictionalism in ontology, and Kalderon says that Hermeneutic Fictionalism should be used for morality (Kalderon, 2007). More precisely, Moral Hermeneutic Fictionalism is describing moral discourse as it stands now as a fiction (Kalderon, 2007, p. 136).

In this section, I place Fictionalism in the overall debate of metaethics. First, I consider Fictionalism's relationship to Error Theory. Error Theorists are cognitivist

but not realist (Daly and Liggins, 2010; Joyce, 2001; Joyce and Kirchin, 2010; Miller, 2013; Olson, 2014; Streumer, 2017). They claim that "moral judgements describe the world as having objective value, but that the world does not contain any objective value. Consequently all moral judgements are systematically and uniformly false"(Fisher, 2011, p. 178). This means that Error Theorists think that no moral facts exist. Thus, the Error Theorists hold that all assertions that there are moral facts, are false or mistaken. In other words, when we make a moral claim, such as 'you should not kill', one is actually uttering a false claim or is mistaken by their claim. Moreover, Error Theory is a rival theory to Moral Hermeneutic Fictionalism (but is compatible with Revolutionary Fictionalism – Revolutionary Fictionalism will be explained later in this chapter, in Section 2). Error Theory and Moral Hermeneutic Fictionalism agree that moral judgements are semantically representational and that moral properties are semantically represented by sentences that are literally false (van Roojen, 2015, p. 177). However, Error Theorists and Hermeneutic Fictionalists disagree about the way that we use our moral language; Error Theorists think that moral sentences express false beliefs, whereas Hermeneutic Fictionalists think that when we express a moral proposition it is in fact a make-believe claim.

There is also another theory called Moral Expressivism, which is an important rival to Moral Hermeneutic Fictionalism. Moral Expressivism is "a form of noncognitivism that analyses moral terms as linguistic devices for the expression of some mental state" (van Roojen, 2015, p. 294) (see also Blackburn, 1984; Ridge, 2007; Sinclair, 2021, 2009). Moral Expressivism is a rival to Moral Hermeneutic Fictionalism as Moral Expressivism explains moral sentences in terms of expressions, whereas Moral Hermeneutic Fictionalism describes morality as already being a fiction. Moreover, one form of Moral Expressivism is Emotivism which I will now discuss as a rival to Moral Hermeneutic Fictionalism. Emotivism is another form of Non-Cognitivism which understands moral linguistic terms in terms of the expression of emotion (van Roojen, 2015, p. 294) (See also Ayer, 2001; Hare, 1952a; Miller, 2013; Stevenson, 1937). Emotivism is a rival to Moral Hermeneutic Fictionalism as Emotivism says that all moral utterances are expressions of emotion whereas Moral Hermeneutic Fictionalists claim that all moral terms are typically used to make fictional claims. The last theory that I will discuss in relation to Moral

Hermeneutic Fictionalism is Moral Realism. Moral Realism is the view that "moral properties and/or facts exist and are, in some way, independent from people's judgements."(Fisher, 2011, p. 180). (See also Finlay, 2007; FitzPatrick, 2009; Miller, 2021, 2013; Sayre-McCord, 1986; Shafer-Landau, 2003). Moral Realism is a rival theory to Moral Hermeneutic Fictionalism, as Moral Realism claims that moral facts or properties exist and are independent from individuals' judgements (or beliefs). Moral Realism is different to Moral Hermeneutic Fictionalism which says that one is uttering fictional propositions.

I propose that one could be interested in Moral Fictionalism because it seems to offer a way of being an anti-realist without getting into the apparent difficulties that other anti-realist theories face (for example Moral Expressivism). This is because Moral Hermeneutic Fictionalism, seems to avoid some of the problems (such as the Frege-Geach problem) associated with other anti-realist theories (such as Moral Expressivism).

3. What is Fictionalism?

As stated above, Fictionalism can be applied to any discourse such as maths, identity, or negative existentials.[1] The main premise of any discourse being able to be described by Fictionalism is that that discourse D can be or will be redescribed in terms of a pretence. For example, if one takes Fictionalism to describe what the discourse of maths is doing, then it is the case that one is describing maths in terms of a pretence and therefore describing maths in a Fictionalist way. To put it more precisely, Fictionalism can be "characterised as the view that claims made within that discourse are not best seen as aiming at literal truth but are better regarded as a sort of 'fiction' " (Eklund, 2024). In other words, fictional claims should be understood as not asserting something that is literally true, for example, if I utter that

[1] There are many different types of Fictionalism such as Mathematical Fictionalism (see Yablo, 2002, 2001, 2000a, 2000b), Fictionalism about Scientific Theories (van Fraassen, 1980), Moral Fictionalism (which incorporates Moral Hermeneutic Fictionalism, but also includes Moral Revolutionary Fictionalism), for example Joyce (2005a, 2001)), Modal Fictionalism (Hale, 1995; Hinckfuss, 1993; Kim, 2005; Rosen, 1990), Fictionalism about Truth (Armour-Garb and Woodbridge, 2015; Burgess and Burgess, 2011), Fictionalism about Negative Existentials, Identity Statements and Fictional Characters (Kim, 2005; Rosen, 1990; Walton, 2000, 1990, 1985), and Fictionalism about Ordinary Object Discourse (Dorr and Rosen, 2002; van Inwagen, 1990).

one plus one equals two, then in a mathematical fiction, one plus one does equal two, however this may not be true outside of this fiction. Similarly, the general principle of Fictionalism can be applied to many discourses (that the discourse in question operates as a fiction). Moreover, to provide some context, there is a question about what kind of pretence.

The first distinction that should be made about Fictionalism concerns the Linguistic and Ontological Theses. The Linguistic Thesis says utterances of the sentence are not best seen as efforts to say what is literally true but as some useful fiction (Eklund, 2024). For example, the utterance that 'murder is wrong' is not an assertion but is a useful fiction. In contrast, the Ontological Thesis of Fictionalism says that entities of a discourse do not exist or they have the ontological status of fictional entities (Eklund, 2024). In other words, the Ontological Thesis concerns the entity that the given moral utterances appear to be about. For example, if I utter the phrase 'murder is wrong', I appear to be saying that murder is a wrong action (in instances of murder), but the Ontological Thesis says that you appear to be referring to (a wrong action) which does not exist (or are fictional entities), and therefore the utterance of 'murder is wrong' is a fiction.

Eklund (2024) notes that one is able to endorse the Linguistic Thesis or the Ontological Thesis separately or together. Moreover, Eklund claims that one reason why the Linguistic Thesis and the Ontological Thesis can run together is because the Linguistic Thesis is motivated by ontological concerns (Eklund, 2024). This is because when the Linguistic and Ontological Theses are advocated together (see Nolan et al., 2005, p. 308 for further details), we can say that claims are literally false, but it is still useful to utter such claims for theoretical purposes. In other words, Moral Hermeneutic Fictionalists can say that, although claims are literally false, they can still be useful in moral talk. For example, consider a claim interpreted by the Linguistic and Ontological Theses of 'killing is wrong'; together, these theses would claim that the moral sentence was literally false, but the claim still has some practical purpose in moral talk as this claim is preventing people from killing and killing is not beneficial for our society. Moreover, the claim avoids ontological concerns, since a combined thesis is useful without implying that wrongness, or indeed wrong actions, exist.

A way of clarifying the Linguistic Thesis is the distinction between Force and Content Fictionalism. Content Fictionalism says that when making an utterance, some content is asserted, but what is asserted is different to the literal content of the assertion (Eklund, 2024). For example, if I were to say, 'Harry Potter waves a wand to do magic', I am actually asserting something in accordance with the book of Harry Potter. Therefore, I am asserting something like 'according to the book from J.K. Rowling, Harry Potter waves a wand to do magic'. According to the Content Fictionalist, this proposition is asserting something about the book of Harry Potter, but not directly asserting that Harry Potter waved a wand. It should be noted most forms of Fictionalism are Content Fictionalism apart from Instrumentalism.[2]

In contrast to this, the Force Fictionalist claims that when uttering a moral statement, the content expressed by the utterance of the sentence is not asserted, but instead some other speech act is performed (what that speech act is, is decided by the kind of Force Fictionalism one is endorsing) (Eklund, 2024).[3] For example, the Force Fictionalist would say that the content of the sentence of 'giving to charity is right' is not asserted but some other speech act is performed. The content of the above sentence, according to the Force Fictionalist, however could still be 'giving to charity is right'. In other words, I am doing something else other than asserting the sentence when I say, 'giving to charity is right,' for example, I might be conveying the content or performing another speech act. Force Fictionalism is compatible with saying the content of the speech act is still conveyed (as conveying is a speech act) i.e. the utterance that 'giving to charity is right' can still convey that giving to charity is right, one is just not asserting the content.

There does seem to be a difference between conveying and asserting a proposition. For example, if I am conveying a story, I am simply telling one a story, I am making no assertion whether the story is correct or incorrect (in other words, I am taking no stance on whether the story should be taken seriously or not). For example, if I am telling you a story about Little Red Riding Hood, I might say Little Red Riding

[2] Instrumentalism is the theory that the speaker is not really asserting anything, but only pretending to do so ((Eklund, 2024).
[3] Other types of speech acts include requesting, warning, inviting, apologising and predicting. See Green (2021) for further details.

Hood wears a red hood, and there I am merely conveying the claim that Little Red wears a hood rather than asserting that Little Red wears a hood, but in order for one to understand the characteristics of the character in the story, one needs to be told that Little Red wears a hood.

On the other hand, critics might say that for one to be conveying a message, one is standing by it. For example, if one is conveying a message that 'Little Red Riding Hood wears a hood', then they are actually asserting that 'Little Red Riding Hood wears a hood'. However, I do not think that is plausible, as one can simply tell another person a story without suggesting that the meaning of the story is true; one could utter words, and be, therefore, conveying words without asserting the claims that one has conveyed. For example, consider an agent who utters 'break a leg'. When the individual utters 'break a leg', it is possible that the agent is not asserting anything, but still conveying (or reporting) something else. For example, perhaps the individual speaking the metaphor is actually conveying that they hope that the agent will perform well on stage. In other words, they are not literally asserting that the actor should break a leg but instead conveying that they wish the actor luck on stage.

Moreover, one can endorse a Fictionalist theory that combines Content Fictionalism and Force Fictionalism together. This means that in a moral sentence, the literal content is conveyed but not asserted and other content other than the literal content of the sentence is asserted (Eklund, 2024). For example, if I say, 'Sherlock Holmes wore a black coat', I am asserting some content here, but it is not the literal content of what is asserted. In other words, I am not literally saying that 'Sherlock Holmes wore a black coat', but I am saying that 'in the fictional book of Arthur Conan Doyle, Sherlock Holmes wore a black coat.'

There is also a psychological issue of what mental state one is in when one accepts a fictional sentence. There are two main schools of thought of what mental states are occurring when one utters a fictional sentence, and I discuss them now. The first interpretation is Cognitive Fictionalism which is the thought that Fictionalism (which is a judgement about the discourse) involves a belief about a sentence describing the content of a fiction. One should note that Cognitive Fictionalism is similar to Content Fictionalism as both views hold that when we utter the sentence

'X is F' we do not assert a belief that the content is 'X is F', but we do assert a belief of some other content (Maslen, 2006).[4] In other words, Cognitive Fictionalists believe that all moral judgments are beliefs (and the belief is not a fiction), but the belief is a belief about the content of a fiction i.e., the belief has fictional content. For example, Cognitive Fictionalism can be demonstrated by saying 'in the fiction murder is wrong'. However, in contrast, Non-Cognitive Fictionalism claims that there is a non-belief like attitude when a fictional utterance is expressed, and this non-belief like attitude is the type of attitude that we have when we engage in fiction. This is demonstrated by the fact that when we are uttering 'Sherlock Holmes wears a black coat', according to the Non-Cognitive Fictionalist, we are engaging in a non-belief like attitude when we are engaging in the fiction of Sherlock Holmes.

To further illustrate the contrast between Cognitive Fictionalism and Non-Cognitive Fictionalism, I now discuss the example of the utterance 'Sherlock Holmes lives on Baker Street'. Cognitive Fictionalism would claim I am *talking* about a fiction when I say 'Sherlock Holmes lives on Baker Street' whereas Non-Cognitivist Fictionalism would say that when I am talking about 'Sherlock Holmes living on Baker Street', I am make-believing that 'Sherlock Holmes lives on Baker Street' (there may be other possible states here that an agent is engaging with, but make-believing is one that Fictionalism endorses, so I refer to make-believing here). In sum, the difference between Cognitive Fictionalism and Non-Cognitive Fictionalism is that according to Cognitive Fictionalism an agent believes that content about a fiction is true, for example, if I utter 'murder is wrong', then I will believe that 'according to moral fiction, murder is wrong'. On the other hand, Non-Cognitivist Fictionalism claims that when one is making a moral utterance, one is expressing a Fictionalist attitude, for example, when I utter that 'murder is wrong', I am expressing a non-belief like Fictionalist state.

There are two different questions that one needs to grasp to fully understand Fictionalism. The first is how we could understand Fictionalism linguistically and the second is how we understand Fictionalism psychologically. Therefore, Fictionalism needs to be describing what is happening when one is uttering a moral

[4] There is a very close connection between an assertion and a belief. See, for example, Truth and Objectivity (Wright, 1993) for further details.

sentence both linguistically and psychologically. I claim that Force Fictionalism (which is a linguistic theory) fits well with Non-Cognitivist Fictionalism (which is a psychological theory), and Content Fictionalism (which is a linguistic theory) fits well with Cognitivist Fictionalism (which is a psychological theory). This is because the combined Force Fictionalist and the Non-Cognitivist claim that moral judgements are not assertions, and when one accepts a moral judgement, it is a non-belief like state. In contrast, Content Cognitivist Fictionalism states that when one utters claims one is asserting something other than the content of the fiction, and when one accepts a moral judgement, one has a belief.

The type of Fictionalism that I will be looking at in this chapter will be Moral Hermeneutic Fictionalism. Moral Hermeneutic Fictionalism is the claim that moral discourse, as it currently exists, is fictional. More precisely, Moral Hermeneutic Fictionalism can be formulated in the following way: "[…] we never really believed P, but that we have all along, in some sense, made-believe that P." (Hussain, 2004, p. 152). Moral Hermeneutic Fictionalism is in contrast with Revolutionary Fictionalism, which insists that "when engaging in D we ought only to make such pretend-assertions" (Eklund, 2024). This means that Moral Revolutionary Fictionalism claims that when we engage in a given [moral] discourse, we ought to make statements in a pretence. This is in contrast to Moral Hermeneutic Fictionalism where they say that we are already in a fiction when we utter moral sentences, and we are already making fictional claims.

Moreover, more precisely, Moral Hermeneutic Fictionalism is "a description of some actual [moral] discourse" (Kalderon, 2007, p. 136).[5] Consider the following example: Imagine two children pretending that a cardboard box is a spaceship, when they proclaim the sentence that, 'we are going to the moon,' they are not literally saying that they are going to the moon, but they are going to the moon figuratively.

[5] Moral Hermeneutic Fictionalism might adopt different views of what pretence amounts to. For example: Metarepresentational, Behaviouralist and Intentional. A Metarepresentational view is the view that the concept of pretending "underlies the ability to engage in pretend play" (Liao and Gendler, 2011, p. 81). Behaviouralist view is "hold the ability to engage in pretend play is driven by the process of behaving-as-if" (Liao and Gendler, 2011, p. 81). The Intentionalist view holds that engaging in a pretence requires behaving *as-if* as well as [the agent's] intention to do so; in other words, the agent acts in terms of *as-if* rather than behaves *as-if* (Liao and Gendler, 2011, p. 81). See Liao and Gendler (2011) for further discussion on different views of pretence.

In other words, it is true that they are going to the moon in the game, but it is false that they are literally going to the moon in the real world. Moreover, if we utter the phrase that 'Sherlock Holmes lives on Baker Street', it would be the case that Sherlock Holmes lives on Baker Street in a fiction, but it would not be the case that Sherlock Holmes lived on Baker Street in the real world. These two examples are analogous to how Moral Hermeneutic Fictionalism would see morality – they would say that the utterance of 'murder is wrong' is only true in the fiction which, in this case, would be in our moral system. In other words, the utterance of 'murder is wrong' would not literally be true as we are just pretending that 'murder is wrong'. In contrast to Moral Hermeneutic Fictionalism, I now briefly consider Moral Revolutionary Fictionalism. Moral Revolutionary Fictionalism aims to describe the way that we should use moral discourse in the future, given that something is wrong about our moral discourse now. For example, if we say that 'murder is wrong', one should be speaking in a pretence, and this is the correct way to understand moral practice. This is in contrast with Moral Hermeneutic Fictionalism as Moral Hermeneutic Fictionalism claims that one is describing moral discourse as it actually stands now, whereas Moral Revolutionary Fictionalism argues that in order to see progress, we should be describing morality as a fiction in the future. Thus, it is significant to briefly explore the relationship. (For further discussion on Moral Revolutionary Fictionalism (see Fisher, 2011, p. 157; van Roojen, 2015, pp. 194–195).[6]

The distinction between different discourses when applied to Hermeneutic Fictionalism is relevant as in my argument, which is presented in the next chapter, I will be presenting an objection which is primarily given by Stanley (2001, pp. 47–50); this is called the Autism Objection. In Stanley's paper, the Autism Objection is raised against Hermeneutic Fictionalism about ontology, not against Hermeneutic Fictionalism about morality, which means that to date the Autism Objection has not been investigated when put forward against Moral Hermeneutic Fictionalism. This is important as different types of Fictionalism are varied from one another. Therefore, I

[6] It should be noted that Moral Revolutionary Fictionalism links to other metaethical theories in varied ways to Moral Hermeneutic Fictionalism. For example, as discussed above Moral Hermeneutic Fictionalism and Error Theory are rivals, whereas Moral Revolutionary Fictionalism and Error Theory are compatible.

argue in the next chapter that, building on Stanley's work, the Autism Objection may not be defensible for Moral Hermeneutic Fictionalism as some autistic people (who meet certain conditions which will be explained in Chapter 2) may not understand morality.

However, in order to set the scene for my next chapter, I now discuss why one should find Moral Hermeneutic Fictionalism plausible before moving to discussing some objections to the theory, and I offer some replies to these objections on behalf of Moral Hermeneutic Fictionalism. I do this because I claim that understanding its plausibility and defending Moral Hermeneutic Fictionalism against criticisms is vital to understanding why one should take the theory seriously in metaethical discourse.

4. Attractive Features of Moral Hermeneutic Fictionalism

I now lay out some reasons why one should find Moral Hermeneutic Fictionalism attractive, to make the overall claim that it is a theory that one should take seriously in metaethical discourse. One reason to find Moral Hermeneutic Fictionalism plausible is that the view seems to take off some of the semantic burden, which is usually placed on Non-Cognitivists, because Moral Hermeneutic Fictionalism creates a positive way of looking at sentences. Non-Cognitivists (particularly Psychological Non-Cognitivism - the view that the state of mind of accepting a moral claim is not a cognitive representational state; and Semantic Non-Representationalism – that moral sentences do not describe or represent the world in the same way as other sentences do (van Roojen, 2015, p. 178)), seem to make negative claims about the state of moral thought. Additionally, for Moral Hermeneutic Fictionalism it is an advantage that these claims are logically separable (other Non-Cognitivist theories claim that Psychological Non-Cognitivism and Semantic Non-Representationalism cannot be separated) (van Roojen, 2015, p. 179).

This is because Moral Hermeneutic Fictionalism can benefit from the advantages of Non-Cognitivism, without having to change the view too much, and without entailing Semantic Non-Representationalism (van Roojen, 2015, p. 179). A Moral Hermeneutic Fictionalist (who is also a Force Fictionalist) claims that a moral predicate has a semantic meaning (i.e. moral properties are in the sentences that they

represent), but in claiming that for example ‘X is wrong’, we do not express the belief that X possesses the properties of wrongness. In other words, moral claims can have Semantic Representationalism (the view that the meaning of a moral predicate is the property that it represents), with having Psychological Non-Cognitivism (the view that when one utters a moral sentence one expresses a belief – or makes the assertion – about moral states of affairs). This means that moral sentences understood by Moral Hermeneutic Fictionalism have semantic meaning (their meaning is what they represent) but uttering these sentences does not have full assertoric meaning and this is advantageous to Moral Hermeneutic Fictionalism. A good analogy to this is the concept of metaphors (van Roojen, 2015, p. 179). Consider the metaphor of ‘Riana is over the moon’. It is not literally the case that Riana is over the moon, but it is representing an idea, that Riana is ecstatic, which is different to Riana being literally over the moon (as this would be impossible). In other words, believing in Psychological Non-Cognitivism allows somebody who does not believe in moral properties but does say that moral sentences represent some event or object in the world, and Moral Hermeneutic Fictionalists allow for this (van Roojen, 2015, p. 179).

Moreover, one more compelling reason to find Moral Hermeneutic Fictionalism plausible is that it avoids the Frege-Geach problem. The Frege-Geach problem is the problem that theories such as Moral Expressivism face. The problem begins with how Moral Expressivists understand moral claims. Moral Expressivists analyse moral claims in terms of non-belief-like states; in other words, when I utter ‘murder is wrong’, the Expressivist is claiming that I am expressing a non-belief like attitude of, ‘Boo! Killing!’. The Frege-Geach problem however is a problem for Moral Expressivists because, when we set out an argument involving a moral claim that the Moral Expressivist interprets, in *modus ponens* form, it does not seem to make a valid argument.

Consider the following:

1. If murder is wrong, then hitting somebody is wrong.
2. Boo! Murder! (the Moral Expressivist’s understanding of ‘murder is wrong’)
3. Hitting is wrong.

This is following a standard *modus ponens* formula, but P3 does not follow on from the combination of P1 and P2 (and it should do in order for the argument to be valid), and therefore this is the essence of the Frege-Geach problem (in other words the conclusion does not follow on from the premises, making the argument invalid). This is because Moral Expressivists claim that moral sentences express non-belief like states; therefore, when one utters a sentence, one expresses a stance rather than a belief-like state, 'Boo! Murder' and therefore Moral Expressivists cannot explain why arguments containing moral claims which appear to be valid actually are valid.[7]

However, one reason why one should endorse Moral Hermeneutic Fictionalism is that Moral Hermeneutic Fictionalism meets the criteria of not having moral facts, while also not facing the Frege-Geach problem. This is because Moral Hermeneutic Fictionalism (which is using a Force Fictionalist thesis) accepts Semantic Representationalism which means that indicative moral sentences describe or represent the world. As already stated, Moral Hermeneutic Fictionalism does not require moral facts, because when I utter 'murder is wrong' I am uttering this only in the pretence, and also expressing propositions rather than just expressing a non-belief- like state such as 'Boo, Killing'. This is because other Non-Cognitivists such as Moral Expressivists reject Semantic Representationalism which means that other Non-Cognitivists say that moral sentences do not convey propositions, but rather moral sentences convey expressions. Therefore, Moral Hermeneutic Fictionalism can allow for linguistic expressions in the *modus ponens* formula that look like assertions such as 'murder is wrong' whereas other Non-Cognitivists cannot. This leads to Moral Hermeneutic Fictionalism not facing the same problem that Moral Expressivism has with the Frege-Geach problem.

More precisely, Moral Hermeneutic Fictionalism avoids the Frege-Geach problem, because as already stated above (in the first reason) in Moral Hermeneutic Fictionalism, one can accept Semantic Representationalism with, at the same time,

[7] The Frege-Geach problem is a prominent problem for many metaethical theories (for example Ayer (2001), Stevenson(1937, 1944) and Hare (1952a)), and it is one which has been discussed in much philosophical literature, for example see Fisher (2011, p. 92) for a full description on the Frege-Geach problem. Moreover, for further discussion on the Frege-Geach problem, see Eklund, Joyce and Schroeder (2009; 2017, pp. 226–242; 2008). Many responses have been given to the Frege-Geach problem, for example Blackburn (1998*)*.

denying Psychological Representationalism. Therefore, with Moral Hermeneutic Fictionalism propositions can stand in logical relation. This means that Moral Hermeneutic Fictionalism can accept that the meaning of a moral predicate is in fact the property that it represents e.g. the meaning of 'right' is the property of rightness, and the meaning of moral sentences are states of affairs i.e. the meaning of the sentence that 'giving to charity is right' is the state of affairs of *giving to charity-being-right.* At the same time, one can deny Psychological Representationalism i.e. one can deny the belief that 'giving to charity is right'. This is because Moral Hermeneutic Fictionalism accepts Semantic Representationalism, and thus can maintain that moral propositions can be conveyed but not asserted in moral sentences. In contrast, Moral Expressivism denies that the meaning of a moral predicate is the meaning that it represents because it rejects Semantic Representationalism. Therefore, Moral Expressivism cannot use the same solution that the Moral Hermeneutic Fictionalist can use to solve the Frege-Geach problem, as the Moral Hermeneutic Fictionalist can maintain that sentences convey propositions whereas Moral Expressivism cannot and therefore Moral Expressivism still faces the Frege-Geach problem.

One further advantage is that Moral Hermeneutic Fictionalist statements coincide with Non-Fictionalist statements. For example, Joyce (2017, p. 75) discusses the case of a moral statement 'stealing is wrong' and a real-life claim of 'Amy stole last Tuesday'. The fictionalist believes that a moral claim can influence behaviour and what people think about what they are doing morally and this moral fictionalist claim does in fact have a connection to the real world. This is advantageous as the Moral Hermeneutic Fictionalist does not have to commit to moral facts but can still say that moral statements can influence individual's behaviour.

A final advantage of Moral Hermeneutic Fictionalism is given by Van Roojen. He says that an advantage of Moral Hermeneutic Fictionalism is that the Moral Hermeneutic Fictionalists need not commit to the facts, properties, and entities actual existence in moral discourse (van Roojen, 2015, p. 180). Furthermore, the fact that Moral Hermeneutic Fictionalism does not commit to moral facts, properties and entities is advantageous as it seems to support J. L. Mackie's Argument from

Queerness (1990, p. 38).[8] Broadly, Mackie's Argument from Queerness claims that moral properties are "utterly different from anything else in the universe" (1990, p. 38). This means that if we are aware of moral properties then we would need some special features that describe moral perceptions (1990, p. 38). This aligns with Moore's argument (1922, p. 113), where he speaks of non-natural properties. In order to understand these non-natural properties, we need moral intuition – how we understand moral properties is sometimes called the Argument from Epistemology (see Mackie, 1990, p. 38 for further details on the Argument from Epistemology). I claim that there are some similarities between the advantages of Moral Hermeneutic Fictionalism and Mackie's argument that some entities are odd or queer, and this seems to support the claim that we should not commit to moral facts, properties or entities in moral discourse. This is because if Mackie's argument stands and moral entities and facts are queer; perhaps we should not try to understand moral facts, entities or properties. Therefore, Moral Hermeneutic Fictionalism gives us a way to not commit to moral facts or moral entities on moral discourse, while still enabling society to benefit from moral claims.[9] Furthermore, although the moral sentences make fictional descriptive claims about the world, Moral Hermeneutic Fictionalists deny that the correct use of the sentences commits the speaker to the truth of those sentences (van Roojen, 2015, p. 180). In other words, the Moral Hermeneutic Fictionalist says that uttering the propositions does not commit the speaker to the belief or stance in the proposition expressed in the fictional sentence and thus does not commit us to moral facts (van Roojen, 2015, p. 180).

Now that we have discussed some attractive features of Moral Hermeneutic Fictionalism, I move on to discussing and defending the theory against some criticisms, before starting to lay my groundwork for my next chapter which will

[8] I do not include all claims made by the Argument from Queerness in this chapter (see Mackie, 1990, pp. 38–42). I only include Mackie's claim that moral claims are utterly different and thus this makes Moral Hermeneutic Fictionalism more plausible as it allows us to simply deny (as moral claims, according to Mackie, are utterly different) the existence of moral facts, properties and entities, while still allowing one to give moral propositions. I only include the claims about the Argument from Queerness insofar as it supports Moral Hermeneutic Fictionalism's plausibility.

[9] It is beyond the scope of my chapter to discuss the Queerness Argument in depth. See Olson (2014, pp. 79–135) for further details on the Queerness Argument. Moreover, for a critique of the Queerness Argument, see Shepski (2008).

argue that the Autism Objection does not work against Moral Hermeneutic Fictionalism.

5. Objections and Replies for Moral Hermeneutic Fictionalism

In the next chapter of my thesis, I defend Moral Hermeneutic Fictionalism against the Autism Objection. However, in this section, I first consider some other objections and replies to the theory, showing that we should take Moral Hermeneutic Fictionalism seriously.

The first objection I consider is called the phenomenological objection. The objection is given by many different authors against Moral Hermeneutic Fictionalism so it is important to discuss the objection (see for further details Brock, 2014, p. 578; Chrisman, 2007, p. 7; Eklund, 2024 see section 4.1; Fisher, 2011, pp. 159–60; Liggins, 2010, p. 768; Reynolds, 2009, p. 315; Stanley, 2001, p. 46; van Roojen, 2015, p. 189). The objection starts from the fact that Hermeneutic Fictionalists imply that people do not know that they are pretending when they utter moral statements, but according to the Phenomenological objection it is implausible to think that we do not know we are pretending when we are actually pretending. The above two claims taken together imply that such people (who are supposedly engaging in a pretence) do not believe themselves that they are engaging in a pretence. However, Moral Hermeneutic Fictionalism implies that whether or not individuals are engaging in a pretence is inaccessible to them, and the phenomenological objection claims that the theory of Moral Hermeneutic Fictionalism makes it hard to show that people are engaging in a pretence as we must take people's testimony seriously, when they say that they are not engaging in a pretence.

Moreover, for clarity, the phenomenological objection can be laid out in the following way:

P1 Moral Hermeneutic Fictionalism claims that people uttering moral sentences are engaged in pretence.
P2 We assume such people do not believe themselves to be engaged in pretence.

P3 They think they are not pretending, but Moral Hermeneutic Fictionalism implies they are.

C This implies they cannot tell whether or not they are pretending which is an objection for Moral Hermeneutic Fictionalism.

Van Roojen notes that a criticism of the phenomenological objection is that the phenomenological objection does not adequately use all Fictionalist accounts to combat the objection (van Roojen, 2015, p. 190). In other words, van Roojen (2015, p. 190) claims that the phenomenological objection only uses one type of fictionalism and thus he worries that this is being uncharitable to the Hermeneutic Fictionalist. He claims that some Hermeneutic Fictionalist accounts might work and might be able to provide a solution to the objection. For example, there is an account of Fictionalism that says that Fictionalists know when others are speaking in a fiction. My response on behalf of the Moral Hermeneutic Fictionalist would be that there would be no need for the listener and the speaker to realise that they are speaking in Fictionalist terms in an ideal account of Moral Hermeneutic Fictionalism. In other words, I agree that Moral Hermeneutic Fictionalism implies that whether someone is engaged in pretence may be inaccessible to them but deny that this is a problem for the theory. This is because Moral Hermeneutic Fictionalism does not claim there is any need for the speaker to know that they are not speaking in literal truths but rather they are speaking in figurative truths.[10]

Moral Hermeneutic Fictionalism and the phenomenological objection reach a deadlock. The Moral Hermeneutic Fictionalist is going to say that the speaker and the listener need not realise they are in a fiction, whereas critics of Moral Hermeneutic Fictionalism are going to claim that the speaker and the listener should know they are speaking in a pretence, if they are engaging in a pretence when they are engaging in the discourse in question. It seems like there is no way to resolve this issue and therefore we need another approach in which to solve this problem. I propose that such an approach would be looking at some individuals with Autism

[10] It should be noted that someone who adopts the metarepresentational view will disagree with this claim.

and investigating how they understand pretence and individual discourses (which (Stanley, 2001, pp. 47–50) has also suggested).

This is because some autistic people struggle to understand fiction. This means that if the discourse in X is a fiction they will struggle to understand it. More precisely, Stanley (2001, p. 48) notes that the same psychological mechanisms are used both to understand make-believe and the fictional discourse in Hermeneutic Fictionalism. Therefore, the phenomenological objection appears to be supported by empirical evidence. This is because some autistic people that have problems understanding make-believe should also have problems understanding the discourse that the fiction is referring to, but it seems like, they *actually* do understand the discourse that the Hermeneutic Fictionalist is referring to. However, I leave this objection open as I plan to discuss this objection in my next chapter, in regard to Moral Hermeneutic Fictionalism.

The second objection is that Moral Hermeneutic Fictionalism is not clear on what type of pretence they are explaining when they talk about the pretence in the theory (the closest that we have is Kalderon (2007, p. 136) as he explains what he means by pretence, but even he does not precisely spell out the specific details of what the pretence amounts to). However, I claim that this is not a problem with the theory, rather it just reflects the fact that we need to have a clear idea of what pretence amounts to and it is possible to get this, we just need to lay out the conditions of the pretence. Moreover, there are many plausible theories of pretence to choose from when considering Moral Hermeneutic Fictionalism, and this further suggests that Moral Hermeneutic Fictionalism is a theory worth taking seriously. For example, Kendall Walton's (1990, p. 138) account of pretence or Stanley's (2001, pp. 31–40) account of pretence, who also builds on Walton's account of pretence (1990, p. 138). These authors all give a sound account of what a pretence might amount to. For example, Stanley (2001, p. 38) thinks that in a make-believe game, certain real world fact makes the linked proposition fictionally true. Additionally, Walton claims that all discussions of the ontological status of fictional entities (which can be realist or irrealist) begin with an observation that appears to be about fictional entities that also happens to be (in many cases) assertions of the literal truth (Walton, 1990, p. 391). Walton also notes that individuals use pretence as a way of describing verbal

participation in make-believe (Walton, 1990, p. 391). We can use both of these plausible accounts to describe what a pretence amounts to and thus Moral Hermeneutic Fictionalism is still a theory worth taking seriously.

The third objection is that we may abandon Fictionalist discourse when it is too hard to follow (Joyce, 2017, p. 80). For example, if our Fictionalist discourse says that abortion is wrong, we might be tempted to abandon this discourse if circumstances prevail (e.g. the mother's health is at risk), and it is a dubious moral situation. But this does not seem to happen with moral discourse (Joyce, 2017, p. 80). This is because we seem to value and hold up a certain moral standard in our society and it seems to have serious consequences if one does not follow that moral standard. For example, many people might say that killing is wrong. However, when this is put to the test and one wants to be euthanised, this is taken very seriously and judged accordingly. Furthermore, it is the case that we have serious consequences for not following moral discourse, for example, in the UK, one can go to prison for committing a crime (and in other countries, for example, Singapore, one can receive a death penalty for committing certain crimes). The gravity of the consequences of committing a crime implies that we do take moral discourse very seriously and we do hold people to moral standards. Furthermore, if we do not adhere to moral standards then society invokes punishment and this is evidence that we do not abandon moral discourse and in fact, we do take moral discourse seriously. The objection is therefore if we are living in a Moral Hermeneutic Fictionalist discourse then we would not take moral discourse as seriously as we do now, as according to the Moral Hermeneutic Fictionalist, morality is a pretence.

However, the Moral Hermeneutic Fictionalist could respond to this by saying that, in a pretence, we could still keep our moral standards and there is no reason that we should abandon a moral discourse, if morality gets difficult to deal with, if we are living in a pretence. This is because it is possible to take pretence seriously, even if it is a pretence. Consider a movie: we seem to take the plot seriously and we seem to care about the character, even though we know that it is a pretence. Similarly, a gambling game with money or with property could be considered as a pretence, but it has serious consequences, particularly for the losers who might lose their properties or their life savings. Moreover, the Moral Hermeneutic Fictionalist could

further respond, which I think is a good response, by saying that there are some benefits in using a pretence to describe moral discourse, in terms of our morality we need to pretend in order for people to act well to one another in society. For example, by pretending that we should not kill we protect others in our society and therefore there are still some benefits. Moreover, this is because on occasion people do have very good motive for engaging in a pretence, for example by using a metaphor. We may have good reason for using a metaphor in some circumstances, for example, using a metaphor to explain and express concepts clearly and concisely and therefore one is engaged in a pretence. This means that a concept of a pretence can be helpful in our language and therefore a pretence could be beneficial in a moral situation. Just as a metaphor is useful or perhaps the most useful way of explaining a concept or idea, a moral pretence may be better for society to maintain a moral standard and to prevent individuals from harm.

There is a fourth objection to Moral Hermeneutic Fictionalism which I will call the 'Seriousness Objection' (Lillehammer, 2004, pp. 104–105).[11] The objection is that we want to think of moral discourse as a discourse that should be taken seriously – in other words, we think that it is important to be moral. Individuals agonise over moral decisions, people are severely punished if they act immorally, and people fall out with one another over moral disagreements. If moral discourse is a pretence, then the seriousness in which people engage with morality seems hard to explain. However, I claim that it is possible to take some pretences seriously, even if it is a pretence. It simply depends on the players of the game and how seriously they treat the rules. In other words, just because morality is a pretence does not mean one cannot take it seriously; pretences are not intrinsically games that we cannot take seriously. Therefore, this objection does not show that Moral Hermeneutic Fictionalism is not a plausible theory.

[11] It should be noted that my 'Seriousness Objection' is based on Lillehammer's objection where he discusses Moral Hermeneutic Fictionalism failing to provide a serious basis for morality, see Lillehammer, 2004, pp. 104–105.

6. Conclusion

I have argued that Moral Hermeneutic Fictionalism is a theory worth taking seriously in metaethical discourse. I have argued this in the following way: first I have outlined the landscape for Hermeneutic Fictionalism and Moral Hermeneutic Fictionalism (and have highlighted how Moral Hermeneutic Fictionalism relates to other metaethical theories). Next, I have given some reasons why we would find Moral Hermeneutic Fictionalism attractive as a metaethical theory. Then, I have considered some objections and some replies to Moral Hermeneutic Fictionalism. In conclusion, I argue that Moral Hermeneutic Fictionalism is a theory worth taking seriously in metaethical discourse. I do this, as in the next chapter I will focus on one objection, the Autism Objection put forward by Stanley (2001, pp. 47–50) and I will argue that this Autism Objection may not work when applied to Moral Hermeneutic Fictionalism.

Chapter 2: Does the Autism Objection work against Moral Hermeneutic Fictionalism?

1. Introduction

In this chapter, I argue that the Autism Objection presented by Stanley (2001) may not work against Moral Hermeneutic Fictionalism. I argue this because some autistic people may not understand morality, and this means that one of the key premises presented in the Autism Objection (that autistic people understand morality) fails when put against Moral Hermeneutic Fictionalism. This is because much psychological literature shows that some autistic individuals may not understand morality because of their lack of empathy which is shown by their lack of theory of mind. I argue this in the following way: I first recap on what Moral Hermeneutic Fictionalism amounts to (section 1); then I define the Autism Objection as it is presented by Stanley (2001) (section 2). Next, I propose my argument that the Autism Objection may fail against Moral Hermeneutic Fictionalism as some autistic people may fail to understand empathy (because of their lack of Theory of Mind) and thus may struggle to understand morality, and this provides evidence against one of the key premises of the Autism Objection, when applied to Moral Hermeneutic Fictionalism.

2. Recap: What is Moral Hermeneutic Fictionalism?

Moral Hermeneutic Fictionalism uses the idea of pretence to describe moral discourse as "it actually stands" (Kalderon, 2007, p. 136). Moral discourse will be broadly defined as the practice of ascertaining what is right and wrong in our society (which occurs in the real world). I will adopt Stanley's view of a pretence which is, "In a game of make-believe, [pretence is where] certain real world facts make certain propositions fictionally true," (Stanley, 2001, p. 38). In other words, the practice of make-believe is where the person is imagining an entity (like an object or person) to be the case in the real world (this phenomenon is not already occurring in the real world) and is acting as if it was the case in the real world. Stanley's

quotation above is alluding to the fact that, Moral Hermeneutic Fictionalism claims that morality is in fact reflective of a pretence. For example, when I claim that 'murder is wrong' I am making a moral claim; this is a claim that occurs in the pretence of morality. This is further demonstrated by if I think that it is wrong to murder in the pretence and therefore, I do not murder in the real world. This means that in a pretence, certain propositions in the pretence are linked to propositions that represent real-life events, but the fictional propositions are not real-life events themselves.

I briefly discuss an example which is analogous to what is happening when Hermeneutic Fictionalism is applied to the discourse of morality. Imagine children playing a make-believe game of Cops and Robbers; a game in which 'robbers' have to retrieve a certain object, and 'cops' have to stop them. If a child utters 'I am going to steal the diamond,' they are only stealing the 'diamond' in the game, and thus this is a proposition that would only be true in the pretence and not in the real world. The proposition of 'I am going to steal the diamond' is reflective of what is occurring in the real world as the child performs certain actions in order to steal the diamond in the game (in this case, this might be running to get the diamond, avoiding other children who may stop them getting the diamond, etc.).

Moreover, this example is analogous to Hermeneutic Fictionalism when applied to morality (Moral Hermeneutic Fictionalism), as when I utter the claim that, 'I must not kill,' I am claiming that I must not kill in the pretence, because morality is in fact a pretence. This means that when I say, 'I must not kill,' I am uttering this phrase in the pretence, and thus, this is reflective of events happening in the real world, but these events might not actually be occurring in the real world exactly like they are happening in the pretence. Furthermore, I may not actually be asserting in the real world that one 'must not kill,' but I am merely uttering this proposition as we are pretending that we should not kill, because I am uttering 'do not kill' in the pretence. The most important thing to remember about Moral Hermeneutic Fictionalism is that

when one is uttering propositions, one is not asserting propositions.[12] This is because they are uttering these propositions in the pretence.

Furthermore, Moral Hermeneutic Fictionalism is similar to when one acts in a play. For example, if an actor (who is playing Romeo), says 'if I am forbidden to be in love with Juliet, then I do not want to live', the actor is actually saying that 'if I am forbidden to be in love with Juliet, then I do not want to live' in the play of Romeo and Juliet, rather than actually claiming this statement in the real world. However, this proposition might have an impact on the real world because one can say, 'in the play the actor playing Romeo claims that if he is forbidden to be in love with Juliet, then he does not want to live' and this statement may impact real-life events i.e. in this case it might dictate the narrative of the play. This case demonstrates an example in which we need to scale up in order for one to understand how Moral Hermeneutic Fictionalism impacts the real world. Moreover, to understand the theory of Moral Hermeneutic Fictionalism, one can then say with taking this example, that when one utters any moral sentence, one is not asserting the moral proposition but merely uttering the proposition in the pretence. This is synonymous to the way that Romeo utters 'if I am forbidden to be in love with Juliet, then I do not want to live' in the play of Romeo and Juliet. The actor is not actually asserting that he wants to commit suicide if he is not allowed to be in a relationship with Juliet, the actor is merely uttering this for the purpose of the play. These concepts are vital to understanding my criticism to the Autism Objection which will be presented in the latter part of the chapter.

However, now I turn to defining the Autism Objection. I first discuss it as it is presented by Stanley (2001) and then how the Autism Objection would be presented if it was applied to Moral Hermeneutic Fictionalism. I do this to show why the Autism Objection to Moral Hermeneutic Fictionalism may not work.

[12] It should be noted that I am adopting a Force Fictionalist view here rather than a Content Fictionalist view as this statement would not work with Content Fictionalism.

3. Defining the Autism Objection to Hermeneutic Fictionalism

The Autism Objection which is used in this chapter is given by Jason Stanley (2001, p. 48) (for other discussions of the Autism Objection, see Brock, 2014, pp. 588–89; Eklund, 2024, p. Section 4.1; Kim, 2014, pp. 322–323, 2013, pp. 158–60; Liggins, 2010, pp. 768–70; Tallant, 2013, pp. 819–20)). It should be noted that the Autism Objection was framed by Stanley as a worry for Hermeneutic Fictionalism about ontology. However crucially, Stanley did not intend for the Autism Objection to be a worry for Moral Hermeneutic Fictionalism (or at least, he suggested that the individual needs to do more work when the Autism Objection is applied to Moral Hermeneutic Fictionalism (Stanley, 2001, p. 66)). To give an entire account of what the Autism Objection amounts to, I will first spell out the objection and then discuss what Autism is.

First, I highlight the Autism Objection as presented by Stanley (2001, p. 48). Stanley gives his objection, as already explained before, against Hermeneutic Fictionalism about ontology (and other discourses such as negative existentials and identity), therefore what follows below is a generalised Autism Objection generated from Stanley's Autism Objection. The objection by Stanley is presented in the following way:

P1 Autistic people in set P (set P is the sub-set of Autistic people that do not understand pretence) do not understand, or cannot engage, in pretence.

P2 If discourse X can be explained by Hermeneutic Fictionalism, then discourse X is a pretence, and autistic people (in set P) would not understand or engage in it.

P3 If X is a pretence, then people in set P cannot engage in X or understand X.

P4 Some people (in set P) can engage in or understand X.

C Discourse X cannot be explained by Hermeneutic Fictionalism.

A few clarifications to make: the Autism Objection can be applied to all types of Hermeneutic Fictionalism; this includes Fictionalism about Ontology, Fictionalism about Mathematics, Fictionalism about Identity (one can simply plug in the desirable

discourse in Stanley's argument). However, my precise focus in this paper is only with Moral Hermeneutic Fictionalism. Importantly, Stanley's Autism Objection relies on the key premise that some autistic people cannot understand pretence but can understand discourse about X (for example discourse X could be about negative existentials, identity or ontology). Coupled with this, I build on Stanley's (2001, pp. 53, 66) remarks where he claims that the Autism Objection would not be able to be applied to more controversial topics such as morality. This is because I argue that the Autism Objection may not work against Moral Hermeneutic Fictionalism as a sub-section of autistic people (those who both cannot understand morality and those who do not understand pretence) exist, and this means that the Autism Objection to Moral Hermeneutic Fictionalism may fail. This is because if some autistic people do not understand these two concepts (pretence and morality) then the Autism Objection cannot hang on P3. From now in this paper, I refer to the sub-section of autistic individuals who do not understand pretence (set P) and do not understand morality, when I say autistic individuals unless otherwise specified.

The argument below can be laid out in the following way (the argument shows the premises that would have to be true if the Autism Objection were to be a viable objection to Moral Hermeneutic Fictionalism):

P1 Autistic people in set P do not understand pretence
P2 If discourse X can be explained by Hermeneutic Fictionalism, then discourse X is a pretence, and autistic people in set P would not understand it.
P3 If morality is a pretence, then autistic people in set P cannot understand it.
P4 Some people in set P understand morality.
P5 Morality cannot be explained by Hermeneutic Fictionalism.
C Therefore Moral Hermeneutic Fictionalism is not true.

This argument relies on two key premises – the claim that some autistic people do not understand pretence (P1), and that some autistic people understand morality (P4). The fact that some autistic people do not understand pretence is uncontroversial in the literature, and there is much psychological literature indicating this (for example, Baron-Cohen et al., 1985; Happé, 1995, 1994; Leslie and Frith, 1988). I will discuss this later in the chapter. However, the claim that

some autistic people understand morality is controversial and I elaborate on why it is controversial later in this chapter as this impacts the application of the Autism Objection to Moral Hermeneutic Fictionalism.

However, first I turn to what Autism Spectrum Disorder (ASD) amounts to and then I explain the two premises in turn. The first attribute that one should understand about ASD is that it is a spectrum. In other words, individuals who have ASD will have symptoms to varying degrees and may have different combinations of symptoms. Symptoms of ASD may include individuals with ASD having deficits in socio-emotional reciprocity. For example, they may have trouble with social interaction and find it hard to engage in back-and-forth behaviour (APA, 2013, p. 50). However, more notably to the Autism Objection people with ASD may find it hard to share interests and emotions with other individuals (APA, 2013, p. 50). They may also struggle to engage, initiate, or respond to, social cues (APA, 2013, p. 50).

Less relevantly to our discussion (however I mention it here as it is important to give a full account of ASD), they may have deficits in non-verbal behaviour, which is used for social interaction. For example, they may find it hard to engage in eye contact, they may face deficits in understanding facial expressions or bodily cues of other individuals that would be understood by non-autistic people, and sometimes autistic people can have a total lack of understanding of non-verbal behaviour, both by themselves and from others (APA, 2013, p. 50). However, perhaps more importantly to our discussion, however, some autistic people may have difficulty in understanding, maintaining and developing relationships or difficulty in adjusting their behaviour to fit social contexts (APA, 2013, p. 50). Crucially, however, to the Autism Objection, people with ASD may have difficulty in understanding imaginative or pretend play, as well as trouble making friends. They may also lack an interest in their peers (APA, 2013, p. 50).

The second criterion of ASD is the restricted and repetitive patterns of behaviour, interests or activities, as manifested by two of the following ways: autistic people may have stereotyped or repetitive motor movements, or they may repeat the use of objects, or repeat certain words or phrases (APA, 2013, p. 50). Moreover, they may have an insistence on sameness or routine, or patterns which are repetitive of verbal

or physical behaviour (APA, 2013, p. 50). They may have highly restricted or fixated interests that are abnormal in intensity. For example, a strong attachment to certain objects (APA, 2013, p. 50). They may also have a hyper-reactivity to sensory input or unusual interest in sensory aspects of the environment. They may, for example, have an excessive response to sights or smells, or have a visual fascination with lights (APA, 2013, p. 50).

The third criterion for ASD is that symptoms must be present in an early developmental period (APA, 2013, p. 50). However, sometimes they may not manifest until autistic people struggle to engage in social situations, or they may be masked by strategies learned in later life. They must also cause clinical significant impairment in social, occupational, and other areas of functioning (APA, 2013, p. 50). These disturbances must also not be better explained by intellectual disability, and can exist with other conditions (APA, 2013, p. 51).

From the general explanation of ASD, one can pick out certain features which are highly relevant to the Autism Objection. The most relevant features are the reduced sharing of interests, emotions, or the feelings of others, and difficulty to engage in social interaction. Moreover, the Autism Objection (and my argument) rely on some autistic people's inability to understand gestures or social cues of another person or oneself (which may make it hard for them to understand pretend play, as there is a need to understand others' body language and gestures when engaging in pretend play with others). Most importantly for my argument, however, they have difficulty with developing and maintaining relationships. For example, they may have difficulty adjusting behaviour to certain situations and an absence in sharing in imaginary (or pretend) play. Furthermore, they fail to understand fully the interest of others, or their peers. This is vital, because Moral Hermeneutic Fictionalism relies on pretence and pretend play as a concept heavily.

As already stated above, ASD is a spectrum, so individuals who suffer from ASD will suffer from varying symptoms within the disorder. However, in this chapter, I focus on two key attributes of ASD: the inability to understand pretence and the lack of understanding of morality or empathy. Some autistic people may struggle with understanding both pretence and morality. There is a caveat to this which is that

some autistic people may not struggle with understanding pretence or understanding morality, but some will. By understanding morality, I mean when an agent understands the reasons why an action is wrong both affectively and cognitively. The agent who understands morality will understand the belief of why the action is wrong (cognitively) and also understand and be able to feel the emotions associated with the action (affectively). For example, an agent who understands morality both cognitively and affectively would understand why the belief that hitting is wrong (cognitive understanding) and understand the emotions that the victim (of the hitting) feels when they are hurt, (affective understanding). More precisely I am focusing on the subsection of autistic people, who struggle to understand pretence and morality at the same time. Moreover, it should be noted that autistic individuals may struggle with the concepts of morality and pretence to varying degrees and combinations may be different among various autistic people.

There is an important question on how pretence is defined as Hermeneutic Fictionalism refers to how one understands and knows one is in a moral pretence. The first thing to note is that the literature claims that some autistic people may not understand a certain kind of pretence and the literature often assumes the metarepresentational view. Broadly a metarepresentational view is where the concept of pretence underpins pretend play (see for example Jarrold et al., 1994; Leslie, 1987). This is also the theory of pretence that is used in Moral Hermeneutic Fictionalism. More specifically, Leslie (1987, p. 423) claims that the deficit in pretend play in some autistic children is a result of a problem in metarepresentational pretence as well as a result of the lack of Theory of Mind.[13]

As pretend play is an important concept for my paper, it is important to explore what it amounts to and investigate why some autistic people might struggle to understand pretend play. First, it is vital to note that there are different types of pretend play. Moreover, there are some common types of play which clinicians often identify as leading to children engaging in pretend play (see Lin et al., 2017; Westby, 2022). There are four main types of play, which I highlight here. First, is exploratory play, which involves children using their senses to examine objects by, for example,

[13] Leslie (1987, p. 423) says that this deficit can be found in autistic individuals with both high and low support needs.

touching, smelling or feeling the object. Furthermore, in exploratory play, the child feels no connection between the world and the object (Westby, 2022, p. 7). There is also another type of play, relational play, this is where children begin to understand how objects relate to one another (Westby, 2022, p. 7). The third type of play is functional play, which involves using toys the way they were intended to be used, for example, putting a real phone to one's ears (but not calling anybody) (Westby, 2022, p. 7). Moreover, there is a fourth type of play which is called symbolic play. Symbolic play is when the child gives an object an alternative meaning, for example, giving a cardboard box in the earlier example the status of a spaceship (Westby, 2022, p. 7).

I now focus on symbolic play as I claim this is the kind of play the autistic people in set P struggle with. Three types of play can be incorporated into symbolic play, which is the type of pretend play that some autistic people struggle the most with and which I refer to in this chapter (Westby, 2022, p. 7). In other words, symbolic play is defined as the ability to imagine pretences and thus to imaginatively transform objects into something else (Westby, 2022, p. 7). This type of pretend play process can be object substitution where a child will substitute an object in place of something else. For example, you may have ribbons representing spaghetti (Westby, 2022, p. 7). Additionally, there is also another type of play which some autistic people may struggle with, which is imaginary play. This is where the agent imagines an absent object or assigns an absent attribute to an existing object. For example, in the illustration which is mentioned in Chapter 1 (where the children are flying to the moon), the cardboard box referred to represents a spaceship. This is because the cardboard box would be assigned the status and attributes of a spaceship. This is further demonstrated by the children assigning imaginary engines to the box to help it fly (Westby, 2022, p. 7). Finally, there is a kind of pretend play which is called agent play which is where the object becomes the agent of an action. For example, a doll might be assigned to go to work, so the child may pretend the doll is driving to work in a toy car (Westby, 2022, p. 7). Crucially, agent play may be hard for an autistic agent to understand as they may not be able to understand the doll's imaginary thoughts and feelings. In this chapter, I will be referring to agent play as I am mainly looking at how autistic people understand others' thoughts, feelings and actions in pretend play.

Now I have discussed what pretend play amounts to, I turn to why exactly some autistic people fail to understand pretence. In sum, they do not understand pretence because of lack of theory of mind (Baron-Cohen et al., 1985; Frith, 1989a; Premack and Woodruff, 1978). The theory of mind is the ability to understand and interpret others' thoughts, beliefs, feelings, and actions, and to be able to predict or manipulate others' beliefs or feelings, and react accordingly and appropriately to these beliefs or feelings (Lin et al., 2017, p. 1188; Zarei et al., 2022, p. 1). This is because, as said above, the deficit of theory of mind shows that one may struggle with empathising with other individuals' beliefs or feelings and this is what is required for pretence (Baron-Cohen et al., 1985; Frith, 1989a; Premack and Woodruff, 1978). I claim this is aided by the explanation of symbolic play (which I define above) which is the most relevant type of play when discussing the lack of theory of mind in some autistic people. Consider the example presented in the first chapter – imagine a case where children are playing on a pretend spaceship. One non-autistic child pretends to be an astronaut, the other child who has Autism (and struggles with understanding pretence), would struggle to empathise with the non-autistic child. More precisely, the autistic child may struggle with cognitive empathy, i.e. they may wonder why their counterpart is playing with a cardboard box, and they may struggle with understanding why they may do this; or affectively; they may struggle to feel the emotions that the other child is experiencing when they pretend that the card box is a spaceship, i.e. excitement (Baron-Cohen, 1997; Baron-Cohen et al., 1985; Happé, 1995, 1994). In the experiment that Baron-Cohen (1989, p. 286) conducted where the children were told a story concerning two protagonists Mary and John, the sample of autistic children were asked to predict what the protagonist was thinking in parts of the story. Baron-Cohen found that 80% of autistic children in the study lacked the ability to tell the psychologist what the protagonist was thinking, showing that autistic people may lack theory of mind (Baron-Cohen, 1989, p. 286).

To spell this out further, there are two types of levels of understanding here. There is cognitive empathy where the agent can empathise with the pretence happening and realises the narrative of the pretence. However, the second type of understanding, known as affective empathy, is where the agent can empathise with how the other agents feel. This means that if an agent struggles to understand other agents in an

emotional capacity, they lack affective empathy. Furthermore, this means that some autistic individuals might understand that others are engaging in a pretence, and they might understand from a theoretical point of view that they are in a pretence, but they may fail to understand how to engage in the pretence, or how other agents feel cognitively in a pretence. Just to be clear, I am interested in empathy concerning other people's perspectives as this is what the literature looks at (see Baron-Cohen, 1989; Baron-Cohen et al., 1985; Happé, 1995 for further details of the type of empathy psychologists use). The reason why some autistic people fail to understand pretence on a cognitive level (or an affective level) is because of their lack of theory of mind which I turn to next. The lack of theory of mind means that a person will struggle to understand another individual's point of view or another individual's thought process and moreover they will struggle to predict future events. This means that as people who lack theory of mind who struggle to understand another individual's thought process or point of view, they will struggle to understand pretence or pretend play, as pretence involves understanding situations from another perspective.[14] For example, imagine that two children are playing doctor and nurses, one pretends to be a doctor, one pretend to be a patient. In order to be able to understand that pretence, one needs to be able to understand the perspective of the doctor or the nurse in order to fully participate. They also need to understand the perspective of the character they are playing and the perspective of the character their partner is playing. For example, if they were playing the part of the doctor, they will have to understand the situation from the doctor's point of view, i.e. that I have to cure the patient and they also have to understand the situation from their counterpart's point of view, from the patient's point of view.

The lack of theory of mind which some autistic people exhibit shows how they lack understanding of other people's points of view (Baron-Cohen et al., 1985; Frith, 1989a; Happé, 1995, 1994). This is exemplified by the Sally-Anne test (Baron-Cohen et al., 1985, p. 41).[15] The Sally-Anne test was discussed by Baron-Cohen in

[14] It should be noted that I refer to a type of pretence which involves understanding other point of view as this is the type of pretence, which is referenced in the literature, in conjunction with some autistic people's lack of Theory of Mind. See, for example, Baron-Cohen (1989; 1985) and Happé (Happé, 1995) for further details on the type of pretence I am referring to.

[15] This type of pretence further indicates that the autistic people (in set P) struggles with non-actual pretence as they struggle to understand other people's point of view in the pretence.

his paper (1985, p. 41). The Sally-Anne test was conducted as follows; there were two protagonists, Sally and Anne (who are dolls) and the experiment was conducted with children with Autism. Before the experiment started the researcher checked that the children knew which doll was which (presumably to make sure that the children were commenting on the right doll). Sally placed a marble in her basket and then left the scene. The marble was then transferred to Anne (without Sally's knowledge) and hidden in Anne's box. When Sally returned, the researcher asked the children where should Sally look for the marble? If the children pointed to the previous location, then they passed the belief question (indicating that they understood Sally's point of view as Sally would have thought that the marble would have been put in the previous location). However, if the children pointed to the current location of the marble (Anne's box), then they will have failed the question because they did not take Sally's belief into account. This Sally-Anne test was conducted with autistic people as well as non-autistic people, and the autistic people in the study failed to answer the belief question accurately (although this question is called the belief question, it implies that the autistic people fail to understand the beliefs and feelings of others accurately because they struggle to understand others' perspectives), which suggests that some autistic people have a lack of theory of mind, which may explain why they may struggle to understand pretence (as said above, individuals that lack theory of mind struggle to understand events and situations from another person's perspective, that means they may struggle to get into the mindset of the people that they are embodying the character of in the pretence, and they may also struggle to understand what their counterpart(s) in the game is doing). More precisely, 80% of autistic children failed the Sally-Anne test and the Smarties test [(which is a similar type of test but asks the children to identify the location of the Smartie rather than the location of the marble, and both tested autistic people's ability to predict future events and thus testing the theory of mind (see Baron-Cohen et al., 1985, p. 41; also see Frith, 1989b, p. 40) for further details on results of how autistic people reacted to the tests)]. This is because as stated above, pretence requires an affective and cognitive understanding of other people's point of view and autistic people's lack of theory of mind shows that they may fail to understand situations cognitively and affectively.

Moreover, the lack of theory of mind may also help one to understand why some autistic people may not understand morality because they lack empathy (as they lack the understanding of others' perspectives), but I will return to this later in the chapter. In this section I have described Stanley's objection and why some autistic people may not understand pretence. I will now, in the next section, present how the Autism Objection would be applied to Moral Hermeneutic Fictionalism before I move to explain why applying the Autism Objection to Moral Hermeneutic Fictionalism may not be viable.

4. Applying the Autism Objection to Moral Hermeneutic Fictionalism

Now that I have discussed the Autism Objection, presented by Stanley, against Hermeneutic Fictionalism about Ontology, I now recap on the Autistic Objection in terms of moral discourse:

P1 Autistic people in set P do not understand pretence
P2 If discourse X can be explained by Hermeneutic Fictionalism, then discourse X is a pretence, and autistic people in set P would not understand it.
P3 If morality is a pretence, then autistic people in set P cannot understand it.
P4 Some people in set P understand morality.
P5 Morality cannot be explained by Hermeneutic Fictionalism.
C Therefore Moral Hermeneutic Fictionalism is not true.

However, there is much evidence to suggest that P4 may not be the case. This is because it has come to light that some autistic people may not understand morality. This depends on two individual claims about morality. The first is the claim that understanding morality requires empathy and the second is that some autistic people (in set P) may lack empathy and therefore may not understand morality. Moreover, one needs to understand the exact definition of empathy, therefore later in this chapter, I discuss this.

Firstly, I explain why some autistic individuals may not understand morality. They may not understand morality as I claim that empathy is a key part of morality and

some autistic people may struggle to understand empathy because of their lack of theory of mind.[16] In other words, empathy is a key component of my definition of morality i.e., one needs to have both types of empathy (cognitive and affective empathy) to understand morality (at least on my account). Moreover, my definition of morality is that one must both understand others' beliefs and feel how they are causing harm to another individual, for them to understand morality. We should accept this definition of morality because it seems intuitive that if one is causing somebody unjust harm that is intrinsically bad. Furthermore, a key indicator of whether and why we are causing somebody harm is empathy i.e. one needs to know the harm that one is causing another person to be empathetic towards them.

I claim that for an agent to understand morality and therefore be moral, they need to understand the reasons behind why an action is wrong. It seems intuitively right to say that if somebody just merely copies or does an action because it is wrong in society's eyes, this is not them understanding what morality amounts to, because they cannot understand and give reasons as to why something is wrong. It is important to distinguish and define different types of empathy in order to establish what precisely some autistic people struggle with when they fail to understand morality. This is where I turn to next.

As we have seen above, empathy involves the capability to feel the mental states of others (Read, 2019, p. 2) (see also Blair, 1995, p. 4; Hoffman, 1987, p. 48 for further details on empathy). I now discuss two different types of empathy in more depth to give the landscape of the concept of empathy but will focus on cognitive empathy in the majority of this chapter as this type of empathy is what some autistic people struggle with (and the evidence regarding the lack of theory of mind points to the fact that they lack cognitive empathy). Cognitive empathy is the capacity to understand another person's state of mind from their perspective (Spaulding, 2017,

[16] It should be noted that critics disagree that autistic people lack empathy. One school of thought is the 'Double Empathy Problem' which is the claim that some autistic individuals may understand empathy between each other but in a different way to non-autistic individuals. It is, therefore, better to consider empathy between pairs of people rather than between individuals. This is beyond the scope of my argument as I argue that there is enough evidence at this present time that some autistic people may lack empathy or the ability to empathize. This is why I only make the claim that some autistic people may lack empathy rather than a definitive claim that some autistic people do lack empathy. For further discussion on the 'Double Empathy Problem', see Milton (2016, 2012), Milton et al., (2022), Mitchell et al., (2021).

p. 13) (see also Blair, 2005; Kauppinen, 2017, p. 216; Leslie, 1987; Read, 2019, p. 3; Zaki, 2014). This is in contrast to affective empathy which is where one can feel what others feel ((Hoffman, 2001, p. 30) (see also Maibom, 2017, p. 22; Read, 2019, p. 3)). In other words, affective empathy involves sharing in another's emotional experience ((Read, 2019) (also see Slote, 2007; Sreenivasan, 2020)). The difference between affective empathy and cognitive empathy is the fact that cognitive empathy has more to do with understanding other people's reasons cognitively, whereas affective empathy is feeling another's emotions e.g. anger, sadness and embarrassment for oneself (Maibom, 2017, p. 22).

I have highlighted different types of empathy, which are important to distinguish when discussing why some autistic people may not understand morality. Consider the following example. If I hit somebody, and I regret the fact that I hit somebody, then it may be because I have empathy for that person. I may have cognitive empathy, because I understand that hitting another person will cause pain to them cognitively, but if I have affective empathy, I will feel the other person's pain. To be precise, some autistic people may lack cognitive empathy, but most have affective empathy. For example, in the case of hitting an agent, the autistic person will be able to affectively empathise with the person that they have hit, but they will struggle to cognitively empathise with the person that they've hit (see for further details on how autistic individual (in set P) struggle to understand cognitive empathy but have little deficit in affective empathy (for example see Dziobek et al., 2008; Fatima and Babu, 2022; McKenzie et al., 2022).

To my knowledge, there has not been a clear study done on the differences between understanding affective and cognitive empathy in autism in psychology. However, from the seminal psychological literature (Baron-Cohen, 1989; Baron-Cohen et al., 1985; Happé, 1995, 1994), we can infer that some autistic people may struggle to understand the thoughts and the feelings of other individuals and thus may not understand why something is wrong. More precisely, some autistic people struggle to understand cognitive empathy (understanding of cognitive recognition and processing) rather than distress or concern ((Mazza et al., 2014, p. 2) (see also Hirvelä and Helkama, 2011; Jones et al., 2010; Lockwood et al., 2013, p. 5; Samson

et al., 2012; Schwenck et al., 2012)).[17] Similarly, Georgiou et al (2019, pp. 1869–1870) found that autistic people struggled with cognitive empathy. Moreover, Georgiou et al (2019, p. 1871) suggested that although there are impairments in autistic people with affective empathy, autistic people are primarily impaired in cognitive empathy. The lack of theory of mind, as suggested previously, highlights that some autistic people may struggle to understand cognitive empathy as the theory of mind is asking people to discuss other individual's beliefs and some autistic people cannot identify other individuals' beliefs. An example of an autistic person failing to understand cognitive empathy is the Sally-Anne test discussed above, where the autistic individual is asked to say where Sally has put the marble, and the majority of autistic people struggle to answer this question. Moreover, most of the autistic people experience affective empathy (but the minority might struggle with understanding affective empathy). For example, in Baron-Cohen (1989) (see also Happé, 1995, 1994; Leslie and Frith, 1988) some autistic individuals are asked to identify signals of distress among others and these autistic individuals struggle to do this.

I claim that cognitive and affective empathy are needed for understanding morality. One needs affective empathy to understand the feelings of harm and one needs cognitive empathy to give reasons as to why an action is wrong. In other words, if I hit somebody, it is morally wrong, because it is wrong to harm other individuals (and because hitting physically and emotionally hurts others). Empathy provides the means to understand why it is wrong to harm the other person, and that it is wrong to harm the other person.[18]

However, in order to understand the relationship between empathy and morality, one needs to consider how, precisely, morality and empathy fit together so this is where I turn next. The first premise that one needs to consider is that empathy is needed in

[17] It should be noted that there is much debate on whether autistic people understand cognitive and/ or affective empathy. There are some papers that suggest autistic people do struggle with both. See for example Fatima and Babu (2022) for further details on a different perspectives on whether some autistic people understand affective empathy and cognitive empathy (Fatima and Babu (2022) claims that autistic people are impaired when it comes to both cognitive empathy and affective empathy.)
[18] It should be noted that there are other definitions of morality, like utilitarianism (where morality is based on consequences for further details see Mill (2014, pp. 8–10)) and deontology (where morality is based on intentions see Kant, 2011, pp. 57–63), but I argue that morality should be based on empathy.

order to understand the theoretical principles of morality (theoretically) but not to act morally (or perform the practical principles). In other words, I may understand morality theoretically by understanding the rules or the ideas of the particular concept without understanding how to do that concept, or vice versa. For example, I may know how to play cricket – I may know when I see the ball I must hit it with a bat and run (this refers to practical principles), but I may not know why the rules of cricket are as they are (theoretical principles - in this case I may not understand why I must hit the ball with my bat, I just witness all the other players doing this action and perform the act myself). In this cricket case, it is evident that I have knowledge-how rather than having knowledge-that.

Another example is when an agent donates money to charity. Imagine an agent donates £100 to charity, they may have the moral know-how to give to charity, but when asked why they gave to charity, they reply that they merely want to look good in their society and this is the sole reason as to why they give to charity. It looks like the agent does not have the know-that knowledge as they cannot give appropriate moral reasons (appropriate moral reasons in this case might be, for example, that the agent wants to help other individuals) as to why they do that action. This is the difference between doing an action which is moral and understanding the reasons behind doing the action. I argue that if the agent does the moral action, they have the knowledge-how to do the action (or the practical knowledge), however, if the agent has appropriate moral reasons as to why they do the action, they have the knowledge-that the action is the right thing to do (theoretical knowledge).

I claim that the agent needs both know-that and know-how knowledge, in order to be moral as the agent needs empathy to be moral, and to understand empathy one needs both know-that and know-how knowledge. For example, in the donation case, the agent lacks the know-that knowledge, on why they gave to charity. This means that they lack appropriate moral reasons as to why it is right to give to charity, even though they do the deed. They also fail to give appropriate reasons as to why they should give to charity. This could be for many reasons, but one reason might be that they lack the ability to empathise with another person's perspective. I claim that empathy is the reason why agents fail to give appropriate know-that moral reasons and therefore struggle to understand morality.

Moreover, I claim that neither theoretical (know-that knowledge) nor practical principles (know-how knowledge) are enough for agents to be moral by themselves, but they are both needed together for an agent to be moral. This is because empathy is needed to have theoretical knowledge for morality as it allows one to explain why one should be moral. In other words, empathy gives one appropriate normative reasons as to why we should be moral. This is important as some autistic people lack empathy and therefore, they may lack understanding or the ability to give reasons as to why an act is morally wrong, and therefore they lack theoretical knowledge associated with morality. For example, the autistic individual that does not understand the theoretical knowledge of why it is wrong to kill, may still not kill in society, as they understand that it is wrong on a practical level, so they do not do the action of killing people. This is not because they understand why it is wrong to kill people, but rather, they struggle to feel empathy towards another person being killed. Therefore, they struggle to understand why this action is wrong, in terms of other people's emotions. This is because some autistic individuals fail to understand why other individuals feel distress (and they may even struggle to give moral reasons in terms of harm). Instead, some autistic people choose not to kill because they may think that if they do kill, they may go to prison, which is not an empathetic reason not to kill (or not a reason which incorporates harm), and therefore they fail to understand morality (Leslie et al., 2006; Leslie and Frith, 1988; Nichols, 2002; Shulman et al., 2012).

An obvious objection to the claim that some autistic people do not understand morality would be that some autistic people act morally so they must understand morality and there is a question of why they are acting morally. I argue that when acting morally they are in fact making conventional judgements rather than moral judgements as they fail to engage in theoretical judgements because of their lack of empathy. In other words, they are just acting on what society finds acceptable rather than acting on moral reasons. For example, if an autistic person, who struggles to understand morality, was asked why they think it is wrong to kill, they might reply that they must not kill because society thinks it is unacceptable to kill and they will go to prison if they kill. This reply indicates a conventional reason for not killing rather than a moral reason as this reason lacks empathy. In the next section, I go into depth on the difference between moral and conventional norms and then explain

why I think some autistic people may be making a conventional judgement rather than a moral judgement, when they are seemingly making a moral claim.

5. What is the Difference between a Moral and Conventional Judgment?

Blair identifies that there are two types of transgressions: moral and conventional.[19] In other words, when somebody is making a moral judgment, they are, in fact, treating something as a conventional judgment. The difference between these two judgements is important to understanding what some autistic people may be doing when they seemingly make moral judgements (James and Blair, 1996). This is because I claim that some autistic people may make conventional judgements when they are seemingly making moral judgments.[20] I first consider a quote from Blair (1996, p. 572) as he gives a clear description of moral and conventional transgressions.

*"*Within the literature, moral transgressions (e.g., hitting another, damaging another's property) are defined by their consequences for the rights and welfare of others. Conventional transgressions (e.g. talking in class, dressing in opposite-sex clothes) are defined by their consequences for the social order" (James and Blair, 1996, p. 572).

In other words, moral transgressions are defined by their implications of harm towards others (or their impact on the welfare of others). For example, a moral

[19] Blair (1996, p. 572) uses the term 'transgressions' to describe the actions of what individuals do. However, I claim that judgments lead to transgressions and furthermore, moral judgments would typically lead to moral transgressions and conventional judgments would typically lead to conventional transgressions. For now, I will be using the term 'transgressions' in this section (before I identify transgressions as reasons later in the chapter). Therefore, from now, when I use the term 'transgressions' I mean transgressions and the judgment associated with it.

[20] A good example is from the novel *The Curious Incident of the Dog in the Night-Time*, where Christopher steals his father's credit card: "I had Father's cashpoint card in my pocket and it was illegal to steal things, but he was a policeman so I had to tell the truth, so I said, "I have a cashpoint card," and I took it out of my pocket and I showed it to him. And this was a white lie. But the policeman said, "Is this your card?" And then I thought he might arrest me…" This quote is showing that Christopher (Haddon, 2003, pp. 185–186) is concerned with being arrested rather than the morality of stealing his dad's credit card. In other words, he does not feel guilty for stealing his dad's credit card because it may hurt his dad but rather, he feels concerned about going to prison. This is an example of where an autistic individual seems to act morally, but for a conventional reason / norm.

transgression would result in physical or psychological harm to another person. This is illustrated by the claim 'do not kill' which seems like a moral transgression, because if an agent does kill, they would cause psychological and physical harm to another person. This is in contrast to conventional transgressions, which are defined by the society that we live in. For example, if one does not wear clothes in the street, it is an action that society has decided to implement, and if one did not wear clothes it would not cause harm to another person.

It should be noted that one can have a transgression which is both moral, and conventional. The two categories can co-exist simultaneously for the same transgression. For example, if I insult somebody this can be a moral transgression because it is harmful to another agent, but it can also be a conventional transgression as it is frowned upon in society; this can be different depending on different societies' moral and conventional transgressions. Another example of this may be a driving rule in the UK that all drivers must drive on the left side of the road. This rule has both a moral and a conventional norm embedded in it. This is because it is arbitrary whether we drive on the right or left side of the road, as long as all drivers have the same rule i.e., that they all drive on the same side of the road. However, at the same time, this driving rule can be considered as a moral norm, as if an agent decides to drive on the wrong side of the road, it might infringe on somebody's safety as one is increasing the risk of a collision and therefore this rule is also affecting welfare.[21]

As you can see, these categories can be overlapping, and one transgression can be both a moral and conventional transgression at the same time for different reasons; in other words, they are not mutually exclusive categories and can exist together.

[21] There is also an alternative category which is a legal law, which is a law decided on by the legal system in our society. This can look the same as a conventional law, as the population may govern laws, but may not necessarily be (particularly if we do not live in a democratic society). For example, consider a society ruled by the Taliban: it may not be the case that the population endorses the Taliban, but the Taliban might still enforce legal laws. In this case, conventional laws might look vastly different to legal laws. Similarly, it might be the case that some conventional laws are not legal laws, for example, adultery. This is a conventional law in the UK, but certainly not a legal law. This is shown by the fact that the agent will not be prosecuted for committing adultery, but adultery is frowned upon in society.

However, these categories can be distinguished, and some transgressions are only moral, and others are only conventional.

There might be a transgression that becomes a moral transgression depending on the society that the transgression occurs in. For example, take the law to not drink alcohol in Saudi Arabia. This, I would argue, is both a moral and conventional transgression in this particular society. It is a moral transgression in that society because it is believed to affect people's welfare both for religious reasons (Saudi Arabia is a Muslim state and therefore they believe that drinking alcohol is against God's will) and a conventional transgression as it is frowned upon in society because of the religious stigma attached and also for other societal reasons. One could also have a moral transgression that is not a conventional transgression like adultery. Imagine that a spouse has committed adultery against another spouse in the UK, this may be conventionally acceptable in certain cultures like UK, but this may be morally unacceptable to certain cultures because of religious reasons like Idaho, Massachusetts (as it is a Christian state), or because they think that marriage is sacred.

Moral transgressions are defined by their consequences for the rights and welfare of others, whereas conventional transgressions are defined as violations of social norms within society (Blair, 1995, p. 5). Moral transgressions are judged as being less rule contingent than conventional transgressions – this means that moral transgressions are not based on societal laws but based on what is considered moral in that society. In other words, moral transgressions have less authority justification (meaning that they are less dependent on authority figures deciding what is right or wrong) than conventional transgressions (Blair, 1995, p. 6).

Blair describes moral and conventional transgressions in terms of transgressions, but one could easily define these concepts in terms of reasons (in fact I claim that it would be more philosophically clear to define these concepts in terms of reasons). For example, one can describe the decision of not to hit somebody in terms of a moral reason. When an agent is asked to describe why they do not hit somebody they usually answer that, 'It is wrong to hit somebody because it causes the person harm and harm is wrong'. Moreover, the action performed could be the same if

somebody acted on conventional reasoning. This is demonstrated by the fact that one can say that they do not hit somebody because they do not want to be arrested. I note that this is the same action which is done for two different reasons, and it is difficult to know without discussing with the agent, which kind of reason is motivating their action.

Moral and conventional reasons can also be used to explain how psychopaths appear to act morally. Psychopaths appear to act morally without being moral and this is because they lack the understanding of moral reasons and treat moral reasons as conventional reasons. In other words, psychopaths appear to act morally as they may perform the moral action but when asked to explain why they did this action, they will fail to give genuine moral reasons. This results in psychopaths failing to be moral even though they appear to act morally. Experts believe that one of the distinguishing factors between a psychopath and a non-psychopath is that a psychopath will fail to make a distinction between a moral and conventional judgment and that he will not make victim-based justifications of why moral transgressions are not good to do (Blair, 1995, p. 20). In other words, the psychopath will treat a moral judgement as a conventional one, so they may still act morally, but they will give a conventional justification as why they will not do a certain action (Blair, 1995, p. 20). In other words, the justification they give will not be in terms of the victim's welfare, but will be in terms of non-victim based justifications (Blair, 1995, p. 20) (see also Nichols, 2002, p. 223).[22] Whether a reason is moral or conventional can be distinguished in terms of consequences that the agent gives i.e. the victim will give different kinds of consequences to doing the action (Cited from Durkheim in Turiel, 1983, p. 42) (see Durkheim, 1953 for more details).[23]

The question remains of what distinguishes moral reasons from conventional reasons. This is because an agent's ability to give moral judgements requires that, the agent is able to understand moral reasons. Furthermore, I claim that agents must have affective and cognitive empathy to make moral judgements. In other words,

[22] The transgressions that harm the victim's welfare are often referred to as transgressions of disgust, but I argue that the disgust transgressions are the same as moral transgressions.
[23] There is a full criterion of the difference between moral and conventional norms which is laid out clearly in 'Harm, Affect & the Moral Conventional Distinction'. See Kelly et al., (2007, p. 118) for further details (it should be noted that Kelly refers to 'reason' rather than 'transgressions').

they must understand why it is wrong to make a moral judgment and be able to sympathise with the other person, they need to be able to put themselves in other individuals' shoes. Otherwise, a seemingly moral judgment without empathetic justification is actually a conventional judgment. This is because to make a moral judgement, one must give reasons of why something is wrong. They must be able to give both cognitive reasons and affective reasons (as to how one feels) for why that action is wrong. For example, if one says that killing is wrong, one must both understand and be able to give reasons why killing is wrong (i.e. because it causes people harm) and affectively, where they explain why killing is wrong in terms of other peoples' emotions. Moreover, these reasons cannot be arbitrary reasons, but rather they need to be in terms of harming the other person or people. For example, if an agent makes the claim that it is wrong to kill innocent babies, and the agent is asked to give reasons as to why it is wrong to kill babies, for the claim to be a moral judgment, one needs to be able to describe the reasons in terms of harm. However, if an agent makes a claim that killing babies is wrong because they will go to prison, this means that the claim of killing babies is wrong turns into a conventional judgement.

However, a problem arises as it is difficult to know when an agent is making a conventional claim when they are seemingly making a moral claim. The only way of knowing this fully is if the agent was asked why the actions referred to in the claim was wrong, and they were able to give moral reasons (i.e., in terms of harm) as to why the claim was wrong. However, simply giving a moral reason is not going deep enough in ascertaining who makes a moral claim. After all, it could be the case that the psychopath merely utters the words 'killing babies is wrong' without actually meaning those words. Therefore, I claim that a distinguishing factor between a moral and a conventional action is empathy. The reason for arguing this is that the difference between a moral and a conventional action is the agent's ability to identify harm to another agent; in other words, a moral action will be moral in virtue of it not causing pain to another agent (I mean by 'pain' physical or mental anguish). Moreover, for a moral claim to, in fact, count as a moral claim, one needs to feel empathy; they need to have both cognitive empathy (understanding why something is wrong from the other person's perspective) and have affective empathy (feeling why something is wrong from another person's perspective).

This is why I argue that some autistic people, when they are seemingly making moral claims, are actually making conventional claims. This is because they lack empathy, as shown by their lack of theory of mind. In order to have a proper understanding of why an action was wrong -- in other words, in order to feel why something was wrong-- they would need to have empathy with the other person, which some autistic people lack because of their lack of theory of mind. Therefore, some autistic people are making conventional judgements rather than moral judgements when they are seemingly making a moral claim.

6. Objections to Defending Moral Hermeneutic Fictionalism from the Autism Objection

I now consider some objections to my argument that some autistic people may not understand morality due to their lack of empathy, and some responses.[24] I do this to defend my argument against critics who may endorse the Autism Objection to Moral Hermeneutic Fictionalism.

The first objection may be a clarification of my argument. This is because I claim that some autistic people may not possess empathy, thus they may ask me to clarify whether I mean that empathy is a key component of morality, or whether empathy should be part of our understanding of morality at all. I claim that empathy should be counted as a key component of morality, but I acknowledge that this is only one version of morality. In other words, if you do not endorse empathy as a key component in morality, then my argument will not be appealing. However, the reason why I claim that empathy is a key component of morality is that in order to understand why one has harmed another person (which I argue is a definition of morality), one needs to be able to empathise with the other agent. In other words, they should be able to understand, why an action is wrong from another point of view.

[24] It should be noted that some of my argument and the objection that I discuss can also be applied to psychopaths. However, psychopaths are beyond the limits of my discussion. I do think that psychopaths would be an interesting topic to further research in regard to the Autism Objection, and Moral Hermeneutic Fictionalism, but I argue they need to be considered separately to Autism as psychopaths have different characteristics to autistic people.

The next objection that one may pose to my argument is the fact that social conventions may in fact require a theory of mind like morality does. Therefore, those who lack empathy and therefore morality may also lack the understanding of social conventions. According to this argument, some autistic people do not seem to lack an understanding of social conventions, and they have the ability to follow conventions, therefore, some autistic people do not lack empathy to a large extent. However, I claim that social conventions do not require empathy or require a different kind of empathy to the one that morality does. Social conventions instead require a mimicking or a copying of other people's actions or following a set of norms and values (like the law). One does not need to understand why these exist, or whether they are right or wrong; one only needs to follow them. This following process, I claim, does not require empathy, as one is not following the rules because one thinks that it is wrong from the other person's perspective, but rather following the rules because it is the law and there are consequences for not following the law.

The next criticism that one could have against my argument is based on Yablo (2001, pp. 97–99), which is that we do not necessarily need to understand that we are acting in a fiction, we just need to be able to act in accordance with a fiction. It should be noted that Yablo is talking about Mathematical Fictionalism, whereas I am talking about Moral Hermeneutic Fictionalism. However, this objection could also be posed to my argument. My response would be that it would really depend on your definition of engaging in a pretence. One could act in a pretence, but I think that to engage in a pretence, one needs to be able to understand that they are engaging in a pretence and why they are engaging in a pretence in order for them to be fully immersed in the pretence.[25] In other words, I think in order to engage in a pretence, one needs to understand morality and be able to engage in morality. Some autistic people can only engage in morality therefore they are not being moral in a full sense.

[25] It should be noted that this claim in this paragraph could throw some doubts on my response to the Phenomenological objection. However, the Phenomenological objection is not the focus in my thesis, so I leave the objection open.

7. How does Casting Doubt on Some Autistic People not Understanding Morality Impact the Relationship Between Moral Hermeneutic Fictionalism and the Autism Objection?

In Section 5, I established that some autistic people may not be able to make moral judgements because of their lack of theory of mind as this shows that some autistic people lack empathy. I have also explained what some autistic people are doing when they seemingly make moral judgements. I have claimed that autistic people when they are seemingly making moral judgements, are actually making conventional judgements. This impacts the viability of the Autism Objection against Moral Hermeneutic Fictionalism, as if some autistic people struggle to understand morality, then it may mean P4 fails. This means that some autistic people cannot understand pretence, but they also may not understand morality. In other words, the Autism Objection may not work against Moral Hermeneutic Fictionalism as it is reliant on the premise that some autistic people do not understand pretence, but do understand discourse X, but in this case, some autistic people may not understand discourse X (morality in this case), so the objection may not work.

This argument is not a definitive argument (which is why I said the objection may not work, and I do not propose a more definite claim) as it is unclear whether some autistic people understand morality or not. There is much literature on both sides, and the debate is still very much current, and findings might change with further research. This argument is also not a definitive argument, because one might have different definitions of what morality amounts to, and some of them may not require empathy. If the definition of morality does not require the concept of empathy, then it might be the case that some autistic people understand morality (and act morally), as they may empathise fully. However, I argue that any definition of morality needs to include empathy. Therefore, further research needs to be done on whether morality should incorporate empathy as this will impact the relationship between Moral Hermeneutic Fictionalism and the Autism Objection (Stanley, 2001, p. 48).

8. Conclusion

In conclusion, I have argued that the Autism Objection to Moral Hermeneutic Fictionalism may not be viable because one of the key premises in the argument may not work, which is that some autistic people understand morality (P4). This is because my account of morality relies on empathy as a key component, and some autistic people lack empathy because of their lack of theory of mind. To respond to the largest objection to my argument that these autistic people act morally so they must be moral, I have also given an explanation as to what I think some autistic individuals might be doing when they seemingly make moral judgements; I have argued that they make conventional judgements, and this is why they seemingly act morally without actually being moral. This psychological and scientific research is changing rapidly in this area and our understanding of some autistic people having a deficit in empathy is changing, so this discussion will need to change depending on the findings of the psychological research. However, at this time, I can conclude that the Autism Objection may not work against Moral Hermeneutic Fictionalism, as there is doubt as to whether some autistic people understand morality.

Part 2

Chapter 3: What is Motivational Internalism? An Investigation of the Key Debates

1. Introduction

The purpose of this chapter is to explain Weak Motivational Internalism and to explore some arguments for and against it, to set the context for the discussion in the next chapter. I motivate the credibility of Weak Motivational Internalism to set the scene for my argument, which I propose in the next chapter, which is that agents with depression who have certain psychological conditions, (which are that a person judges that X is right to do in C, and at the same time is not motivated at all to do X in C, and at the same time is rational) may pose a counterexample to Weak Motivational Internalism. I propose that Weak Motivational Internalism should be taken seriously by doing the following in this chapter: first, I give an overview of the landscape of Motivational Internalism (and Weak Motivational Internalism). I then discuss why one might find Weak Motivational Internalism attractive. Next, I discuss some objections to Weak Motivational Internalism and consider some replies to those objections, to position and convince the reader that Weak Motivational Internalism is a plausible theory in metaethical discourse.

2. What is Motivational Internalism?

Internalism is a family of theories which exist in philosophical discourse. Internalism is broadly the theory that a judgement about X (at a particular time) necessarily leads to a motivation to do X (at that time).[26] The first thing to note is that any reference to a judgment in this chapter will be defined as a mental act or a mental state. I specify my meaning of judgment because other scholars have referred to judgment in alternative ways, such as referring to a judgment as a speech act

[26] It should be noted that different forms of Internalism (which have different philosophical claims) have important ramifications for many branches of philosophy such as moral epistemology, semantics and ontology (for further details on different kinds of Internalism see Audi, 1998; Bjornsson et al., 2015, p. 2; Brink, 1989; Darwall, 1995, 1983; Parfit, 1998; Shafer-Landau, 2003, pp. 144–145; van Roojen, 2013).

rather than a mental state (Bjornsson et al., 2015, p. 2) (for further discussion on how Motivational Internalism claims that there is a necessary connection between moral speech acts and motivation (see Joyce, 2002; Ridge, 2006) and Ridge (2006)). Other versions of Internalism require moral judgements to be motivational states; in other words, this means that motivation is internal to moral judgements (Bjornsson et al., 2015, pp. 2–3) (see Zangwill, 2015 for questions concerning the motivational efficacy of moral judgements). There are also different discourses which can be applied to Internalism, such as Existence Internalism or Reasons Internalism which should be distinguished from Motivational Internalism (Bjornsson et al., 2015, p. 3) (for further discussion on Existence Internalism see Darwall (1983) and for further discussion on Reasons Internalism see Goldman (2005) and Spaid (2021). My focus in this chapter, however, will be regarding Weak Motivational Internalism, but I explain briefly about the other types of Internalism to give sufficient background.

There are many forms of Motivational Internalism (such as Strong, Weak, Weakest, Conditional, and Deferred) but Weak Motivational Internalism will be the theory that is focused on in this chapter (see Björklund et al., 2012).[27] Weak Motivational Internalism is the claim that "if an agent judges that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is practically irrational" (Smith, 1994, p. 61). However, to understand the landscape, I discuss the key theories surrounding Weak Motivational Internalism and the relationship between the theories and Weak Motivational Internalism.

First, as Motivational Internalism is a rival to Motivational Externalism, I turn to Motivational Externalism. Motivational Externalism says that a moral judgement does not necessarily have to lead to a motivation. The connection is in virtue of the agent's desires (Fisher, 2011, p. 129). In other words, the moral judgement is not

[27] For further discussion on other kinds of Motivational Internalism, see Björklund et al. (2012) and van Roojen (2013). It is important to note that Darwall (1997, p. 308) made a helpful distinction between different forms of Internalism i.e. Moral Judgement Internalism, Existence Internalism, Perceptual Internalism and Metaphysical Internalism. There are also other forms of Internalism such as Deferred Internalism where the judgement and the motivation that is necessarily connected are made by the same person but at different times, for example, see Blackburn (1998, pp. 61–65), or Conditional Internalism (see Greenspan, 1998; Tresan, 2009, p. 180 who advocate for a communal version of Deferred Internalism; see also Bjornsson et al., 2015, p. 11). I only talk about Moral Judgement Internalism in this chapter and so focus on this. For further elaboration on the different types of Internalism see Darwall (1997), van Roojen (2015), or Bjornsson et al (2015).

necessarily connected to the motivation. Although the moral judgement can motivate an agent to do an action, it does not need to. Moreover, if an agent judges that Φ is right but fails to be motivated to do Φ, this is consistent with Motivational Externalism. For example, suppose I make a moral judgement in the morning that getting out of bed is the right thing to do. In that case, I may be motivated to get out of bed (in the morning), or I may not (the fact that I am not motivated may be because of other features of my psychology, such as having a desire not to do the action).

Strong Motivational Internalism is a competing version of Motivational Internalism to Weak Motivational Internalism. Strong Motivational Internalism claims that "if an agent judges that it is right for her to Φ in circumstances C, then she is motivated to Φ in C" ((Smith, 1994, p. 61) (for further examples of Strong Motivational Internalists see McDowell, 1979, 1985; McDowell and McFetridge, 1978; Nagel, 1970; Platts, 1980, 1979)).[28] For example, if I judge that helping people is right to do in certain circumstances then necessarily in those same circumstances I will be motivated to help people. This also means that with Strong Motivational Internalism, even if I have an irrational moral judgement at a particular time, I will still be motivated to act according to the irrational judgement at that time. For example, if I judge that it is right to jump off a building at nine o'clock in the morning (for no rational reason), necessarily I will be motivated to jump off that building at nine o'clock in the morning. Furthermore, it should be noted that both Strong (and Weak) Motivational Internalism do not entail that the agent making the moral judgement actually performs Φ but only that they have some motivation to Φ. They may not Φ for many different reasons, for example, it may not be feasible to Φ, or it may be impossible to Φ. This is demonstrated by the following case: an agent judges that it is right to keep a promise in certain circumstances at a particular time and thus will have (at least) some motivation to do that promise in those circumstances (at that time). It may be the case that motivation does not lead to action (and this is

[28] Motivational Internalism also can be applied to other moral claims other than the judgements whereby 'the agent judges what is right to do X in C'. In other words, Motivational Internalism can be spelt out in different ways, for example, the agent judging that X is good in C and therefore necessarily being motivated to do what is good in C. For further details of examples of Weak Motivational Internalism being spelled out in other ways see for example, McNaughton (1988, p. 118).

consistent with Motivational Internalism). In the case above, for the agent's promise to be fulfilled, they may require money, but they may not have enough money at that time to fulfil the action. Thus, they may not be able to perform the action (in this case the action is fulfilling the promise).

However, critics have proposed a counterexample to Strong Motivational Internalism which is that some individuals that seem to be irrational, appear to make genuine moral judgements (at a particular time) but fail to be motivated by these moral judgements at that time. This is demonstrated by the following example: imagine that an agent judges that it right to give to charity at 9 o'clock in the morning (on a particular day). If I do not have some motivation to give to charity at 9 o'clock on that particular day for any reason (including reasons involving practical rationality), then, according to the Strong Motivational Internalist, I am not making a genuine moral judgement that giving to charity is right at 9 o'clock in the morning (on that particular day). Many philosophers have seen problems with the strength of this claim and thus have developed a response which is Weak Motivational Internalism.

Weak Motivational Internalism makes a weaker claim to respond to this counterexample. As previously stated above, Weak Motivational Internalism is the claim that "if an agent judges that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is practically irrational" (Smith, 1994, p. 61) (examples of Weak Motivational Internalists are (Smith, 1994), (Pettit and Smith, 1993), (Blackburn, 1984)).[29] Examples of cases of practical irrationality include individuals who suffer from weakness of will, the inability to concentrate or fatigue or through mental illness such as depression, apathy, or despair, (see Stocker, 1979, p. 744; Svavarsdottir, 1999, pp. 163–164 for further details). More precisely, the

[29] Practical rationality is when an agent is sound of mind and is not suffering with any conditions that constitute irrationality, for example, weakness of the will and body, accidie and physical (and spiritual) tiredness. For further discussion on examples of practical irrationality, see Stocker (1979, p. 744)

version of Weak Motivational Internalism, which is spelt out in this paragraph, is the version which my thesis will focus on and criticise.[30]

Next, I discuss the wider significance of Motivational Internalism and how it relates to other theories in Metaethics. According to Bjornsson (2015, p. 3), metaethical literature has been focused on Motivational Internalism because of Motivational Internalism's role in arguments against Cognitivism (Cognitivism is the view that moral judgments are, or express, beliefs ((van Roojen, 2015, p. 294) (also see Fisher, 2011, p. 6)). This is because critics (to Smith) suggest that if we accept the Humean Theory and Motivational Internalism then we will need to reject Moral Cognitivism, because the Humean Theory of Motivation tells us that beliefs alone do not motivate and Motivational Internalism tells us, that moral judgement necessarily motivates (Smith, 1994, pp. 9–10).[31] Therefore moral judgements cannot be beliefs and thus we have to reject Cognitivism (Smith, 1994, pp. 9–10). Moreover, some argue that Cognitivism is compatible with Internalism (for example, see Bjornsson et al., 2015) (also see Bromwich, 2010; Dancy, 1993 Chapter 1-3; Garrard and McNaughton, 1998; McDowell, 1979; McDowell and McFetridge, 1978; Nagel, 1970; Pendlebury, 2002; Scanlon, 2000 Chapter 1; Shafer-Landau, 2003 Chapter 5; Tenenbaum, 2006; van Roojen, 2002 for further details). This is because scholars argue that Cognitivism and Internalism are compatible as Cognitivism states that moral judgments express beliefs and Internalism says that moral judgments necessarily motivate, meaning that one can say that moral judgments, which are beliefs,

[30] It should be noted that most of the recent philosophical discussion looks at Weak Motivational Internalism rather than Strong Motivational Internalism because Weak Motivational Internalism is thought to be more plausible. Furthermore, Miller (2008, p. 234) claims there is another name for the distinction between Weak and Strong Motivational Internalism, which is Unrestricted Motivational Internalism (Strong) and Restrictive Motivational Internalism (Weak). Unrestricted Motivational Internalists argue that the relationship between moral judgement and motivation is true of all moral agents "no matter what the makeup of their characters or their rational capacities happen to be." (Miller, 2008, p. 234). This contrasts with the Restrictive Motivational Internalist, who only commits themselves to "the truth of the relevant thesis when it comes to the motivation of a particular class of agents – so-called normal agents" (Miller, 2008, p. 234).

[31] Moreover, Smith notes that if we accept the Humean Theory and Moral Cognitivism then we will reject Motivational Internalism (Smith, 1994, pp. 9–10). This is because Moral Cognitivism says that moral judgements are beliefs and the Humean Theory says that the relevant belief and the appropriately related desire motivates an agent, and therefore we cannot say that moral judgements motivate and thus this is incompatible with Motivational Internalism. Moreover, if we accept Moral Cognitivism and Motivational Internalism then we reject the Humean Theory of Motivation because moral beliefs cannot motivate themselves according to the Humean(Smith, 1994, pp. 9–10). This is what Smith calls The Moral Problem.

motivate.[32] Others argue that Cognitivism and Internalism are compatible with each other, even though beliefs are inert (they are descriptions of the world, so they just sit there and therefore they do not influence the world).[33] Different authors say that Cognitivism and Internalism are related in different ways – for example, Smith, (1994) says that Cognitivism and Internalism are related in the content of the moral belief. Moreover, Dreier (1990, p. 14) says Internalism and Cognitivism are related because of the way that the content of the moral motivation is dependent on the content of the moral belief. Moreover, Dancy (1993, p. 193) says that there is a special type of motivational state which relates to both belief and motivation.[34]

The next chapter (Chapter 4) will argue that certain cases of depression with certain conditions (that the agent judges Φ to be right in C and at the same time lacks motivation and at the same time is rational) may provide a counterexample to Weak Motivational Internalism. Therefore, in what follows, I discuss some attractive features and some objections (and some responses) to Weak Motivational Internalism, to show that Weak Motivational Internalism is a plausible theory in metaethics.

3. Attractive Features of Weak Motivational Internalism

I first discuss some attractive features of Weak Motivational Internalism, to show why Weak Motivational Internalism is plausible and is worth taking seriously. It should be noted, however, that my aim is not to defend Weak Motivational Internalism entirely, but merely to show how Weak Motivational Internalism is a plausible theory that is worth taking seriously in metaethics.

[32] However, the view of Cognitivism and Internalism combined is thought to be incompatible with the Humean Theory of Motivation (however Smith (1994) argues that all three theories can fit together: Cognitivism, Internalism and the Humean Theory.
[33] For further discussion on how beliefs are inert see Fisher (2011, p. 134)
[34] It should be noted that some authors (such as Brink (1989) Chapter 3) argue that if Internalism is false then so are Moral Non-Cognitivism and Moral Expressivism (Bjornsson et al., 2015, p. 4).

3a. Appealing to An Intuitive Thought

Weak Motivational Internalism appeals to a natural thought we have about how we are motivated, which is that, when we judge X is right to do in C, then we are necessarily motivated to do X in C (or we are practically irrational). It seems to be the case that Weak Motivational Internalism appeals to a natural thought that occurs and there seems to be an expectation that one would be motivated in the way that the Motivational Weak Internalist describes. In other words, one can imagine a PhD student who judges that it is right to write their PhD thesis for a particular deadline will feel motivated to finish their thesis by a particular deadline. In other words, Internalism seems to be a good descriptor of how we think about motivation, i.e. if I judge something to be right (at a particular time) then I am motivated to do it (at that time). Therefore, Internalism seems to be a good predictor of how one is motivated.

More precisely, I now separate the argument into steps. The first step is when one hears that an agent has a moral judgement; we have certain expectations of what they are going to do (i.e. we have certain natural expectations that people are going to be motivated by that judgement). The second step is that one needs to ask oneself why one has this expectation and what explains this expectation. The third step is the answer to the second step, which is that Internalism seems to be able to explain the pattern of behaviour illustrated when somebody is motivated (i.e. an agent judging that an action is right at a particular time and then necessarily the agent being motivated by that judgement) is something intuitive to agents when individuals discuss how people are motivated. Therefore, if I judge that giving to charity is right at a particular time then one would expect me to be motivated at that time to give to charity.

One case which may be able to support the idea that Internalism fits with a natural thought pattern is the case of cognitive behaviour therapy (CBT). CBT is used to treat many mental health conditions because of the practice of separation of thought, feeling and action. This is because the CBT therapist makes a natural assumption that a judgement and a motivation are linked, and this is how they treat patient disorders (for example clinical depression, clinical anxiety, schizophrenia). There is evidence of this in psychological literature (for example, see Gilson et al., 2009, pp.

25, 27). In other words, if judgements were not necessarily linked to motivation this raises the question of why CBT is useful in treating mental health disorders (for example, CBT is commonly used successfully in cases of depression (for evidence of this, see Beltman et al., 2010; van Straten et al., 2010)). This is because in CBT the therapist relies on the fact that if an agent is able to change one mental state, then this influences the other mental states associated with it. More precisely, CBT is broken down into thoughts, emotions, behaviour and actions (this can vary depending on the source of information that one is referring to, but it is generally split into these mental states). The idea with CBT is that if you change one of the mental states above, then this will have an impact on the others. I claim there may be a resemblance between CBT and Motivational Internalism as the broad idea of both CBT and Motivational Internalism is that if one changes an agent's judgment then one changes an agent's actions or motivation.[35] There is no reason why this pattern of CBT would not apply to moral judgments either, and although there is no direct evidence that CBT works with moral judgments, there is much evidence that CBT affects agents' judgments as these judgments affect an agent's motivation. This is why CBT as a therapeutic practice is effective.

Another case study that indicates the necessary connection between moral judgement and motivation is when you are teaching a child not to do a certain action (because it is wrong). Imagine a case where an agent has a small child, and the child has done something wrong. A parent/guardian may discuss the wrong action with the child in the hope of changing the judgement of the child, so they will not be motivated to do the wrong action again. In other words, if there is a relationship between a judgement and a motivation, then this would account for why we adopt practices like discussing with a child why certain actions are wrong or CBT therapy when trying to teach and help individuals. Furthermore, if changing a judgement did not change a motivation then there is a question of why we would behave in this way. Therefore, the reason that one finds Weak Motivational Internalism attractive

[35] I acknowledge that this argument could also apply to the Motivational Externalist as well as the Motivational Internalist. This is because there is also a connection that the Externalist endorses but the connection is not necessary between a judgment and a motivation. However, for my purposes, there is a pattern in psychology, which is shown by CBT therapy, that shows a resemblance to Motivational Internalism.

is because it marries up with how we expect people to be motivated (unless people are practically irrational).

3b. The Debate Between Weak Motivational Internalism and Motivational Externalism: The Explanation for the Striking Fact

The second reason to find Weak Motivational Internalism (and Motivational Internalism) attractive is because the Motivational Externalist's explanation of the striking fact may seem unsatisfactory (Smith, 1994, p. 71). Smith claims that the striking fact is a claim about how one is motivated. More precisely, the striking fact claims that "*a change in (moral) motivation* follows reliably in the wake of *a change in moral judgment*" (Smith, 1994, p. 71). Smith (1994, p. 71) chooses to use this claim, because this is a claim that both the Motivational Internalist and the Motivational Externalist agree with. However, Smith claims that the Motivational Externalist does not give a plausible explanation as to why the agent is motivated when they make a moral judgement, whereas Smith claims that the Motivational Internalist does give an adequate explanation. This is because Smith claims that the Motivational Internalist can explain the link between the moral judgement and the motivation, however, Smith claims that a Motivational Externalist has a sub-standard explanation (Smith, 1994, p. 71). This is because the Motivational Externalist would describe the striking fact in terms of a good and strong-willed person and thus deny the necessary connection.

More precisely, there are two possible answers that one could provide towards the striking fact: the first one comes from the internalist theory – the Internalist can say that motivation can be explained internally i.e., judgement necessarily leads to motivation (and it follows directly from the content of the moral judgement itself) - the belief that the act is right produces a corresponding motivation (Smith, 1994, p. 72). However, the second answer (which comes from the Externalist) says that the striking fact can be explained in terms of the content of the motivational disposition possessed by a good and strong-willed person. This is because, as the Externalist claims, that the good and strong-willed person is someone who aims to act in a right way and because they are strong-willed, they will successfully do whatever they believe to be right. In other words, if one gets the agent to change their behaviour

and then the agent changes their view of what is right, one will be motivated to do what they judge to be right. According to the Motivational Externalist, rather than there being a necessary connection between judging what is right and being motivated to do what is right, there is a contingent connection between an agent judging something to be right and being motivated to do what they judge is right, and this exists in the good and strong-willed person.

Smith claims that the first answer is plausible, that judging that something is right (at a particular time) necessarily leads to a motivation at that time. This is because with the Internalist, if somebody judges that it is right to do ϕ in C, then they are motivated to do ϕ in C or they are practically irrational (Smith, 1994, p. 72). However, Motivational Externalists would claim that Motivational Internalists leave out a crucial part, which is that there is no need for a necessary connection between judgments and motivation, but rather, according to the Motivational Externalist, individuals are motivated by just being a good and strong-willed person (and thus, there is no necessary connection required for motivation), and this is enough to provide the explanation for the robust connection between the change in judgement and the change in motivation in the striking fact (Smith, 1994, p. 73).

However, Smith claims that Motivational Internalism gives a more plausible explanation, as the Motivational Internalist can explain the striking fact in terms of the that judging that X is right at a particular time, necessarily leads to the motivation to do X at that time (Smith, 1994, p. 71). This is because the Motivational Externalist uses the explanation of a good and strong-willed person and thus, they must explain the striking fact in terms of both *de dicto* and *de re* explanations. The *de dicto* and *de re* distinction is about how one interprets the description of how we interpret propositions, one needs to be able to describe a proposition in terms of both *de dicto* or *de re*. Consider the example of the desire to do the right thing. On a *de dicto* reading, one could read the claim that 'Riana desires to do the right thing' under the guise of doing the right thing. This means that Riana will need to have the concept of rightness on a *de dicto* reading of the claim that Riana desires to do the right thing. However, on a *de re* reading, one could read Riana's desire to do the right thing as Riana having a particular desire to do the right thing. More precisely, on a *de re* reading, Riana may desire to give to charity - Riana

may not have the concept of doing the right thing but may have a particular desire to do something that happens to be the right thing. Therefore, on a *de re* reading, the thing that an agent does happens to be the right thing, but this is not the explanation given when one asks why Riana desires to do that action.

The difference between a *de re* and *de dicto* desire is about the content of the desire, i.e. what the concept of desires is focused on (Smith, 1994, p. 74). However, Smith argues that the Motivational Externalist has a problem when they are explaining the striking fact. This is because the Motivational Externalist needs to describe the desire to do the right thing both in *de dicto* and *de re* terms (because of their explanation involving the good and strong-willed person). Therefore, Smith's objection to the Motivational Externalist is that their explanation results in an unacceptable moral fetish and thus, their explanation is unsatisfactory according to Smith (1994, p. 76). This is because Smith thinks that one gets the psychology of the agent wrong as the Motivational Externalist thinks that an agent derives a specific desire (*de re*) from a general desire (*de dicto*) and the *de dicto* desire is therefore fetishistic (Smith, 1994, p. 76) . For example, Riana desires to do what is right (*de dicto*) and she desires to give to the poor (*de re*). Her desire to give to the poor (*de re*) is derived from her desire to do what is right (*de dicto*). Smith therefore argues that this explanation (the Motivational Externalist explanation) posits a moral fetish by claiming the agent desires to do the right thing because it is right. This is because Smith thinks that the good agents just desire to do right things (*de re*) for example, Riana just desires to give to charity without thinking that she should do it because it is right. Moreover, the problem is, according to Smith, that we think of a good agent as someone who is motivated to just do right things (*de re*), for example, in the above case the agent could think that it is right to give to charity, without thinking that they should do it because it is right (*de dicto*). Therefore, Smith claims that the Motivational Externalist does not give a satisfactory explanation of the striking fact.

Contrastingly, Zhang presents an alternative version of the Externalist explanation of the striking fact, which he calls the co-present objection to the fetishism argument (the fetishism argument is that the externalist can only account for the striking fact by referring to a *de dicto* desire to do what is right). Zhang's objection derives from the fact that Smith, in his Externalist description of the striking fact, claims that the

Externalist commits a moral fetish as they can only account for an agent's change in motivation after a change in judgement by pointing to a *de dicto* desire to do what is right and thus Smith claims that the agent does not have direct concerns to do what is right (Zhang, 2021, p. 306). This is why Smith thinks that the Externalist explanation of the striking fact turns agents into moral fetishists (Zhang, 2021, p. 306). However, Zhang argues that we can find different *de re* desires that vary from the psychological mind of the good and strong-willed person (Zhang, 2021, p. 306). It might therefore be a feature of a good and strong-willed person that they have a range of *de dicto* and *de re* desires (Zhang, 2021, p. 307) (See Copp, 1995, p. 212 for further details). Consider an example: imagine that I have a *de dicto* desire to do the right thing, the only difference in the Externalist explanation is that I also still have a *de re* desire to do a particular right thing (Zhang, 2021, p. 307). Copp (1995, p. 212) says that I would be influenced by my *de re* desires (to do a particular right thing) rather than my *de dicto* desire, and the fact that I have a *de dicto* desire makes no difference at all (and thus there is no reason to regard this *de dicto* desire as a moral fetish) as I am influenced by the *de re* desires ((Zhang, 2021, p. 307) (also see Copp, 1995)).[36] For example, suppose I believe that stealing is wrong, and I have a *de dicto* desire to do the right thing and I at the same time believe that stealing a Dairy Milk chocolate bar is the wrong thing to do, and have a specific *de re* desire not to steal the chocolate bar. In this case, it does not really matter that I believe that stealing the chocolate bar is wrong (and have a desire to do the right thing (*de dicto*)), because all that matters is that I thought that stealing that particular Dairy Milk chocolate bar was wrong from that shop, and thus I will be motivated by that particular desire.

There is a second objection to Smith's explanation of the striking fact, which is made by Lillehammer (1997); (see also Fisher, 2011, pp. 133–4). Lillehammer (1997, p. 192) presents the case of a father informing on his son (and if he does so

[36] Some would claim that Copp (1995, p. 212) is mistaken. Critics may claim that there is a respect that the de dicto desire would make a difference if the de dicto desire generates the agent's de re desires. This is demonstrated by the following example. Imagine if what generates an agent's desire to keep a promise is that the agent thinks that keeping this promise, on this occasion, is the right thing to do. In this case, the de dicto desire is generating the agent's de re desire to keep a promise. Moreover, if one changed their de dicto desire then one would also change their de re desire. For example, if I changed my de dicto desire that keeping a promise (at a particular time) was the right thing to do because of an emergency, then my de re desire (at that time) about keeping the promise would also change.

his son will get sent to the gas chambers). In this instance what motivates the father is a general desire to do what is right (*de dicto*), rather than a specific desire (*de re*). In other words, in this example, what motivates the father to do what is right is the general desire to do what is right (as his son has killed and he thinks that it is right to hand murderers to the authorities), whereas he does not have a specific desire to hand his son into the authorities because the father (presumably) feels a duty and love for his son (Fisher, 2011, pp. 133–4). This is a case where a general desire to do what is right motivates an agent rather than a specific desire to do what is right. Therefore, this Externalist explanation of the striking fact seems to be the right explanation and does not posit a moral fetish. Looking at this case then, it might be that the Motivational Externalist explanation is plausible because the Motivational Externalist says that the agent can have a *de dicto* desire and a *de re* desire to do what is right (or they can just have a *de dicto* desire) and therefore the Motivational Externalist explanation of the striking fact may be just as plausible as the Motivational Internalist explanation, as the Motivational Externalist can say the agent has a *de dicto* and a *de re* desire to do what is right (Fisher, 2011, pp. 133–4) and this can be plausible. However, this does not mean that Motivational Internalism is not a plausible theory, rather I claim that the Motivational Internalist has a simpler explanation, which is that judging that X is right in C necessarily leads to a motivation to do X in C.

Fisher (2011, p. 134) also raises an objection to Smith's criticism of the Motivational Externalist explanation of the good and strong-willed person. More precisely, Smith's (1994, pp. 74–75) criticism is that the Motivational Externalist's explanation involves the *de dicto* desire to do the right thing and this desire is a fetish. However, Fisher (2011, p. 134) worries that the Motivational Externalist does not have to claim that this *de dicto* desire has some psychological role to explain the striking fact, which the agent is constantly aware of, and thus is motivated by the *de dicto* desire (2011, p. 134). Moreover, perhaps we need a general desire to do what is right as well as a specific desire to do what is right, which the Motivational Externalist advocates for, but the Motivational Internalist does not advocate for. Furthermore, it is possible that we may have background desires that we may not be aware of but may influence our behaviour and this would be consistent with the *de dicto* desire that the Motivational Internalist describes in the striking fact explanation

(2011, p. 134). However, I think the Motivational Internalist explanation of the striking fact is more intuitive and straightforward, as it only focuses and uses (in their theory) desires that we have consciously, and this further supports the thesis that we should take Motivational Internalism seriously.

However, although there are many objections to Smith's argument (which is about the striking fact), the striking fact explanation shows that Weak Motivational Internalism is a theory worth taking seriously in metaethics as Internalism does give a straightforward and plausible explanation. Although it is debatable whose explanation is better, the Motivational Internalist or the Motivational Externalist, the Internalist still gives a strong explanation that judging X is right to do in C at a particular time necessarily leads to motivation to do X in C at that time and this explanation is simple and fits with how we understand motivation. For example, we think that when we make a judgement that writing an essay is right (at a particular time) then we are motivated to write an essay, and this seems like a simple, plausible explanation.

4. What are the Objections to Weak Motivational Internalism?

This section will establish some objections to Weak Motivational Internalism (and Strong Motivational Internalism) and some potential replies to those objections. I do this to show the landscape of the Internalism/Externalism debate, and I also defend the claim that we should take Weak Motivational Internalism as a plausible theory in Metaethical discourse.

4a. The Amoralist Challenge

The first apparent counterexample, the Amoralist Challenge, is related to my depression counterexample that I will give to Weak Motivational Internalism (in Chapter 4). The reason why the Amoralist Challenge is a potential counterexample to Weak Motivational Internalism, is that an amoralist is an agent who seemingly makes a moral judgement at a particular time but fails to be motivated at that time. My depression counterexample will argue that there are agents with certain conditions that judge X to be right at a particular time but fail to be motivated at that

time, and at the same time are rational, and this seems similar to the amoralist. However, the difference between my counterexample and the amoralist depends on the details of the amoralist discussed. For example, some so-called amoralists like psychopaths judge that X is right to do in C but are not motivated to do X in C (and seem to be rational), and thus the type of amoralist depends on the features of the individual one is discussing.[37] I now discuss the amoralist objection in more depth below.

David Brink (1986, p. 30) makes a crucial objection, which is commonly referred to as the Amoralist Challenge. He puts forward this challenge towards both the Strong and the Weak Internalist. The amoralist is "somebody that recognises the existence of moral considerations and remain unmoved" (Brink, 1986, p. 30). However, the Internalist would claim that the amoralist is impossible and thus, does not exist, as the Internalist claims that there always needs to be a conceptual link between judgement and motivation which the amoralist does not seem to have. More precisely, a conceptual link, which the Internalist endorses, is that if an agent judges it right to X at a certain time, then they are necessarily motivated to do X at that time or they are practically irrational.

However, as previously stated, the Internalist denies the existence of the amoralist. This is because the Internalist claims that all people who are not motivated by their moral judgements, and are not practically irrational, merely use these moral judgements in the inverted comma sense; in other words, one cannot make a moral judgement at a particular time without being motivated at that same time ((Smith, 1994, p. 67) (also see Shields, 2016, p. 14)).[38] For example, when an amoralist makes a moral judgement of 'I should give to charity', they are actually claiming something like 'Most people think that 'I should give to charity' and look down on those who do not'. This means that the amoralist uses moral terms in an *inverted comma* sense (Hare, 1952a, pp. 124-6 165-70). This is further demonstrated by the

[37] There has been much literature on why psychopaths are not morally motivated when they seemingly make a moral judgement (and how this links to Motivational Internalism). This is because it seems to be the case that psychopaths make a moral judgement but fail to be motivated. For discussion on psychopaths, see Roskies (2008, 2006); Cholbi (2006); Kennett and Fine (2008); Smith (2008).

[38] This has been a key objection first established by Hare (1952b, pp. 124–6, 163–5). See Hare for further details.

example in which an amoralist utters ‘murder is wrong’ – an amoralist may actually be conveying ‘there is a convention (or a rule) where people say that murder is wrong’. Furthermore, the amoralist can make the claim about a convention without having any motivation to conform to the convention, and this may be what is occurring with how the amoralist is dealing with morality – they may know and utter the convention, but they may have no motivation to conform to the convention. Another way that this could be described is the amoralist switching moral demands for conventional demands (but still using moral terminology to express conventional demands) and this is why the amoralist remains unmoved by an apparent moral judgment (Smith, 1994, p. 67).[39]

Smith’s view is not quite as plausible as it may first appear, because I claim that one can make moral utterances without being motivated to be moral. This is because there seem to be many cases, for example, psychopaths, where individuals for various reasons make a moral judgement (at a particular time) but fail to be motivated (at the same time) by that moral judgement. However, Weak Motivational Internalism can reply to the amoralist objection by saying that the amoralist must have some motivation (albeit minimal) to do action X (at a particular time) because they are making a moral judgement to do X at that time, or they must be practically irrational. The Weak Motivational Internalist says that agents who have a genuine moral judgement (at a particular time) but do not appear to be motivated by the judgement (at the same time), are actually motivated at that time, but their motivation is so minimal that they do not appear to be motivated at that time. I can see three possible replies to the amoralist objection that the Weak Motivational Internalist can give. The first is that the Weak Motivational Internalist may claim that the agent is not really making a moral judgement to do X at a particular time, and therefore they are not motivated to do X at that time. The second is that the amoralist, according to the Weak Motivational Internalist, is in fact motivated, despite their appearance of not being motivated. The third is that the amoralists, according to the Weak Motivational Internalists, is practically irrational and in fact

[39] James Lenman (1999, p. 443) describes the inverted comma objection, by describing the Amoralist as somebody who is making a moral judgement as a sociological claim rather than a moral claim (Lenman, 1999, p. 443).

the amoralist does fit into the Weak Motivational Internalist claim of how one is motivated.

Brink (1986, p. 30) disagrees with the Weak Motivational Internalist by claiming that the idea of the amoralist is not taken as seriously as it should be. This is because Brink believes that as a condition of making moral assertions, one needs to have a good sense and understanding of how to use moral terms in our society. However, Smith disagrees with Brink, saying that when somebody judges that X is right to do in C and therefore necessarily, they are motivated to do X in C, needs to be adhered to if an agent is making genuine moral judgements. This is because Smith proposes that Brink is wrong when Brink claims that the only condition for making moral utterances is that one needs to be able to use moral terms in society (Smith, 1994, p. 70).

However, the Externalist and Internalist reach an apparently unbreakable deadlock about the amoralist. The Internalist would say that the amoralist does not exist, that the agent is simply failing to make a genuine moral judgement (at a particular time) and therefore fails to be motivated (at that time). Moreover, the Internalist could also claim that the agent (who appears to be an amoralist) is actually somewhat motivated to do what they judge to be right, even if they do not appear to be. The Externalist however would say that the amoralist does in fact make a moral judgement at a particular time, but this moral judgement does not necessarily lead to motivation at that time. At this time, this deadlock is hard to solve, and thus in the next chapter, I will discuss my counterexample of depression (where an agent has certain conditions, which I will specify in the next chapter) to help solve the debate between the Externalist and the Internalist.

4b. The Restricted Class of Explanation Objection

Svavarsdottir (1999, p. 179) (see also Bjornsson et al., 2015, p. 14 for further details on this objection) makes an objection against Weak Motivational Internalism. She claims that the Weak Motivational Internalists exclude many people in their explanation of how people are motivated by only describing how rational people are motivated, rather than providing an explanation for both the rational and irrational

person (which a Motivational Externalist does).[40] In other words, Weak Motivational Internalists are limited to rational people when they describe motivation. For example, imagine an agent called Adam (who has weakness of will) who judges it is right to get out of bed at a certain time, but is not necessarily motivated to get out of bed at that time. The Weak Motivational Internalist explanation only gives two possible explanations for why he is not motivated. The first is that Adam does not make a genuine moral judgement (at a particular time) so therefore fails to be motivated at that time, or the second is that Adam is practically irrational. Therefore, the Weak Motivational Internalist is limited in the explanation that they can provide for cases of practical irrationality. However, the worry for the Motivational Internalist is that the Motivational Externalists have a more inclusive explanation as to who is motivated as they do not need the clause of practical irrationality. In other words, there is no need for Motivational Externalists to call agents practically irrational as they do not rely on a necessary connection between a genuine moral judgement and motivation. This is because they claim that when an agent makes a moral judgement, the agent merely *might* be motivated by their moral judgement, and this is dependent on other factors (e.g. practical rationality, clinical depression). Furthermore, Svavarsdottir (1999, p. 181) says that the burden of proof lies with the Weak Motivational Internalist, as Weak Motivational Internalism is a theory that is

[40] It should be noted that some authors such as Olinder (2012, p. 579) say that Svavarsdottir might be using a different kind of Motivational Internalism to Smith. Svavarsdottir is using the definition, "moral judgments are a conceptual necessity connected to motivation to pursue or promote what is judged favourably and to shun or prevent what is judged unfavourably, except in individuals suffering from motivational disorders that affect them more generally" (Svavarsdottir, 1999, p. 165). As you can see, Svavarsdottir is not discussing an agent who is motivated because they judge an action to be right, rather the agent is judging what they deem to be more favourable. The distinction between what is judged more favourably and what is judged to be right is subtle yet important. This is demonstrated by the fact that, for example, I could judge that I would not like to (and thus not find it favourable to) finish my work tomorrow (at a particular time) and thus, from the Internalist viewpoint, not feel motivated to do my work. This is a slightly different thought to saying that I judge it is not right to do my work (as there may be a reason why I think it is not right at a particular time, for example, it may be that I should be spending time with my family and therefore not feel motivated to do my work at that time). Moreover, it is consistent to say that I judge something to be right but think of it as unfavourable, for example, I may judge that not smoking is right at a particular time, but because I am a smoker I may judge it more favourable to smoke. However, Svavarsdottir's objection does not rely too heavily on the word "favourable" in her definition of Motivational Internalism, so one can still apply her objection to my version of Internalism.

more restrictive with their explanation of how people are motivated, as the type of individuals that it gives an explanation for are only rational individuals.[41]

Dreier (1990, p. 12), however, has ideas that we can use to respond to Svavarsdottir's burden of proof criticism. He claims that we need to carefully distinguish which cases of abnormal motivation are practically irrational before claiming that the Weak Motivational Internalist limits the scope of individuals. Furthermore, the worry that Drier (1990, p. 12) presents is that the Weak Motivational Internalist has to provide an answer as to why these conditions are abnormal, otherwise, the Weak Motivational Internalist may look like they are merely passing judgment on what constitutes a moral judgment, rather than how one makes a moral judgment. For example, say an agent makes a judgment that giving to charity is right in circumstances C but fails to be motivated in circumstances C, then we need to explain why this motivation pattern occurs rather than saying that the judgment was not a moral judgment (Dreier, 1990, p. 12). Furthermore, Drier (1990, p. 12) claims that the Motivational Internalist looks to 'list-like' as there is no explanation as to why the conditions the Weak Motivational Internalist specifies as abnormal are actually abnormal. Moreover, the worry is that we might merely rule these conditions as abnormal because they are not motivated by the agent's supposed judgment, which is not giving an adequate explanation as to how agents are motivated. Drier (1990, p. 12) worries that any list that we come up with could be re-defined by the Weak Motivational Internalist as abnormal conditions if the agent is not motivated in C by those judgments. However, Drier (1990, p. 12) does support the Internalist by saying it looks as if all the cases of practical irrationality the Internalist defines do look like they have something in common, but he questions what the commonality between these conditions amounts to. To Dreier, the solution might be to differently define what we mean by abnormal or practically irrational to incorporate more individuals into the Motivational Internalist explanation (Dreier, 1990, p. 12). Dreier claims that if we make the Weak Motivational Internalist explanation more inclusive and make practical irrationality include more individuals,

[41] It should be noted that Svavarsdottir (1999, pp. 180–182) claims that her objection settles the deadlock between the Motivational Internalist and Motivational Externalist. This is because, as said above, the Motivational Externalist is less restrictive on the type of motivation they can include. For further discussion, see Svavarsdottir (1999, pp. 180–182).

then the Weak Motivational Internalists may be able to defend themselves against Svavarsdottir's objection highlighted above. Moreover, if Weak Motivational Internalists take on Dreier's suggestion, this may result in a more defensible version of Weak Motivational Internalism, and this could serve as a response to Svavarsdottir's objection.

Mabrito ((2013, p. 193) (also see Dreier, 2000)) additionally notes that Motivational Externalists also have a problem with Svavarsdottir's objection as they also restrict the individuals who they provide motivational explanations for. This is because Motivational Externalism has to explain how agents are morally aware i.e., how they are motivated to perform the right action. This explanation is both available to the Motivational Internalist and the Motivational Externalist. However, the Motivational Internalist's explanation is clearer and simpler; because they claim a reliable connection which is that judging that X is right to do in C necessarily leads to a motivation to do X in C (unless they are practically irrational). The Motivational Externalist, however, relies on the explanation of the good and strong-willed person which some critics would say is vague and open to interpretation (Mabrito, 2013, pp. 202–3). Mabrito notes that if the Motivational Externalist wants to exclude the Motivational Internalist explanation they must give us a clear reason to do so and a clear reason as to why their theory is more plausible, because at present, Mabrito claims that the Motivational Externalist is unable to. This is because Mabrito claims that the fact that the Motivational Internalist explanation is inconsistent with the Motivational Externalist explanation is not enough of a reason to discredit the Motivational Internalist ((Mabrito, 2013) see also (see also Olinder, 2012, p. 584 for further discussion on how the Motivational Externalist also has a burden of explanation in regards to limiting explanations to only eligible individuals). Therefore, as the Motivational Externalist has the same burden of proof as the Motivational Internalist (and one cannot settle the debate between Motivational Internalism and Motivational Externalism using Svavarsdottir's objection), this means that Motivational Internalism remains a plausible theory in metaethical discourse.

4c. Empirical and Conceptual Evidence for Motivation

There is a question about whether the debate between the Motivational Externalist and the Motivational Internalist is ultimately about empirical evidence or is a conceptual claim or both. More precisely, if one considers the Motivational Internalist and Motivational Externalist debate as an empirical claim, the focus of the debate is about what one should count as making a moral judgement, for example, what is the empirical evidence that people judge that X is right to do in C at a particular time, and therefore are necessarily motivated to do X in C at that time. For example, should we count somebody who says that giving to charity is right at a particular time but has no motivation to give to charity at the same time as making a moral judgement. Should we count this person as in the empirical evidence investigation, as to whether Motivational Internalism or Motivational Externalism is plausible? Moreover, this raises the question of how one should think of real-life examples where agents are motivated to do actions in a way that Motivational Externalism and Motivational Internalism predict. This results in a deadlock between the Motivational Internalist and the Motivational Externalist, as it is difficult to know when an agent is actually making a moral judgement as many agents may look like they are making moral judgements (at a particular time), but they fail to be motivated by the judgement at the same time. The Motivational Internalist would claim that the agent in question is not making a moral judgement at that time (as they are not motivated at that time to perform the action). This contradicts what the Motivational Externalist claims as the Motivational Externalist would claim that the agent (in this case) is making a judgement at that time but simply fails to be motivated by that judgement at the same time. It is therefore difficult to ascertain what counts as plausible in the debate between the two theories, and how we ascertain which of the narratives is more plausible. However, I propose that there is enough evidence to claim that Internalism is viable and thus it should be thought of as a plausible theory in metaethics.

As we can see, the Motivational Externalist and the Motivational Internalist reach a deadlock if we rely on empirical evidence. Moreover, the worry is that if we rely on conceptual evidence (which is more plausible as it is hard to find decisive empirical evidence to settle the debate) to solve the Motivational Internalist and Motivational

Externalist debate, then this would also result in a deadlock, because at the moment, it is the case that if an agent seemingly judges that X is right to do in C and fails to be motivated to do X in C, then there is no way of resolving whether the agent does not make a genuine judgement that X is right to do in C and is therefore not motivated to do X in C, or whether the agent actually makes this judgement (in C), but is not motivated by the judgement (in C). Therefore, we need another objection or counterexample to adjudicate between Motivational Internalism and Motivational Externalism, as all the objections (including this one) discussed in this chapter result in a deadlock between the two theories. This counterexample, I argue, is my depression case, which I hope will contribute to the debate between Motivational Internalism and Motivational Externalism.

5. Conclusion

In this chapter, I have argued that Weak Motivational Internalism is a theory that should be taken seriously in metaethical discourse. I have also laid out the existing literature on the Weak Motivational Internalism (and Motivational Internalism) and Motivational Externalism debate. The consensus of the key arguments of the debate is that the theories reach a deadlock, and thus we need another counterexample to help move the debate forward. Therefore, in the next chapter, I propose a new counterexample, my depression counterexample (which is where a certain group of individuals with depression judge that X is right to do in C, lack all motivation to do X in C, and at the same time are rational in C), and this counterexample, I hope, will help contribute to solving the deadlock between the Motivational Internalists and the Motivational Externalists.

Chapter 4: Weak Motivational Internalism and Depression: An Exploration of the Relationship

1. Introduction

In this chapter, I argue that a particular kind of case of depression appears to be a counterexample to Weak Motivational Internalism. I do this by laying out a case study of a depressed person which shows that certain conditions, appear to exist together. These conditions are that a person can judge that ϕ is right to do in C, (at a particular time) that they lack motivation to do ϕ in C at the same time and they are rational at that time. If these conditions do exist together, which I argue they appear to do, this appears to provide a new counterexample to Weak Motivational Internalism. I propose this in the following way: first, I briefly recap what Motivational Internalism amounts to, then I investigate what depression is and examine some depression testimony. I then explain my depression counterexample and examine each condition of the depression counterexample in turn and show why each condition appears to be plausible. I then examine some objections and replies to those objections to my depression counterexample in order to defend my depression counterexample. Finally, I conclude that my depression counterexample appears to be an objection to Weak Motivational Internalism.

2. What is Motivational Internalism? A Recap

In this section, I outline what Motivational Internalism is; this will be drawn from the previous chapter. I first start with Strong Motivational Internalism; this is the claim that "if an agent judges that it is right for her to ϕ in circumstances C, then she is motivated to ϕ in C" (Smith, 1994, p. 61). This means that Strong Motivational Internalism says that if an agent judges that ϕ is right in certain circumstances (at any given time), then they will be motivated to do ϕ in those circumstances (at that time). For example, if I judge that fulfilling a promise at a particular time is right then I will be motivated to fulfil a promise at that time.

However, in order to account for cases where an agent makes a moral judgement that X is right in circumstances C but lacks all motivation to do ϕ in C due to practical irrationality, Smith and others (Blackburn, 1984, pp. 187–9; Johnston, 1989; Pettit and Smith, 1993) have developed an alternative theory which is Weak Motivational Internalism. As stated in the previous chapter, Weak Motivational Internalism is the claim that "if an agent judges that it is right for her to ϕ in circumstances C, then either she is motivated to ϕ in C or she is practically irrational." (Smith, 1994, p. 61). In other words, according to Weak Motivational Internalism, if an agent judges that it is right to do ϕ in C at any particular time then they will be motivated to do ϕ in C at that time unless they are practically irrational. For example, if an agent judges that it is wrong to kill in the morning, an agent will be motivated not to kill in the morning unless they are practically irrational. Moreover, Smith gives some examples of practical irrationality, such as weakness of the will and other forms of practical unreason in their motivation (Smith, 1994, p. 61).[42]

Now that I have discussed Motivational Internalism, I turn to what depression is.

3. What is Depression?

There are separate forms of depression, but I focus on Major Depressive Disorder.[43] I focus on this because it is the most common to characterise and this disorder displays symptoms that are also present in the other types of depression (the other types of depression have different causes, but they have the same symptoms, see the DSM-5 (APA, 2013, pp. 155–189) for further details).

[42] It should be noted that Smith does not mention depression here in his description of Weak Motivational Internalism. However, others such as Mele, Stocker and Svavarsdottir (1996, pp. 733–34; 1979, p. 744; 1999, pp. 163–4) interpret practical irrationality to incorporate depression so my argument is responding to their claims about practical irrationality.

[43] Some other forms of depressive disorders include Disruptive Mood Dysregulation Disorder, Persistent Depressive Disorder, Premenstrual Dysphoric Disorder, Substance/Medication-Induced Depressive Disorder, Depressive Disorder due to another Medical Condition, Other Specified Depressive Disorders, and Unspecified Depressive Disorder. See the DSM-5 on depression for further details (APA, 2013, pp. 155–189).

Firstly, I describe Major Depressive Disorder. The diagnostic criteria of Major Depressive Disorder require that depressive symptoms have to be present during the same two-week period and they must represent a change from previous functioning (APA, 2013, p. 160). The person who meets this diagnosis needs to exhibit at least one of the core symptoms. The first is that the person exhibits depressed mood most of the day, nearly every day (feeling sad, empty or hopeless) (APA, 2013, p. 160). This could be defined either by the sufferer or through observations made by others (APA, 2013, p. 160). The second main symptom that could be present in Major Depressive Disorder is markedly diminished interest or pleasure in all or almost all activities most of the day, nearly every day (this could be described by the depressed agent or by others observing the symptoms) (APA, 2013, p. 160). These two symptoms of depression are important as if one does not meet the criteria of one of these symptoms, one cannot be diagnosed with Major Depressive Disorder (APA, 2013, p. 160).

A patient also has to display at least five symptoms (including one of the symptoms above) in order for Major Depressive Disorder to be made as a diagnosis (APA, 2013, p. 160). I briefly discuss some of these symptoms as they give a well-rounded picture of what depression is like to live with (APA, 2013, p. 160). The first symptom is significant weight loss when not dieting or putting on weight (APA, 2013, p. 161). The second is insomnia or hypersomnia nearly every day (APA, 2013, p. 161). The third is psychomotor agitation (psychomotor agitation is a state of tension or restlessness ((Day, 1999, p. 90), (see also Ashcroft et al., 1978; Campbell, 1996 for further details)) or retardation (significantly lower mental functioning) nearly every day (observed by others or oneself) (APA, 2013, p. 161). The fourth symptom is fatigue or loss of energy nearly every day. The fifth symptom is feeling worthlessness or excessive guilt or inappropriate guilt. The sixth symptom is diminished ability to think or concentrate, or indecisiveness. The seventh symptom is recurrent thoughts of death, or recurrent suicidal ideation without a specific plan, a suicide attempt or a specific plan for committing suicide (APA, 2013, p. 161). These symptoms cause clinically significant distress or impairment in social, occupational, and other important areas of functioning (APA, 2013, p. 161).

There are different types of severity of depression-- mild, moderate and severe-- and it is important to discuss the range in which people experience depression (APA, 2013, p. 188).[44] The first is mild depression, where few symptoms are present in the individual. The intensity of the symptoms is distressing but manageable, and the symptoms just result in minor impairment in occupational or social functioning (APA, 2013, p. 188). Moreover, in moderate depression the number of symptoms increases and functional impairment increases (somewhere between mild and severe). Finally, there is severe depression where the number of symptoms is in excess to make the diagnosis. The symptoms are seriously distressing and unmanageable, and they markedly interfere with social and occupational functioning.

4. Depression Testimony

I now discuss some testimonies of depression. I do this because I want to illustrate what the experiences of depression are, and how individuals with depression display symptoms. I also want to illustrate the authenticity of these symptoms and show that there are genuine people who suffer with this condition. Moreover, I do this to aid my argument that certain conditions can exist and can coexist together. Because of space limitations, I will focus on three main symptoms of depression that affect motivation. However, it should be noted that there are many symptoms, as explained in the section above, that affect motivation.

[44] This, in my knowledge, has not been done in philosophical works. The closest philosophical work I can find is from Seligman et al (2016). He writes about depression (and anxiety) and applies it to philosophical ideas (but not Motivational Internalism). There are, of course, others that use psychological literature to discuss metaethical and normative theories, for example Spaid (2021) where he writes about Reasons Internalism and its relationship to depression. However, I am writing on Motivational Internalism and depression, and therefore this is slightly different to what Spaid is doing. I am building on Stocker (1979), Svavarsdottir (1999) and Mele (1996), where they discuss the relationship between Motivational Internalism and practical rationality. However, I am discussing one particular example of how certain cases of depression which meet certain conditions (specified in this chapter) provide a counterexample to Weak Motivational Internalism. There is also Cholbi (2011), Milevski (2017) and Miller (2008) who also discuss depression (and other forms of what they refer to as practical irrationality) as a counterexample to Weak Motivational Internalism. However, they argue that depression is a counterexample (and do not specify certain conditions of depression that the person needs to meet in order to be a counterexample like I do) to Strong Motivational Internalism, but I also argue that depression (with certain conditions) is a counterexample for Weak Motivational Internalism. See their papers for further details on their arguments.

The first symptom I look at is depressed mood most of the day, nearly every day (either subjectively reported by the individual or objectively reported by another individual). This will inevitably affect motivation because the agent will struggle to see the point in doing an action because of their depressed mood. This is demonstrated through the following examples:

"When I start to get depressed, I only filter through the negative messages from friends and family, so even the most benign comment can be perceived as an insult. As a result, they soon learn to step on egg shells around me, they become less affectionate because I'm less receptive to it and it generally compounds the situation…" (Ratcliffe, 2014, p. 279).

"My entire focus is on the negative. I cannot see a positive future. I often find myself thinking I could be anywhere in the world right now, doing anything and I would still be desperately miserable. The world holds no possibilities for me when I'm depressed. Every avenue I consider exploring seems shut off." (Ratcliffe, 2018, pp. 128–9)

"During all of this, I felt deeply alone. Everyone else seemed to be moving through their days peacefully, laughing and having fun. I resented them because they were experiencing such an easy time of it; I felt utterly cut off from them emotionally. I was angry because there was no way they could understand what I was going through. Their very presence seemed to magnify my sense of isolation." (Karp, 1996, p. 59)

These quotes show that when the agent is suffering from a depressed mood, the depression distorts the picture of the world around them. This may affect motivation, for example, in the first quote the agent is filtering through negative messages from family and friends, and they perceive all comments as insults and thus relationships change around the depressed agent as this might change how he interacts with the people around him. Moreover, this may cause him to not be motivated to interact with other people. The authors in the other quotes use the phrases 'isolation' and 'lack of possibilities' which also indicates a depressed mood as they have negative views of the world.

The next symptom of depression is diminished pleasure or interest in almost all activities most of the day, nearly every day. This may manifest in the agent struggling to perform activities that they used to find pleasure in, or that they feel like they need to do. This could be related to any or all, aspects of their life, for example, work or family. This is illustrated by this quote:

"You can't ... even remember what it's like to go and do something and feel pleasure from it. You look at the world, the array of things that you could do, and they're completely meaningless to you. They're as meaningless to you as if you were an earthworm. Because if you can't get any pleasure or satisfaction from something you have no reason ever to do it." (Karp, 1996, p. 98)

"But when you're depressed these things don't do anything for you, they don't, they just, there's nothing, it's just everything's, I don't want to be a cliché and say everything's black, but nothing does… there's no stimulation from anything…"(National Collaborating Centre For Mental Health, 2010, p. 60)

These quotes show a lack of pleasure (when a person is depressed) and how a lack of pleasure corresponds with a negative outlook. This lack of pleasure (which occurs in a depressive state) is usually distinct from when an agent is outside of a depressive state. The DSM-5 uses 'pleasure' and 'interest' interchangeably in the criteria. Note that pleasure and interest are distinct concepts, that although they may at first appear like the same concept, they are quite different concepts, and I claim that both concepts come from different emotions. For example, I can lack interest in playing tennis, this may mean that I have no interest in playing tennis, nor do I have any desire to read about tennis. I may not find pleasure in playing tennis either, but that does not necessarily indicate a lack of interest. Moreover, it could be that I have an interest in playing tennis without finding pleasure in playing tennis. For example, I could be interested in playing tennis for fitness, but I may not take pleasure in playing tennis or vice versa.[45] The loss of pleasure or interest is important in whether one is motivated to do an action. In other words, if I lack all pleasure or interest to

[45] This all depends on how one defines pleasure and interest. Pleasure, for example, does not necessarily mean happy on some accounts. For another account of pleasure (Utilitarianism), see Mill (2014, pp. 8–10).

do ϕ in C it is possible that I will have no motivation to do ϕ in C. For example, if I lack the pleasure or interest to go outside at a particular time, I will not be motivated to go outside at that time.

The next symptoms that I discuss are worthlessness and guilt. Worthlessness or guilt can affect individuals' motivation to perform actions, as they may struggle to see the point in doing daily activities and/or they feel worthless or guilty.

"I assume my family dislikes me and that everyone is looking at me and thinking what a terrible person I am. I begin to feel as if my friends aren't real and secretly they see what I see in myself which is horrible." (Ratcliffe, 2018, p. 132)

"One awful thing about my depression was the tremendous sense of guilt that I was unable to attach to any memory, or action or any part of myself. I was all feeling at that time and no thought—not real thinking, only a slow-motion kind of guilty rumination." (Rowe, 1978, p. 270)

These quotes show that an agent who is depressed may feel guilty or worthless. This is shown by the fact that their self-description is that they are a terrible person, and they may feel guilty that they are a bad person. They feel as though they are worth very little to their loved ones, and thus this may affect motivation. Consider the following example: imagine an agent that believes that their family thinks they are worthless; they may not go and see their family or may avoid them because they feel as though their family does not want them around. This is because of many reasons; for example, they may feel as though they're worthless, or they may feel as though they have done something wrong.

The final feature of depression that might affect motivation is suicidal ideation or recurrent thoughts of death. For example, take these testimonies:

"The world seems pointless because when I am depressed, I can't see the world in a positive way. All I see is a place full of suffering which I often feel I would be better off escaping from… I can't think positively" (Ratcliffe, 2018, p. 129).

“When I’m depressed life never seems worth living… I think that my life will never change and that I will always be depressed. Thinking about the future makes my depression even worse because I can’t bear to think of being depressed my whole life. I forget what my life is like when I’m not depressed and feel that my life and future is pointless.” (Ratcliffe, 2018, p. 129).

“When I am depressed, I am unable to think clearly. I feel sorrow, anger, frustration, sadness, lonely, worthless, despair and mainly I feel like my life is not worth living and I would rather be dead!” (Ratcliffe, 2015, p. 113)

These quotes indicate suicide ideation (suicide ideation refers to when the person has thoughts about committing suicide which the person may or may not act on). Now that I have discussed Major Depressive Disorder (more specifically the symptoms that I have discussed will have an impact on lack of motivation, which will be relevant context for the discussion on my depression counterexample to Weak Motivational Internalism), I will move to discuss how depression poses a counterexample to Weak Motivational Internalism.

5. How are Weak Motivational Internalism and Depression linked?

My opponent’s argument can be laid out in the following way:

1. Weak Motivational Internalism is the claim that if an agent judges that ϕ is right for her to do in circumstances C (at a particular time), then either she is motivated to do ϕ in C (at that time), or she is practically irrational.
2. According to Weak Motivational Internalism if a depressed person judges that ϕ is right to do in circumstances C, but they fail to be motivated to do ϕ in C then they must be practically irrational or else they would be motivated to do ϕ in circumstances C.

However, considering the psychological evidence available, my hypothesis is that the following may occur (but we need to wait until more psychological evidence arises to make a conclusive argument): if somebody, S, judges that it is right to do ϕ

in circumstances, C, but fails to be motivated to do ϕ in C, because she is depressed without being practically irrational then this depression case will result in being a counterexample to Weak Motivational Internalism.

In some of the depression testimony discussed above, some of the testimony may on one reading be indicative of a depressed individual who judges ϕ is right to do in circumstances C but fails to do ϕ in C (at that time) and, at the same time, is rational. For example, imagine that a depressed person judged that it was right to give to charity in the morning, in that same morning, they could fail to be motivated to give to charity, and at the same time be rational. Therefore, I propose that the problem with depression is not that the person cannot judge what is right at a particular time, but rather that they fail to be motivated to do what is right at that time whilst remaining rational.[46]

Now that I have discussed my claim that there is a link between being depressed (or having Major Depressive Disorder) and failing to be motivated, I now precisely lay out why I think my claim appears to be plausible. To do this, I lay out a particular case of depression where a depressed person with certain conditions (specified above) judges that ϕ is right to do in C but, at the same time, fails to be motivated to do ϕ in C and is practically rational.

6. Counterexample to Weak Motivational Internalism: Depression Case

To recap, Weak Motivational Internalism claims that if a person judges that ϕ is right to do in C, then they are necessarily motivated to do ϕ in C unless they are

[46] This is a very important distinction especially compared with other mental health disorders, particularly schizophrenia. This is because I believe in other mental health disorders, that the agent fails to judge that ϕ is right in circumstances C (at that particular time), therefore, is not motivated to do ϕ in C (at the same time). However, I think the problem is that depressed people (with certain conditions specified in the chapter), although can judge what is right to do in certain circumstances (at that time), they may not be motivated to do this action in those circumstances (at that particular time). This means that there appears to be something wrong with the relationship between judging that ϕ is right in C and being motivated to do ϕ in C with a depressed person (with certain conditions) and this may contrast with other mental health disorders such as schizophrenia. For example, in schizophrenia it may well be that agents judge that ϕ is right to do in C and thus are motivated to do ϕ in C.

practically irrational (Smith, 1994, p. 61). However, I argue that there are some cases of depression where one can say that an agent judges that ϕ is right to do in C and they have no motivation to do ϕ in C and they are practically rational in C. I lay out this case (which will now be known as the Hospital Case) in this section. This case seems to have three features (that occur at the same time) which may show a potential counterexample to Weak Motivational Internalism:

> Hospital Case: *Sally has a severe case of depression (a person who suffers with clinical depression). They form a judgement that visiting their loved one in hospital is the right thing to do at a given time (when the loved one is in hospital), but they lack the motivation to do this at a particular time. Yet, by any common-sense standard they form beliefs based on evidence and reach appropriate judgements about which means would promote their ends and are therefore practically rational.*

I now go through each of the features in turn and suggest why one may think that the Hospital Case has these three features.

6a. Sally judges that ϕ is right to do in C

The relevant section in the above example (which is the Hospital Case) is: A person judges that visiting their loved one in the hospital is the right thing to do at a particular time. The Weak Motivational Internalist would say that in the Hospital Case, it looks like Sally judges that it is right to do ϕ in C but fails to be motivated to do ϕ in C, and thus Sally is not making a genuine moral judgement. According to the Weak Motivational Internalist, if Sally were actually making a genuine moral judgment to do ϕ, they would either be motivated to do ϕ in C at that particular time, or they would be practically irrational (at that time). However, I argue that in the Hospital Case, it appears as though Sally is not motivated to do ϕ in C, and Sally also is not practically irrational at that time, but Sally may still be making a genuine judgement that ϕ is right to do in C.

I claim that Sally is making a moral judgement because we need to look at what constitutes a moral judgement, and I argue it is the feelings that coexist with the

judgement, which makes the judgement genuine. It should be noted that I am not simply denying the plausibility of the Motivational Internalist claim that a moral judgement in which an agent judges that it is right to do ϕ in C, and thus is motivated to do ϕ in C, but I am also offering an alternative account of what is involved in the relationship between a judgment and a motivation. Furthermore, I claim that in the relationship between moral judgment and motivation, the feelings of shame and guilt are key.

This is because there is evidence to suggest that in the Hospital Case (and in other similar depression cases which meet the criteria specified in this chapter) the person (Sally) is making a genuine moral judgement. This is because the person feels guilt or shame if they do not feel motivated by, or carry out, this moral judgement. There are cases where a depressed person feels guilt or shame when they make a (moral) judgement, but they do not feel motivated to do the moral judgement like in the case above. More precisely, there are cases where a depressed person feels guilty when they do not feel motivated to do X, and the fact that they feel guilty about not doing X suggests they have a moral judgement that X is right. However, they still have no motivation to do X. The guilt, therefore, is evidence of their moral judgment, but this moral judgment seems to be a moral judgment with no motivation attached to it, just the feeling of guilt. There is much psychological literature that exemplifies that shame/guilt is a large part of depressed individuals failing to be motivated by their judgements in depression (Alexander et al., 1999), (Kim et al., 2011) and (Tilghman-Osbourne et al., 2014). This shows that depressed people are making genuine (moral) judgements because they feel guilty or shameful when they do not feel motivated to carry them out. More precisely, there have been empirical studies done that provide evidence for the association of guilt with depression (Alexander et al., 1999; Ghatavi et al., 2002; Jarrett and Weissenburger, 1990; Walters-Chapman et al., 1995).

These studies show that agents with depression are more likely to feel inappropriate guilt for their actions. This guilt is inappropriate because it is often tied to something that the agent should not feel guilty about or should not blame themselves about. For example, a depressed agent may feel guilty about not being able to get out of bed because of their depression. This is inappropriate as it is due to the mental health

condition that they are not able to get out of bed, but they still feel it. A non-depressed person may well excuse themselves of the fact that they are not able to get out of bed. They may feel guilty about it, but may be able to move past this guilt, whereas a depressed person may find this very hard to get out of bed and thus this would affect their everyday functioning; this is why the guilt is inappropriate for depressed individuals (see Berrios et al., 1992; Kim et al., 2011; Layne, 1980; Peterson et al., 1981). This inappropriate guilt that a depressed person feels looks as if it has an impact on whether they are motivated to do a certain action (ϕ) at a given time when they make a judgement that ϕ is right to do in C at the same time. This is coupled with the fact that there is evidence to suggest that there is an association between depression and guilt as when guilt is controlled or disappears, depression also disappears or is better controlled. ((Fontaine et al., 2001); (Luyten et al., 2002)). Moreover, Orth et al (2006, p. 1608), note that previous studies show the relationship between shame, guilt and depression, therefore if shame is successfully controlled the effects of guilt disappear ((Fontaine et al., 2001); (Tangney et al., 1992)).[47]

These studies show that shame and guilt do play a large part in depression, and this coupled with the DSM (APA, 2013, pp. 160–161) shows how shame and guilt affect motivation. In other words, it appears to be the case that depressed people, in the cases we are discussing, make genuine moral judgements because they have symptoms of shame and guilt. This is a challenge to the Weak Motivational Internalist because the Weak Motivational Internalist would say that the depressed agent (who meets these conditions specified in this chapter above) is not making a moral judgement or is practically irrational, whereas it appears to be the case that the depressed person is making a moral judgement but fails to be motivated because of their feelings of guilt and shame which are associated when they do not feel motivated by the judgement.

[47] It should be noted that there is contradicting evidence that shame and guilt are linked to depression. For example, Orth (2006) said that there is no correlation between shame and depression because there was no effect on depression-related rumination (and shame and rumination are closely linked). They did hypothesise, however, that guilt did have a correlation with depression.

To apply this to the case above, Sally (in the Hospital Case) would feel guilty or shameful when they do not feel motivated to visit their loved one in the hospital at a given time (or when they do not visit their loved one in hospital at that time). To be more precise, Sally judges that visiting their loved one in the hospital at a particular time is the right thing to do, but they fail to be motivated to visit their loved one at that time. This means that they make a genuine moral judgement by visiting their loved one at that time, but they fail to be motivated to do it at that particular time. However, one knows that Sally makes a moral judgement because Sally feels shame or guilt about failing to be motivated to do the action.[48]

In sum, because of the guilt or shame felt by Sally (or Sally beating themselves up), one can claim that Sally is making a genuine moral judgement in cases like the Hospital Case, where Sally judges that ϕ is right to do in C at a particular time.

6b. Sally lacks all Motivation, at a given time, to do what she Judges to be Right at that time because of her Depression

The Weak Motivational Internalist would claim that if Sally is not motivated by her judgment at a particular time, then that judgement at that time, is not a genuine moral judgment or else Sally is practically irrational. Alternatively, the Weak Motivational Internalist could also claim that the person judges that ϕ is right to do in C and so will have some motivation to do ϕ in C (they may even have little motivation to do ϕ in C). More specifically, the Weak Motivational Internalist would claim that it is very hard to detect tiny amounts of motivation, which makes it difficult to be sure, in cases like the Hospital Case, that the agent, actually has no motivation at all to do what she judges to be right.

However, Weak Motivational Internalism does not explain the Hospital Case. This is because in the Hospital Case the person, in fact, judges that ϕ is right and is rational

[48] There are other pieces of philosophical literature that reference the connection between guilt, shame, and moral judgements. See, for example, Gibbard (2002, pp. 136–140), Mason (2003, p. 256) and Joyce (2005b, pp. 102–103). Moreover, there is a definition that psychologists give for the difference between guilt and shame. On a standard taxonomy, shame is global (i.e., directed at the whole person, concerning shortcomings), and guilt is local (i.e., directed at individual actions, concerning transgressions). For further details see Orth et al, (2006, pp. 1609–1610).

but may still lack all motivation to do ϕ in C. The fact that depressed people lack motivation to perform actions is evidenced by much literature, for example, Layne (1980, p. 647) who claims that depressed people suffer motivational defects because of loss of pleasure and erosion of motivation (taken from (Mendels, 1970, p. 7)), or a loss or weakness of will. The fact that agents may have no motivation when judging that ϕ is right to do in C appears to be linked to symptoms of depression, for example, apathy, worthlessness, guilt or fatigue.[49]

Moreover, philosophers should be interested in whether depressed people, as a matter of fact, actually testify to having no motivation at all.[50] Therefore, testimony can play a role in establishing whether a depressed agent is motivated to do ϕ in C when they judge that ϕ is right to do in C.[51] It is more likely that they are not motivated at all which is why they feel guilt or shame, when they have no motivation to do that action (or when they do not actually do the action). Moreover, when talking to some depressed people these depressed individuals claim that they are not motivated to do an action. So therefore, one should take their testimony seriously by suggesting that they have no motivation to do ϕ in C, but at the same time they judge ϕ to be the case in C (and they are rational in C).

In the Hospital Case, the question of whether Sally may have tiny bits of motivation or no motivation at all is hard to settle, therefore there may appear to be a deadlock between the Motivational Internalist and my depression counterexample. The Weak Motivational Internalist is going to claim that in the case of depressed people (who meet the conditions specified above in the chapter), the agent has a tiny bit of motivation, but it is so tiny, that it is hard to detect, or they are practically irrational. For example, in cases like the Hospital Case, it seems that Sally does not have any

[49] It should be noted that it is hard to know whether depressed agents lack motivation or have no motivation when performing an action. I argue that they at least look like they have no motivation and testimony supports this.
[50] I have searched for depression testimony, and it is ambiguous whether depressed individual's testimony meets certain conditions of my proposed counterexample (judging that X is right in C, the agent having no motivation to do X in C and being rational in C), but in principle, real-life depression testimony would be relevant to the philosophical debate.
[51] There are many testimonies of depressed people lacking motivation, see for example Ratcliffe (2018, p. 125) or National Institute For Health And Clinical Excellence (2010, pp. 61, 63) or for further information on depressed persons lacking motivation see Geddes et al. (2012, p. 225).

motivation at all (at a given time) to visit their loved one in the hospital (but at the same time judges that it is right to visit their loved one in hospital and is rational at that time). However, the Weak Motivational Internalist will argue that Sally does in fact have some motivation to visit their loved one (as she is rational and judges that this is the right action to do at that time). However, I claim that Sally has no motivation at that time because it seems that she is not motivated. This is demonstrated by her not being able to act in ways to show that she is motivated, and when we ask her whether she is motivated to visit her loved one she may reply negatively. Whilst I understand the Weak Motivational Internalist position that there might be tiny bits of motivation present in the Sally case, I claim that we need to take Sally's actions and testimony more seriously as she is acting in ways that reflect that she is not motivated at all. This point of whether Sally has a little motivation is hard to solve and more psychological evidence will have to be gathered to determine whether Sally has a little motivation in this case.

As we have seen in the Hospital Case, Sally who is a depressed agent (with certain conditions) may have no motivation at all to perform the action. In the next section (Section 6c), I now turn to the last condition (which is that the agent has to be rational), which the Hospital Case needs to meet, for the case to be a counterexample to Weak Motivational Internalism.

6c. Sally appears to be Rational – there is nothing irrational about her Combination of Attitudes

The relevant part of the Hospital Case is that the agent appears to be rational at a particular time, and at the same time they lack motivation to do ϕ in C, but at that time they also judge that ϕ in C is right. Moreover, Sally may not be irrational in any way as she is not ignoring evidence or possessing contradictory evidence when she fails to be motivated to visit her loved one in the hospital, so the burden of proof lies with the Weak Motivational Internalist because they claim that she is irrational in some way.

There is an account of rationality called means-end rationality which says that if a person thinks that it is necessary to do X, then they will be motivated to pursue the

means to do that end (see Schmidtz, 1994, p. 226 for further details). For example, if an agent is rational and they want to give to charity (the means is that they are giving to charity), they will then be motivated to give to charity to fulfil that end (helping people will be the end). The relevant sense of having X as an end is being motivated to pursue X and, in that sense, Sally does not have visiting as a relevant end in the Hospital Case (even though she judges visiting to be right). This is because Sally is not motivated at all to visit her loved one in the hospital (which is the end in this case) even though she judges visiting her loved one is the right thing to do at this time.[52] However, I claim that one can judge X to be right, without adopting X as an end, and therefore Sally can still be rational in the Hospital Case.[53]

Furthermore, the burden of proof falls on the Weak Motivational Internalist as the Weak Motivational Internalist is yet to describe why people are motivated in this way as when one asks the Weak Motivational Internalist why people are motivated to do ϕ in C, Weak Motivational Internalists would claim that the agent simply had a judgement to do ϕ in C. For example, if an agent judges that keeping a promise is the right thing to do in the morning, then, according to the Weak Motivational Internalists, she will be necessarily motivated to keep that promise in the morning unless she is practically irrational. However, when one asks why the agent is motivated to keep a promise in the morning, the Weak Motivational Internalist would reply that it is because they made a genuine judgement, and this is what Weak Motivational Internalists claim is needed for motivation. Therefore, the Weak Motivational Internalism position in this context is question-begging, and thus requires further explanation on the Weak Motivational Internalist part.

7. Objections to My Depression Counterexample

One objection that can be posed against my argument is this: what is the nature of a moral judgement? This is a question worth considering as one of the key claims in

[52] It should be noted that my interpretation of Sally is only one interpretation based on evidence of what is going on in the Hospital Case. It is important to note that there are other interpretations of what may be happening with Sally. I claim only here that my interpretation is a plausible interpretation that one should consider.

[53] Critics can suggest that she is not motivated and so she has not adopted visiting her loved one in the hospital as an end. However, I claim that she judges that visiting her loved one in the hospital is right and therefore does, in fact, judge it to be an end.

the debate is whether one makes a moral judgement when they fail to be motivated by that judgement (and are not practically irrational). Weak Motivational Internalists might argue that if you are depressed and are not motivated by a moral judgement then the judgement is not a sincere moral judgement (and one may be making the judgement in the inverted comma sense). However, I argue that depressed people are making a moral judgement when they judge that ϕ is right in C even when they are not motivated to do ϕ in C because they endorse the judgement in C and because of the negative feelings associated with their judgement, like guilt and shame.

There is also another key objection that could be posed to my counterexample which is that my account of rationality is wrong. Rationality according to the Weak Motivational Internalist should be based on a reliable connection between a genuine judgement and motivation and any variation on this definition of rationality is not plausible (and thus a means-end rationality is also not viable). Coupled with this, the reliable connection between a judgement and a motivation is intuitive and seems to be how we describe how individuals are motivated in everyday language. However, I claim that Weak Motivational Internalism does not provide enough clarity about how a motivation necessarily follows from a judgement and therefore the burden of proof falls on the Weak Motivational Internalists to provide more explanation on how the reliable connection (between the judgement and motivation) works.

In sum, the Weak Motivational Internalist and the Motivational Externalist reach a deadlock. The Weak Motivational Internalists will claim that Sally in my Hospital Case is not making a moral judgement or has a small amount of motivation to visit their loved one (which is perhaps not easily visible) or is, in fact, irrational. However, the Motivational Externalists will claim that the depression case is a counterexample to the Weak Motivational Internalists because the certain conditions of a depressed person which are presented in the Hospital Case, can exist together. However, my hope is only that the Hospital Case casts doubt on Weak Motivational Internalism and the case helps adjudicate the deadlock between the two theories.

8. Conclusion

In this chapter, I have argued for a modest claim that certain cases of depression (with certain criteria outlined in the chapter) appears to be a counterexample to Weak Motivational Internalism. This is a modest claim as it is difficult to ascertain what the depressed person is doing when they are seemingly making a moral judgement at a particular time but failing to be motivated by their judgement at that time and at the same time look like they are rational and thus there are questions that we need to ask. These questions are: how do we know that the depressed person is making a genuine moral judgement at a particular time when they fail to be motivated at that time? Moreover, how do we know that the depressed person really lacks all motivation to do what they judge to be right at a particular time or how do we know if the depressed person is rational in the state of depression? There needs to be further research done on these questions and further evidence needs to be provided for depression to make clear judgements on whether the cases of depression like the one presented in the chapter, are counterexamples to Weak Motivational Internalism. But for now, with the evidence that we have, I argue that certain cases of depression (with certain criteria specified in the chapter) appears to be a counterexample to Weak Motivational Internalism.

Part 3

Chapter 5: What is the Humean Theory of Motivation, and Why should we take the Theory Seriously?

1. Introduction

In this chapter, I argue that the Humean Theory of Motivation is a theory worth taking seriously in metaethical discourse. I do this by first outlining what the Humean Theory of Motivation amounts to, and the landscape of the Humean Theory. Then I explore in depth different components of the Humean Theory and discuss how they interact in the debate. Next, I consider some objections to the Humean Theory of Motivation and some replies to those objections, to show how my Obsessive compulsive disorder (OCD) counterexample fits into this debate. I, conclude that the Humean Theory is worth taking seriously in metaethical discourse, as the theory has many elements which have merit. My argument claimed in this chapter, will be important as it will provide the landscape required for my next chapter, which will argue that certain cases of OCD (which meet certain conditions specified in the next chapter) may provide a new counterexample to the Humean Theory of Motivation.

2. What is the Humean Theory of Motivation?

A general characterisation of the Humean Theory is that motivation to do X occurs iff there is a relevant belief and an appropriately related desire present. The Humean Theory will say that for every motivation that an agent has, there must be both a relevant belief and an appropriately related desire. If an agent lacks a relevant belief or an appropriately related desire about action X, then they will not be motivated to do action X.

In the Humean Theory, beliefs and desires are modally separable which means that any given desire could (which the word modal informs us of) be separated from another given belief (and vice versa for a belief). This modal separation means that

both beliefs and desires could each exist in the other's absence. The fact that beliefs and desires are modally distinct rules out two things. Firstly, that beliefs and desires can make up a unified mental state comprising of beliefs and desire parts and secondly, that the beliefs and desires are necessarily connected ((Kriegel, 2012, pp. 471–2) (also see Douskos, 2018, pp. 220–1). The claim that beliefs and desires are modally separable is an important qualification, as without this qualification, the Humean Theory is not distinct from Anti-Humean Theories (such as: (Dancy, 1993); (McNaughton, 1988); (Nagel, 1970); (Shafer-Landau, 2003)) which say that a belief can be sufficient to motivate.

A key feature of the Humean Theory is what 'relevant' means. More precisely, the Humean says that motivation to do X exists iff belief Y is a means-end belief to the effect that were the agent to A, he would B (for example, if I were to eat a banana, then I would be less hungry). 'Relevant' in the case of the Humean Theory means that both the belief and the desire have to be appropriately related to each other. For example, I am motivated to eat an apple iff I have a relevant desire and an appropriately related belief that are both linked to eating the apple, such as I have a desire to eat an apple because I am hungry and I believe that if I eat an apple, it will make me less hungry.

There are two main schools of thought of the Humean Theory. The main difference between the two views is that they both differ on what they think desire amounts to. The first position argues for a phenomenological account of desire (famously discussed by Hume (1896, p. 190)) and the second view argues for a direction fit of desire (famously discussed by Smith (1987, p. 45)). I mention this here to lay out the landscape, but I discuss the different accounts of desire later in the chapter (Section 3).

The Humean Theory has a key aspect which is vital to the theory which is intentionality. In other words, the Humean would claim that every action which an agent is motivated to do needs to be intentional. 'Intentional', for my purposes, will mean something that the agent has a purpose to do. This is different to an involuntary response where the agent has no control over their actions or emotions. For example, if I put my hand on top of a burning flame, my hand will instantly pull

away (without me even thinking about doing this action) and this will be an involuntary response. Therefore, I acknowledge that the Humean is describing a motivation which is intentional. The term 'motivation' will be used in this framework.

Michael Smith, whose Humean Theory I focus on, identifies two different claims of the Humean Theory: a weaker and a stronger. The stronger Humean claim says that motivation has a *source* in the presence of a relevant desire and means-end belief. More precisely, the stronger Humean Theory can be expressed in the following way: "R at t constitutes a motivating reason of agent A to ϕ iff there is some ψ such that R at t consists of a desire of A to ψ and a belief that were he to ϕ he would ψ" (Smith, 1987, p. 36). However, in contrast to the strong Humean Theory, is the weaker claim of the Humean Theory of Motivation which requires the *presence* of a relevant desire and a means-end belief. More precisely, the weaker claim of the Humean Theory can be spelt out in the following way: "Agent A at t has a motivating reason to ϕ only if there is some ψ such that, at t, A desires to ψ and believes that were he to ϕ he would ψ" (Smith, 1987, p. 36). Moreover, some Anti-Humeans (McDowell and McFetridge, 1978, p. 15; Nagel, 1970, p. 29; Smith, 1987, p. 37) find that the weaker Humean claim is acceptable because they think that although relevant beliefs and appropriate means-end desires are always present in motivation, they are not themselves the source of motivation. However, other Anti-Humeans, such as Platts (1979, p. 256), think that both the weaker and the stronger Humean claims are not acceptable (Smith, 1987, p. 38).[54] In this paper, I focus on Smith's weaker claim of the Humean Theory of Motivation as my OCD case will be primarily aimed at this claim.[55]

Next, I make a point about the kind of reason that the Humean is discussing. For the Humean Theory to be understood, one needs to distinguish between normative and motivating reasons as the Humean Theory is specifically referring to motivating

[54] However, Smith himself believes the stronger and weaker claims about motivation (that the Humean Theory and other non-Humean theories may accept), which he presents in his paper, are both plausible claims. See Smith (1987, p. 37)

[55] However, the OCD counterexample can also be plausible for the stronger Humean claim about motivation, but I will focus on the weaker claim in my paper as the weaker claim of the theory provides a better defence against my OCD case (which I present in the next chapter).

reasons according to Smith. More precisely, Smith claims that "the distinctive feature of a motivating reason to X is that in virtue of having such a reason an agent is in a state that is *potentially explanatory* of his X-ing" (Smith, 1987, p. 38). A motivating reason is a reason that is 'psychologically real'; this means that this type of reason has the potential to explain the behaviour of that agent. In contrast, a normative reason has a 'justifying role' and thus justifies why an agent should do an action. There is a relationship between a normative and a motivational reason, which Smith notes is prominent, because if a reason plays a normative role to an agent, then they are more likely to be motivated to do the action, and therefore this means that they may be more likely to have a motivating reason. This is because the agent believes that they *should* do the action and therefore they have a *justification* for doing the action, which leads to a motivational reason for them doing the action. Imagine I decide to make gin and tonic and believe that I can do so with mixing ingredients which are in front of me. Suppose that I am mistaken and think that one of the ingredients in front of me is gin, but really, unbeknown to me, it is petrol. From an outsider perspective, I have no reason to mix these ingredients with tonic (as it was not gin). But from my perspective, I believe the ingredient was gin and so I am motivated to do the act of mixing. Crucially, I have the relevant belief that I am mixing the appropriate ingredients for a gin and tonic, and I have the appropriate desire for a gin and tonic. Therefore, I was motivated to do it. However, what this example brings out is that, even though I have a motivating reason to do it, I do not have a normative reason to do it. This is because from an outsider perspective I have no reason to mix the petrol substance with tonic as this is not going to produce my desired gin and tonic ((Williams, 1979, p. 60) (also see Smith, 1987, pp. 37–41))

However, there are many contrasting theories to the Humean Theory. For example, Shafer-Landau has a different theory of motivation to the Humean, defining two ways that only a belief can motivate (Shafer-Landau, 2003). The first is the thought that some evaluative beliefs are sufficient by themselves to motivate and that we do not need desire to motivate (this is called the Kantian interpretation of Anti-Humeanism) (Shafer-Landau, 2003, p. 122; for further consideration see Dancy, 1993, p. 3). For example, if I believe that stealing is the wrong thing to do, then the belief that stealing is wrong is enough to motivate me not to steal.

The second is that beliefs by themselves "can sometimes generate […] motivated or derived desires", which put together with evaluative beliefs are enough to motivate ((Shafer-Landau, 2003, p. 122) (McNaughton, 1988, p. 21)). For example, if I hold the belief that I should not kill, this may lead to a desire not to kill. Notice the word 'lead': this view is not just that an evaluative belief can motivate, but rather that an evaluative belief leads to a desire, and the belief and desire together motivates. Using the example of not killing, the evaluative belief that I should not kill, which results in having a desire not to kill, and this, coupled with the existing belief that by doing X I will not kill, results in me being motivated to do X, and thus the belief and the desire together is sufficient to motivate me to do X. This correlates with Smith (1987), (1994) where he argues that beliefs with the 'right content', give rise to desires, which then both lead to motivation. Moreover, Shafer-Landau says that this idea of motivation is supported by the way that we speak; for example, when I say that "I saw that it needed to be done, and [so I] did it." In other cases, however, a desire *is* present, but it seems to be going against the evaluative belief (evaluative beliefs are beliefs where the individual makes a claim about a moral action, for example 'I believe that this action is right'); one is motivated to do what is right despite facing strong temptations (i.e. desires) to not do the right thing (Shafer-Landau, 2003, p. 123). This is because most individuals would find it plausible to think that claims such as "that simply must be done" or "that is the right way to do [this action]" would be enough to motivate an individual by themselves (Shafer-Landau, 2003, p. 123).

Now that I have considered the Humean Theory and Anti-Humean debate, I will consider the first of the two vital components of the Humean Theory which is belief.

3. What is a Belief?

A belief is a crucial concept in my chapter; not only is it one of the key conceptual components of the Humean Theory, but in the next chapter, I show how a belief is not present in cases where a patient has an OCD compulsion. I argue that we should take a liberal definition of a belief, as when looking at mental health perspectives, one needs to account for the human psyche being varied. Therefore, in this section, I give an account of what a belief is, using David John Owens' conception.

A general characterisation of a belief is a description of the world which seems to be inert; a belief is about what the world is like but does not attempt to change the world in any way (Fisher, 2011, p. 134). Owens claims that there are two key features of a belief. The first is that beliefs can have the ability of being "right" or "wrong", "correct" or "incorrect". The fact that beliefs have these abilities means that they form a normative assessment of the standard of correctness, which governs a belief (Owens, 2003, p. 283). The standard of correctness is that a belief has a goal which it strives towards, or a function, which a belief aims to do (Owens, 2003, p. 283). The concepts of right or wrong (or correct or incorrect) are not the same concepts as true or false. This is because a statement may be correct and/or right to the believer, without having to actually be true or false externally to the believer. This can be demonstrated in the following example: imagine a case in which a person has a delusional belief that aliens exist. This may be right or correct to the believer (as the agent thinks that it is justified given their other beliefs) but could also be false (assuming that aliens do not exist). This means that a belief is endorsed by the believer and is trying to aim at what they perceive as the right description of the world around them. Thus, it is plausible to say that a believer could have a belief that is incorrect (to the external world) but also justified to them or vice versa.

The second feature of a belief is that it is required to have the capacity to be "rational" or "irrational", and "justified" or "unjustified". The two pairs of "rational" and "irrational", and "justified" and "unjustified" can both occur independently of one another, or they can occur together. A believer could have a justified belief which to them is rational or vice versa. The difference is that if a belief is justified, on the whole, there should be some kind of substantial reason for believing the belief (this reason does not necessarily need to be based on evidence but could be based on a feeling or another piece of information accessible to that person at that time). However, if a belief is rational, there is some substantial evidence for believing the belief, and therefore any person who has access to that evidence and is in that mental state would believe that belief. For example, if I believe that monsters exist, then this is an irrational belief, but may be justified because I have a mental illness which causes me to hallucinate about monsters. These two features are vital in understanding precisely what a belief is, and showing what the differences are between how a belief is portrayed to an individual and how a belief represents the

world that it is describing. Moreover, one could also have an irrational belief that is justified. For example, if I were to say (before COVID became apparent), that a serious illness would affect humans globally, then this seems like it was a justified but irrational belief. This is because, assuming that there was no evidence at that time that a global pandemic was on the horizon, this belief had no justification as there was no evidence to support the belief, making it an irrational belief. However, in 2019 we did in fact have a global pandemic so arguably then, the belief was justified. There is another distinction, which is important, which is an agent can have substantial reason without having substantial evidence. For example, in the monster case, the agent can have substantial evidence that monsters exist i.e. they could hallucinate about monsters (their hallucination would feel real to the person who was hallucinating), without having substantial reasons to believe in monsters (the agent might not have substantial reasons to believe that monsters exist).

Now that I have considered what a belief is, I now consider the second key component of the Humean Theory which is desire.

4. What is a Desire?

The second component of the Humean Theory is desire, so this is where I turn to next. I will first describe Hume's phenomenological account of desire. Desire, according to Hume, is "external and internal, passions, affections, sensations, pains and pleasures" appear as impressions or perceptions (Hume, 1896, p. 190).[56] Hume is claiming that desires are a distinct feeling and that desires are prompted by bodily sensations (Hume, 1896, p. 190).[57] For example, if I desire to eat an apple, I may be prompted by a bodily sensation as I am hungry, and only then will I be motivated to eat the apple. Moreover, they have motivational forces that can encourage an agent to perform an action (for further discussion about desires and motivational force see Dancy, 1993, p. 2). Hume is claiming that since all sensations should be known through consciousness, they must appear in the idea that we're discussing.

[56] I understand that Hume is not a modern Humean, but he still gives a sound conception of the phenomenology of desires.
[57] One should note that not all desires are in the form of bodily sensations. Hume would also allow for a desire to be based on a pleasant taste, for example, the pleasant taste of an apple.

Moreover, everything that appears in the human psyche should be felt by the individual. These sensations are independent and eternal of us and they come from immediate sensation or other causes.

Furthermore, I think that the phenomenological account marries up with psychological schools of thought. I think that when we describe ourselves as desiring something, we often describe something that we want, need, or desperately long for. For example, the statement that 'I want to see my grandparents because I love them very much' can be re-described as a desire to see my grandparents. This shows that individuals' feelings (such as love, anger, sadness, happiness) can cause one to desire something. I argue that this account of desire is plausible because it describes how an individual feels, which is particularly important when incorporating mental health sufferers' perspectives such as anxiety, which focus on the individual's feelings rather than looking at mental states from an outside perspective. Furthermore, a mental state of desire should be understood in terms of feelings rather than how an agent interacts with the world (which is the alternative account of desire (and belief)). This account, which is called the direction of fit account, will be discussed in the next section.

However, there are some defenders of the Humean Theory, such as Smith, who deny that all desires are feelings.[58] These defenders argue that if we take that desires should be distinct feelings then the strong phenomenological conception of desires should be rejected because the phenomenological conception of desires cannot be married up with a plausible account of epistemology of desires as the phenomenological conception of desires is inconsistent with the epistemology of desires (Smith, 1987, p. 45).[59] This is because a strong phenomenological conception of desires means that an agent has unconscious desires that they are unaware of. For example, imagine an agent that goes past the same newsagent every day (even though it is out of their way and it takes longer for them to get to) because

[58] Additionally, Smith claims that a desire does not necessarily need to be felt to be present in an individual.
[59] This is a brief account of why the phenomenological account should be rejected. It should be noted that critics of the phenomenological account of desire claim that the phenomenological account should be rejected, as it is unclear what the propositional content of desires is. See Smith (1987, p. 49) for further details.

unknowingly they want to look at themselves in a mirror (Smith, 1987, p. 46). In other words, they have an unconscious desire to look at themselves in the mirror but may not know about this desire. However, Smith claims that to have a plausible theory of desires, then the agent must know that they have a desire, and the strong phenomenological account does not allow for this (because the strong phenomenological account claims we can have unconscious desires) (Smith, 1987, p. 47). Furthermore, if desires do have a phenomenological aspect, then the epistemology of that phenomenological aspect can be based on the epistemology of sensations, and this is why we should reject the strong phenomenological account. This is because the phenomenological account cannot explain that desires have propositional content and thus does not explain the epistemology of sensations. Moreover, Smith also claims the phenomenological account fails to say what the propositional content of desire amounts to (Smith, 1987, p. 50).

Therefore, Smith turns to an alternative account to distinguish between beliefs and desires, which is the direction of fit account, as, according to Smith, it provides the Humean with a clearer distinction between beliefs and desires, this is where I turn to next (Section 4). It should be noted that I do not endorse this view of direction of fit account, I merely report it in the next section. I mention it, however, because it is an important concept to understand when discussing the landscape of belief and desires.

However, different Anti-Humeans propose different accounts of how desire motivates an agent which goes against the Humean Theory of Motivation. This is because the Anti-Humean can define desires differently (using neither the phenomenological account of desire or the direction of fit account) and thus they might object to the Humean Theory. For example, Nagel (1970, p. 29) says that there are two types of desires: motivated and unmotivated. Nagel uses the example of hunger because of a lack of food as an example of an unmotivated desire. This is still a desire, but it does not have the rationality of a motivated desire, but it is governed by impulsivity. An unmotivated desire in this case might be to go to the shops and buy some food, as an involuntary response to hunger. However, an example of a motivated desire could be the desire pick up a newspaper as I want to see what the latest news articles are. According to Nagel (1970, p. 29), an unmotivated desire is instinctual, which means that the unmotivated desire is based

on bodily or mental needs or addictions that one has not consciously decided on, but one merely acts to satisfy. A motivated desire is a desire which one consciously acts on, and the action will satisfy the desire. Furthermore, Dancy (1993, p. 29) argues that many desires are motivated and thus are reached by deliberations. This means that they do not need to immediately compel an agent to perform an action, and some examples of this are emotions or appetites.

5. What is the Direction of Fit Account of Beliefs and Desires?

The direction of fit account is used by Humeans to specify how beliefs and desires are distinct mental states and are separate from each other. A general characterisation of the direction of fit metaphor is that we change our beliefs to fit our world and we change our world to fit our desires (1979, p. 257) . There are two schools of thought of how one should understand the direction of fit account; I now elaborate on each of the two schools of thought, before suggesting that Smith's dispositional account of the direction of fit is the preferable interpretation out of normative and dispositional, as it allows us to explain precisely what a desire is, and how propositional content can be used to explain what a desire amounts to. Moreover, as propositional content is a key concept to understanding what a desire is, and therefore the direction of fit's metaphor, I will discuss this concept later in this section.

There are two schools of thought of direction of fit account. The first is a normative account of the direction of fit. The most plausible account is given by Platts (who uses Anscombe's work *Intention* (1957, pp. 56–57) to explain this direction of fit account) so I lay it out in full here:

> "Beliefs aim at the true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit with the world, not vice versa. Desires aim at realisation, and their realisation is the world fitting with them; the fact that the indicative content of a desire is not realised in the world is not yet a failing *in the desire*, and not yet any reason to discard the desire; the world, crudely, should be changed to fit with our desires, not vice versa." (1979, pp. 256–257)

Platts endorses the metaphor that Anscombe uses in her work; Anscombe (1957, pp. 56–57) discusses a man going round with a shopping list in his hand. A detective is assigned to record the man's purchases in a notepad. This example is meant to illustrate that even if the man and the detective are writing the same list of purchases, both lists will have different relations to the world. The man's shopping list determines what he puts in his basket, whereas the detective's list is based on what the man puts into his basket. Beliefs are similar to the detective's list, as beliefs are supposed to represent things as they are (Swartzer, 2013, p. 976). However, desires are states where the world does not need to fit the desire, and it is not a failing of a desire when the world does not fit it. As the man is changing the world to fit his list, as he is putting items in his shopping basket that he wants to buy, this is a 'world to fit' scenario, so this symbolises a desire.

Moreover, the normative account of the direction of fit can be precisely laid out by the following statement: "the belief that *p ought* to be held only if *p*, whereas the desire that *p* can properly be held even if it is not the case" (Sobel and Copp, 2001, p. 52). In the normative approach one believes that we *ought* to change a belief with content *p*, given that *not p.* For example, it is said that, "if *not p* then it follows that the belief that *p* ought to be changed, but it does not follow that the desire that *p ought* to be changed" (Zangwill, 1998, p. 175). This means that beliefs ought to be revised if they are false, whereas desires need not to be revised if they are not satisfied. Furthermore, Platts (1979, pp. 256–257) notes that the direction of fit account is highly metaphorical which is a problem for Smith when characterising desires.

Smith claims that the normative account is highly metaphorical and does not characterise desires in an adequate way, because Smith worries that in the normative account all desires seem to involve elements of belief (Smith, 1987, p. 51). This means that Platts notes that if we take the characterisations of normative account of belief and desires strictly, then the metaphor is unclear with whether it allows for the sole characterisation of desires as an independent (from belief) mental state (Smith, 1987, p. 51). Therefore, I agree with him that the dispositional conception is better for describing the direction of fit account, so I now turn to describing the dispositional account and why I agree with Smith in saying that this account is better

than the normative conception of direction of fit account, in order to provide a full picture of the debate going on within Humean Theory literature. The dispositional account of the direction of fit is the account that belief and desires *actually do play* different functional roles (Sobel and Copp, 2001, p. 45). (The definition of functional roles will be discussed later in this section).

A general characterisation of the dispositional conception of the direction of fit account is given by Smith, who says that it allows us to see how a desire works through a state in which the *world must fit* (Smith, 1987, p. 54). This is because the difference between belief and desire on the direction of fit account is that "a belief that *p* is a state that *tends* to go out of existence in the presence of a perception that *not p*, whereas a desire that *p* is a state that *tends* to endure, disposing the subject in that state to bring it about that *p*" (Smith, 1987, p. 54).

The dispositional account originates from a problem that Hume faces in his account of desires. There are violent and calm passions according to Hume. Hume's account of the epistemology of desires was that desires can be known by their phenomenology; Smith argued that this concept of desire was inadequate to explain calm passions, because calm passions lack phenomenological content, and therefore Hume needed an alternative account for the epistemology of calm passions (Smith, 1987, pp. 51–52).[60] Hume argues that calm passions are known by their effects rather than their immediate sensation (this is in contrast to the violent passions which can be explained by their phenomenological content) (Smith, 1987, p. 51). According to Stroud (1981, p. 57; see also Smith, 1987, p. 52) desires can be thought of as causes of actions, which calm passions do not incorporate. Therefore, Hume's suggestion is redundant and thus we need a causal conception to be in place (Smith argues there needs to be a reason explanation as a species of teleological explanation, and if we accept this explanation, we lock ourselves into a causal conception (for further discussion see Smith, 1987). For us to accept this concept of desire, therefore, one has to find a different conception, a conception that allows one to remain neutral on whether the desires are causes (Smith, 1987, p. 52). According

[60] It should be noted that, according to Smith, even if one adopted a weaker phenomenological account, it is still problematic, as any phenomenological account would not be able to describe the propositional content of desire. For further discussion see Smith (1987, p. 48).

to this conception, desires are states that have *certain functional roles* (Smith, 1987, p. 52). More precisely, "we should think of the desire to ϕ as that state of a subject that grounds all sorts of his dispositions: like the disposition to ϕ in conditions C, the disposition to ϕ in conditions C', and so on (where, in order for conditions C and C' to obtain, the subject must have, *inter alia*, certain beliefs)" (Smith, 1987, p. 52).

I advocate for the dispositional account of the direction of fit as it has many advantages over the normative account. As I argue that Smith gives the most plausible account of why the dispositional account is better, I lay the advantages out here in brief, before discussing some arguments in favour of the Humean Theory. The main advantage (and reason why I advocate for this conception over the normative account) of the dispositional conception, is that it allows us to explain the cause of the desire adequately. To explain this further, I now need to explain clearly what the difference is between the normative account and disposition account is. The dispositional account would say broadly that the perception that P tend to *cause* the belief that P, therefore the connection between the perception that P and the belief that P is causal. However, in contrast, the normative account says that the perception of P ought to bring about the belief that P (and sometimes things are not as they ought). The dispositional account lays out how propositional content can explain what desire is, as propositional content can be determined by its functional role (Smith, 1987, p. 52). Another advantage of the dispositional account of the direction of fit is that Hume's suggestion of calm passions (discussed earlier) can be translated into the epistemology of desires as dispositional states. The dispositional conception therefore does not commit us to the claim that desires have to be conceived as causes of action, which means we can accept Hume's conception of calm passions (Smith, 1987, p. 52). This is because these calm passions come in a sense that they do not have a strong phenomenology, but these calm passions are still causes.

Because of the reasons stated above, I argue that a dispositional account is more precise than a normative account about how one should deal with the propositional content of desires, and therefore gives us an insight as to what desires are. Now that I have discussed the key components of the Humean Theory, I will now move to consider why one should find the Humean Theory plausible, before turning to some objections and replies to the Theory, which are relevant to my OCD case.

6. Why Find the Humean Theory of Motivation Plausible?

The way we verbalise how we are motivated concurs with the explanation of motivation that the Humean offers. This is because when we express claims about motivation the expression of motivation can be understood in terms of a relevant belief and an appropriately related desire. For example, if I am motivated to eat an apple because I am hungry (although there could be other reasons why one may be motivated to eat an apple), then it is plausible that I express the relevant belief and appropriately related desire by saying, 'I believe that eating an apple will make me less hungry' and 'I desire to be less hungry'. Similarly, if I have a motivation to ride a bike to get me to university, it is highly feasible that I would state that I have a belief that riding a bike will get me to university and I have an appropriately related desire which is a desire to get to university. The mental states of beliefs and desires therefore fit with how we express motivational utterances. The Humean Theory is also plausible as it has explanatory power – the Humean Theory clearly explains what it needs to explain and thus, this makes the theory intuitive, simple and easy to understand. The Humean Theory seems to provide a plausible explanation in everyday cases of how an agent is motivated. Moreover, it is also easy to generalise across all cases of motivation as the key components within the Humean Theory can be applied to all instances of motivation. Van Roojen (1995, p. 37) also notes that the Humean Theory is attractive because a desire is required for intentional action: one needs to want to do an action to be motivated to do the action and this, coupled with the belief, explains the Humean Theory of Motivation. Additionally, Smith (1987, p. 44) claims that there's a key reason to support the Humean Theory which is that the Humean Theory supports a pursuit of a goal. This means that the Humean Theory can explain why people are motivated in the way that they are, and this is shown by the apple example above where the person has a goal not to be hungry and therefore they eat the apple to fulfil that hunger.

McDowell (2005, p. 155), who himself rejects the Humean Theory of Motivation, gives a reason why the Humean Theory is plausible. He says that the Humean Theory explains motivation in terms of a cognitive state (a belief), which is then added to a non-cognitive state, which is a desire. Moreover, it is important to note that desires, according to McDowell, possess causal force whereas beliefs do not

(Smith, 1987, p. 44).[61] However, I claim that beliefs do not give a causal force, as beliefs are merely representations of the world (McDowell does not give an account of why beliefs do not provide a causal force). Instead, beliefs describe what the agent thinks is happening in the world. Consider an example that I believe that aliens are real. I may believe that statement, but I may only be motivated to do an action (e.g., finding an alien spaceship) if I desire to find out whether aliens exist. McDowell adequately shows that the Humean Theory can explain the relationship between a belief (a cognitive state) and a desire (a non-cognitive state) which is a reason, I think, that one should find the theory plausible.[62]

Lewis (1988) argues that the Humean Theory has a similar advantage to those put forth by McDowell in the previous paragraph. He uses Decision Theory to support the Humean Theory. Decision Theory is a well-worked out formula of the theory of belief, which indicates to us what it means to serve our desires according to our beliefs, which therefore tells us how the Humean Theory works (Lewis, 1988, p. 325). More precisely, Lewis says that the Humean Theory about motivation says that we are moved entirely by desire, we will fulfil our desires according to what we believe will fulfil our desires (1988, p. 325). He says this because without desires in the process of motivation, there would never be a mental state to move us, which would then cause us to be unmotivated. This may be another reason why we find the Humean Theory plausible as it gives us a well-supported account of how the functions of belief and desire work.

Now that I have considered why the Humean Theory is plausible, I now turn to considering some key objections and replies which will illustrate the landscape for my OCD objection, which will be made in chapter 6.

[61] It should be noted that Smith disagrees with McDowell because Smith claims that the Humean Theory can endorse a causal conception of reason. See Smith (1987, p. 43).
[62] Coupled with this, Smith (1987, p. 43) uses McDowell's abovementioned reasons to suggest why the Humean Theory is worth taking seriously.

7. Objections to the Humean Theory of Motivation

The main objections to the Humean Theory that I will be discussing find ways to criticise the Humean Theory by suggesting ways that an agent can be motivated without a relevant belief or an appropriately related desire present. This is because a key claim is that motivation occurs iff a relevant belief and an appropriately related desire are present. As this is a key claim on which the Humean Theory is based, if a motivation happens without the distinct mental state of either a belief or a desire, then the Humean Theory is false. Therefore, I first discuss two ways (an *alief,* and a *besire*), in which an agent could be motivated without the presence of a relevant belief or a desire. It should also be noted that this links to other Anti-Humean Theories, such as Dancy (1993), Nagel and Shafer-Landau (1993; 1970; 2003), who claim that individuals can be motivated without either a relevant belief or an appropriately related desire, and this can be linked to my OCD case to the Humean Theory, as I argue that when an agent is motivated to perform an OCD compulsion, they are lacking a relevant belief (but have the relevant desire, which is relevant to the compulsion).

The first mental state which, it has been proposed, can motivate an agent to perform an action, without the presence of a belief or desire, is an *alief*. An *alief* is a distinct mental state which motivates an agent to perform an action (and this mental state is distinct from a belief and a desire) (Gendler, 2008, p. 642). A general characterisation of an *alief* is when one is motivated to do something that does not align with their beliefs. For example, if I was standing on top of an open-top skyscraper and there was a safety barrier around the top of it (meaning that there was no way that I could fall), I may believe that I am safe from falling. However, I may *alieve* that there was danger, and I may fall. A key characteristic of this *alief* is that it motivates me to step away from the barrier and away from the edge of the skyscraper even though I believe that I am safe. In an *alief* there is a belief-behaviour mismatch where an agent's belief is different from what they are motivated to do. Take for example, the skyscraper: I believe that I am safe because there are railings stopping me from falling, but I am still motivated to step back anyway. This means that my behaviour and my belief are not aligned. However, a Humean would respond to a case of an *alief* by claiming that the agent still has a

relevant belief and an appropriately related desire to do the action, it is just a different belief and appropriately related desire than the Anti-Humean thinks that the agent has. For example, in the skyscraper case, the Humean may re-describe this case by saying the agent has a slight belief that they may fall and a desire not to fall. In other words, the agent still has a relevant belief (that they will fall) and an appropriately related desire not to fall and therefore cases like these can still be described by the Humean Theory. However, a further question is whether an *alief* is a 'unified' mental state combining a belief and a desire. If it is, then an *alief* may still pose a counterexample to the Humean Theory of Motivation.

The second apparent counterexample is a *besire* (Zangwill, 2008), which is a unified mental state which is made up of a belief and a desire which can motivate an agent (Park, 2013, p. 5). This mental state displays the combined features of both a belief and a desire (Park, 2013). This is different to the Humean Theory as according to the Humean Theory the relevant beliefs and appropriately related desires that make up motivation are modally distinct mental states, which means that they can be separated. However, the mental state of a *besire* is a mental state in which belief and desire cannot be separated, when motivating an agent.

A *besire*, therefore, is an objection to the Humean Theory as a *besire* can motivate an agent. This is because the Humean Theory states that a motivation occurs iff a belief and an appropriately related desire are present, and a *besire* is one mental state *combining* the characteristics of a belief and desire. An example of a *besire* is hope, as it has two directions of fit. When an agent hopes for something, they have a belief that they may get what they hope for and a desire that they will get what they hope for, but these cannot be separated, thus forming a unitary mental state (Fileva, 2021, p. 1975).[63] Furthermore, hope is a *besire* as it has the features of a belief and a desire, but it is not reducible to a belief and a desire (Fileva, 2021, p. 1978). This means that hope cannot be broken down into the mental state of beliefs and desires, and thus needs to be thought of as one mental state. This can be illustrated in Fileva's example: imagine there are two people who received lottery tickets on their respective birthdays; they both desire to win (therefore we can say that they are

[63] It should be noted here that hope as a besire is a controversial example, but I claim that it shows the characteristics of a besire.

motivated to buy the lottery tickets). Person one (Alex in Fileva's paper) however hopes to win. Person two (Vera) does not hope to win. Alex and Vera both have the same beliefs e.g. 'I may win, but winning is unlikely' and the same desire ('I would like to win') but one of them hopes that they will win, showing that hope is a distinct mental state which is *besire* (a mental state which has the characteristics of beliefs and desires) (Fileva, 2021, p. 1978). A Humean would respond to this example by simply saying that the Anti-Humean is mistaken by the notion of *besires*, and that hope can be separated into two states: belief and desire. Therefore, beliefs and desires are two distinct mental states which together can result in feelings of hope in the right circumstances.

Crucially, it is cases like the *besire* and the *alief* which result in what appears to be a deadlock between the Humean and the Anti-Humean. The Humean will say that cases of *besires* and *aliefs* can, in fact, be separated into two distinct mental states whereas the Anti-Humean would say that we cannot separate the two mental states involved (depending on whether one is talking about an *alief* or a *besire*). It is therefore hard to adjudicate the debate with the Humean and Anti-Humean on this matter and thus, this is where the debate reaches a deadlock. However, I hope that my OCD case presented in the next chapter will help settle this debate as I will argue that there is definitely no belief in certain cases of OCD (which meets the conditions that I specify in the next chapter), which will show that the Anti-Humean is the more plausible theory in metaethical discourse. I now move to other objections (and give some replies on behalf of the Humean) raised by Anti-Humeans in order to further show that the Humean Theory can defend itself against these objections and is thus a theory worth taking seriously. In this section, first, I give a brief overview of OCD, before moving to explain my objection.

Thomas Nagel (1970, p. 29) has a prominent objection, which is based on prudential motivation, to explain why the Humean Theory is mistaken (which Smith is responding to in his 1987 paper). Nagel says that one can desire something now that they do not expect to desire in the future, and thus there will be no reason now to bring this desire to fruition (1970, p. 39). Furthermore, if I expect to be motivated by a desire in the future, then it is true that I will have a necessary reason to do what satisfies the desire in the future (1970, p. 39). But the future reason that I have does

not mean that I actually satisfy any desires now, and therefore because I am not satisfying any desire now, I have no reason to be motivated now by this desire.

Smith criticises Nagel by saying there is no explanation for why an agent's recognition that he will do ϕ in the future gives rise to a present desire that he will do ϕ now. Moreover, Smith responds to Nagel by saying that the agent has a general present desire to promote future ends (Smith, 1987, p. 41). In other words, the Humean claims that the agent has a motivating reason now to promote ϕ-ing in the future (Smith, 1987, p. 42). Coupled with this, Smith worries that Nagel is confusing motivating reasons with normative reasons as motivating reasons can be based on the individual perspective of the agent, and thus can be based on future ends but normative reasons cannot. Moreover, even if Nagel presents a plausible view by claiming that it is irrational to promote a future desire, then the Humean Theory can accept cases of prudential motivation by claiming that their theory of rationality, in fact, requires the agent to have desires to promote their interest in the future (Smith, 1987, p. 42).

Another worry for the Humean Theory is presented by Searle when he claims that the Humean Theory is unable to explain cases of akrasia (see Searle, 2001). Searle (2001) worries that the Humean Theory will not be able to explain how an agent can be akratic (akrasia is when an agent suffers from weakness of will and thus acts against their better judgement) Sinhababu, 2009, p. 496).[64] Furthermore, Sinhababu (who actually defends the Humean Theory in his paper) claims that Searle is saying that there are two different gaps within the process of motivation. The first gap is forming an intention and the second gap is determining whether to perform that action with that intention (Sinhababu, 2009, p. 497). The Humean Theory fails to explain the second gap, i.e. what is wrong when an agent has an intention but fails to act on it. In other words, there seems to be a failure on the Humean part to say why an agent can form a belief and a desire and fail to be motivated to do the action (Sinhababu, 2009, p. 497). Moreover, Searle notes that the problem with akrasia is not that the agent fails to act on their intentions because there are other alternatives

[64] It should be noted that akrasia has different definitions. Some say that akrasia should be characterised as weakness of will (see Davidson, 1969, p. 93), whereas others say that weakness of the will is different to akrasia (see Holton, 2009, p. 73).

available to the agent but rather that there is something abnormal about the psychological processes that are involved in this failure (Searle, 2001, pp. 233–34). However, Sinhababu disagrees (and I find Sinhababu's point plausible) by suggesting that the Humean Theory can in fact account for akrasia by saying that Hume says there are two types of desires in the Humean Theory, violent and calm (2009, p. 500). However, the Humean could suggest that the akratic actions merely have more violent desires in these cases, which means that the agent is in fact motivated to do the akratic actions instead of the non-akratic actions (Sinhababu, 2009, p. 500). Sinhababu also argues that the Humean Theory is simpler than the Anti-Humean Theory because it does not need to bring any other mental state or any alternate desire to explain akratic action, it can just draw on the desires that the agent already has and recognises (2009, p. 500).

Moreover, Scanlon notes that there seem to be cases where individuals seem to be motivated without having any desire to do the action (Scanlon, 2000, p. 39). Scanlon notes that we often talk about desires in terms of pleasant experiences (or the avoidance of unpleasant experiences), but Scanlon argues that these are not the only things that can make up desires (Scanlon, 2000, p. 39). Furthermore, there seems to be cases where the agent is motivated to do an action that they do not want to do, for example, telling a friend that a piece of clothing is unflattering (Scanlon, 2000, p. 39). Here, the agent does not want to tell the friend that they look unappealing, but they feel obliged to do the action. This looks like, at least, that there are cases in which agents are motivated to do undesired actions. However, Sinhababu disagrees with Scanlon by saying that in the cases which Scanlon is referring to, there is in fact, a desire at play but it is, in fact, a negative desire (Sinhababu, 2009, p. 489). For example, in the case of the unflattering clothes, the friend actually desires to tell their friend that they do not look appealing in those particular clothes; in order to save them emotional embarrassment and thus, the agent does in fact have a desire which is motivating them.

However, Darwall (1983, p. 35) raises another objection against the Humean Theory which says that a new desire is attributed after the deliberation has been decided for the motivation, and the previous desire is not necessary in the deliberation process. For example, if I have a pre-existing desire to help an agent in a certain situation and

then if it turned out I had a more specific desire of helping a person to cross the road, who needed it at a particular time (at the same time as my pre-existing desire to help people), then I would be motivated to help that person to cross that road at that time. Moreover, this may explain why some motivations that individuals have can be explained seemingly without desires as the agents act on pre-existing desires (Sinhababu, 2009, p. 483). Furthermore, Darwall argues that any action can be explained by an agent's desire (1983, p. 37). The thought is that when we are moved by a desire we are moved by awareness of that desire, and the desire itself can be moved by other events, which are connected with the object that the desire is about. Darwall supports Nagel (Nagel, 1970, p. 29) in saying that the only way that the agent's desires can explicitly involve reasons is if their reasons refer to the agent's desires, and this gives rise to a motivation and thus the relevant desire (Darwall, 1983, p. 37). However, Sinhababu disagrees as different agents with the same pre-existing desires will be motivated to do different actions in the same circumstances. For example, consider the case of giving to charity; two people may have the same pre-existing desire to give to charity, but their actual desires may be different. For example, one may desire to give to Oxfam and the other may desire to donate to the RSPCA.

Coupled with this, Sinhababu (2009, pp. 473–475) poses the objection to the Humean Theory's explanation of feelings of obligation (although he does overall advocate for the Humean Theory) when it comes to motivation. He says that the Humean Theory cannot adequately explain the feelings of obligation experienced when one is motivated. When an agent feels obligated to do ϕ, they are often not motivated by an appropriately related desire, but they do have a relevant belief, which is that they should do what they feel obligated to do. In this particular circumstance, they feel obligated to do ϕ because they believe that ϕ is the right thing to do. It should be noted that, Sinhababu draws attention to the works by Schurman (1894, p. 643) and Sorley (1935, pp. 66–67), where they claim that there are many cases where the feeling of obligation is present, but when, one is not desiring to do the action. For example, if a student feels obligated to do their homework because their teacher tells them to, and thus they feel motivated to do their homework, then, in fact, they do not desire to do their homework, but they believe that they should do their homework. Moreover, the feelings of obligation

seem different to the feelings of desire. One might think that the feeling of obligation is what one *has* to do, whereas feelings of desire is what one *wants* to do and thus, feelings of obligation seem like a different motivational state which the Humean have not accounted for. This objection is important because the Humean Theory of Motivation fails to explain the genuine phenomenological nature of motivation when agents feel obligated to do an action (Sinhababu, 2009, p. 473). Moreover, the Anti-Humean appeals to the feeling of obligation when an agent is motivated.

In response to this objection Humeans can reply as follows: the Humean can explain feelings of obligation in Hume's terminology by saying that the person has more violent passions towards an action when they feel obligated to do it. For example, if I feel obligated to donate to a children's charity combatting poverty, when physically seeing the children in poverty, my passions may grow more violent due to an emotionally compelling response, which will mean my feelings of obligation become stronger. Therefore, because they are compelled by these passions they have a stronger feeling of obligations. This shows that feelings of obligation are compatible with the Humean Theory of Motivation, and therefore, this further shows that the Humean Theory is worth taking seriously in metaethical debate.

8. Conclusion

In this chapter, I have argued that the Humean Theory of Motivation should be taken seriously in metaethical discourse in order to present and argue for my OCD case in my next chapter (Chapter 6). In this chapter, I have first outlined the debate between the Humean and Anti-Humean regarding how people are motivated. I have also explored some key objections and defended the Humean Theory against them. Moreover, the objections that I have chosen to discuss in this chapter are relevant to my OCD case because they all show either a relevant belief or an appropriately related desire may not be needed for an agent to be motivated and thus show the landscape that I am writing in when I present my OCD case to the Humean Theory of Motivation.

Chapter 6: Does Obsessive Compulsive Disorder provide a new Counterexample to the Humean Theory of Motivation?

1. Introduction

In this chapter, I consider my Obsessive compulsive disorder (OCD) counterexample to the Humean Theory of Motivation. I argue that some agents with Obsessive compulsive disorder (who meet certain conditions specified later in this chapter) may be able to provide a counterexample to the Humean Theory of Motivation. This is because the Humean Theory states that a motivation occurs if and only if a relevant belief and an appropriately related desire are present. The counterexample is that when some patients with OCD (who meet certain conditions) are motivated to do OCD compulsions, although there is a relevant desire present, there may be no appropriately related belief present, and thus this may show that the Humean Theory is false. I do this in the following way: first, I outline precisely what OCD is, then, spell out my OCD case to the Humean Theory of Motivation. Next, I move to consider some objections and some replies to my OCD case. Finally, I conclude that my OCD case may be a counterexample to the Humean Theory as when some individuals with OCD, who meet certain conditions, are motivated to perform OCD compulsions, they may not have the relevant belief to do the compulsion, but they do have the relevant desire (relevant to the compulsion).

2. What is Obsessive compulsive disorder (OCD)?

I have discussed, in the previous chapter, some apparent counterexamples that show that individuals may, in some situations, be motivated without following the Humean belief desire pair pattern, namely *besires* and *aliefs*. I have also discussed other Anti-Humean accounts, but they argue that only a belief can motivate an agent. My OCD objection proposes that motivation to perform an OCD compulsion can occur without a relevant belief present (but with a relevant desire). Furthermore, the agent knows that they have desires, but they do not know that the desires are

motivating them. It may look like people are motivated by a relevant belief and an appropriately related desire, but they might be motivated by an entirely different mental state (and this mental state might not consist of a separate belief and desire), or they might be motivated solely by a belief or a desire both of which are not consistent with the Humean Theory. This is because it seems like my OCD case is a case where an agent has a relevant desire but not an appropriately related belief and is still motivated to perform the OCD compulsion. In other words, it seems like the agent does not have an appropriately related belief (they have a belief, but it seems to be the opposite of a belief that would motivate a compulsion, so it is not appropriately related). This will contribute to the debate between the Anti-Humean and the Humean; to contribute to the debate overall, one needs a clear counterexample of where one is motivated by a relevant desire (or an appropriately related belief) only and I claim that my OCD counterexample does this.

OCD is a mental health disorder in which the sufferer feels anxiety or distress when they experience obsessions (see Fernandez et al., 2024, p. 1 for more details on this, also see Rachman, 1997). They then (usually) turn to a compulsive action to relieve their anxiety. For the purposes of this chapter, anxiety is an unwanted, unpleasant state with similar features to pain.[65] OCD patients feel "driven to perform repetitive behaviours in response to an obsession or according to rules that must be applied rigidly" (APA, 2013, p. 235).[66] The two components, obsessions and compulsions,

[65] One should note that there are different types of anxiety disorders. Recently in the DSM (1969) and the ICD (World Health Organization, 2018, pp. 40–43), Obsessive compulsive disorder has been classed as a different category of disorders than anxiety disorders, but I claim that Obsessive compulsive disorder is similar to other anxiety disorders. Anxiety disorders include, but are not limited to, generalized anxiety disorder, post-traumatic stress disorder and different types of phobias. The question of what anxiety disorder is, is a fascinating question, as is what anxiety itself is, and should be explored further, but is beyond the scope of this chapter. For our purposes, anxiety disorder will be described as excessive fear, anxiety and related behavioural disturbances (APA, 2013, p. 189). However, there is a distinction to be made between fear and anxiety: "fear is an emotional response to real or perceived imminent threats, whereas anxiety is the anticipation of future threats" (APA, 2013, p. 189).

[66] There are different types of OCD. Some different types include, but are not limited to, harm OCD, contamination OCD, sexual OCD, religious OCD (Steketee, 1999, p. 1). OCD has the same symptoms, but it can be about different thought content. For example, if one has religious OCD, they will get anxious about having unwanted thoughts of blasphemy. If one has contamination OCD, they will have unwanted thoughts that germs could make them ill. One school of thought as to why particular types of OCD intrusive thoughts come about to individual patients are what one values most (Rowa and Purdon, 2003). For example, if one values the principle that one should not harm others, then one may have intrusive thoughts of harming others. This is because, although most individuals have intrusive thoughts, because the patient is particularly disturbed by this idea the patient will want to get rid of it,

are important in understanding how the disorder manifests in individuals with the disorder, so I lay them out here:

"Obsessions are defined by (1) and (2):

> 1. Recurrent and persistent thoughts, urges, or images that are experienced, at some time during the disturbance, as intrusive and unwanted, and that in most individuals cause marked anxiety or distress"

An example of an obsession is 'if I do not go to church, then I will go to hell'.

> "2. The individual attempts to ignore or suppress such thoughts, urges, or images, or to neutralize them with some other thought or action (i.e., by performing a compulsion)" (American Psychiatric Association, 2013, p. 237).

An example of a compulsion that may be paired with the previous obsession in (1) is going to church.

These two features are vital to discussing what obsessions are, and how they may cast doubt on the Humean Theory. Point (2) is further enforcing that these thoughts are unwanted. In fact, it is often thought that one of the reasons why people with OCD get anxious about intrusive thoughts is because they try to suppress the thought which leads to them thinking about it more (Steketee, 1999).

I turn to the second part of OCD, which is compulsions. Compulsions are defined by (1) and (2):

> "1. Repetitive behaviors (e.g., hand washing, ordering, checking) or mental acts (e.g., praying, counting, repeating words silently) that the individual feels driven to perform in response to an obsession or according to rules that must be applied rigidly.

---

and because of this, the intrusive thoughts keep reoccurring, which may make an intrusive thought an obsession.

2. The behaviors or mental acts are aimed at preventing or reducing anxiety or distress, or preventing some dreaded event or situation; however, these behaviours or mental acts are not connected in a realistic way with what they are designed to neutralize or prevent, or are clearly excessive" (APA, 2013, p. 237).

The compulsions are a response to anxiety or distress that the agent may have; they are motivated to do these compulsions as they are trying to relieve their anxiety or prevent some dreaded event from happening. These two symptoms (obsessions and compulsions) have to take up a significant amount of time or "must cause significant distress or impairment" to the patient "in social, occupational or other important areas of functioning" (APA, 2013, p. 237).[67]

An example of a thought that can become an obsession would be the thought that they should do something sexually deviant relating to children. This could start as a one-off intrusive thought but would turn into an obsession if the person experiences anxiety in response to this thought.[68] This is an intrusive thought to many, and in fact, it may be the opposite to what the agent wishes to do (it may be that the agent has no deviant sexual urges whatsoever and this is why they are having this intrusive thought). Many people, who do not suffer from OCD, are able to dismiss this thought, without much worry or anxiety, whereas people who suffer from OCD will experience anxiety when having that thought, but do not want to act on it (see Fernandez et al., 2024, p. 1; Rachman, 1997).

The more that the patient tries to get rid of an intrusive thought, the more the thought recurs (see Salkovskis et al., 1995, p. 285; Salkovskis and Campbell, 1994, p. 1). They develop an obsession which then creates compulsive behaviour (such as avoiding children in the above case) that is caused by the intrusive thought. Crucially, most OCD sufferers are aware that their compulsions are not going to prevent their obsession from happening who notes that effective therapy should

[67] The DSM-5 says longer than one hour.

[68] Intrusive thoughts are commonly experienced by the general population. For example, thoughts of harming or injuring a person is a common intrusive thought experience ((Fernandez et al., 2024, p. 1)). For further discussion on evidence that intrusive thoughts are commonly experienced by many in society today, see Fernandez et al.(2024), Kenrick and Sheets (1993), Rowa and Purdon (2003). Individuals with OCD that experience these thoughts go to great length to prevent them from harming others (Fernandez et al., 2024, p. 1; also see Veale et al., 2009, p. 338).

involve the patient examining their compulsions and the OCD patients can recognise that they are being irrational (see for reference that some OCD patients can examine their compulsions Steketee, 1999 for evidence; see also the DSM-5 APA, 2013, p. 238 for further details on how most OCD sufferers have good or fair insight). This compulsive action can be subtle, but also can be very noticeable. An example of a subtle compulsive action is reassurance; a sufferer will seek reassurance for an obsessive thought, even if they know the intrusive thought is not true. An example of reassurance in this case would be for a patient to ask a family member whether they will get ill if they touch a doorknob because of contamination of germs. The family member would then typically reassure the patient that 'they would not get ill' and this would relieve the patient's anxiety temporarily (but adds to the strength of their compulsion) (see Jacoby and Abramowitz, 2017, p. 175). An example of a noticeable compulsive action, however, is one of visible activities. This could be checking that they have turned the stove off 30 times, so that their house does not burn to the ground (APA, 2013, p. 238).

It is often the case in OCD patients that OCD compulsive actions turn into habits and OCD habits can result in a lack of flexibility in performing OCD compulsive actions, when obsessions occur, for the patient (Gillan, 2017, p. 161). For example, if an agent has an obsessive intrusive thought of if they do not touch the knob on the door three times then they will get ill then they will do this compulsion no matter the circumstance or no matter how inconvenient it is. This is demonstrated by an agent with OCD who may be made late for work (which they do not want to be late for), because they feel like they have no option but to do the compulsion (even if they know that the time taken for the compulsion will make them late for work). In other words, there is a lack of flexibility in whether they perform the compulsion; they must perform the compulsion no matter the circumstances. Moreover, it should be noted that compulsive behaviours are not generally focused on achieving positive outcomes for the agent but rather avoiding negative ones, and the habitual action performed by the patient is used to avoid the perceived consequence that the agent fears will happen, if they do not perform the OCD compulsion. For example, consider the case in which the agent has the intrusive thought that if they do not go to church, the negative consequence is that they will go to hell (Gillan, 2017, p. 164). They may avoid objects or stimulus that triggers an intrusive thought; this can

be something that is active like doing an action to reduce the anxiety that comes from the particular intrusive thought, for example, washing your hand an unreasonable number of times, or passive, for example, avoiding places that trigger intrusive thoughts. Moreover, Jacoby and Abramowitz (Jacoby and Abramowitz, 2017, p. 171) agree by saying that one of the diagnostic criteria of OCD should be the inability to cope with unpredictability and ambiguity. Furthermore, there are different thresholds of ambiguity that one may struggle with and thus individuals must learn how to deal with ambiguity when they are suffering with OCD (Jacoby and Abramowitz, 2017, p. 171). Moreover, the severity of OCD in a patient is characterised by the level of ambiguity that one may be able to tolerate. For example, a person with mild OCD may simply have a higher tolerance of ambiguity (Ladouceur et al., 1997) or on a more severe level people with OCD may draw threatening interpretations of ambiguous information (see Reuman et al., 2015, p. 112) or on the most severe level they may find ambiguity to be unmanageable and distressing (e.g. see Buhr and Dugas, 2002; Jacoby and Abramowitz, 2017, p. 171).

An example of an OCD patient drawing problematic interpretations on ambiguous situations may be if an agent does not know whether their environment is clean enough not to make them seriously ill, so, therefore, they may worry that the surrounding environment may make them ill. This ambiguous information causes the agent in question to be afraid of the OCD intrusive thought; the thought that scares them is the OCD thought *may* happen (even though the agent believes it is unlikely – in most cases of OCD patients, they may have good or fair insight into their intrusive thoughts). They draw threatening conclusions from ambiguous information and the threatening conclusions cause their anxiety. As explained above, the struggle to cope with ambiguous information can vary to different degrees. A mild OCD patient may have a small amount of ambiguity surrounding their thoughts but may be able to tolerate the ambiguity, whereas a patient with a severe level of OCD may find ambiguity intolerable and may struggle to live with the ambiguity. These individuals with OCD may also have different and distressing mechanisms to cope with uncertain or ambiguous situations (e.g. (Jacoby et al., 2014, p. 171)). Treatment for OCD is conventionally, therefore, focussed around the patients learning how to live with a level of uncertainty in their daily life, and for the OCD

patients to realise that they cannot be one hundred percent certain that the negative outcomes, which they fear, will not happen (Jacoby and Abramowitz, 2017, p. 175).

Crucially, the DSM has three categories of patients with OCD, which describe how they understand their obsessions (which is important for the chapter as I am only referring to the first thought, the OCD sufferers that have good to fair insights).[69] The first is where the patient understands their obsessions with "good or fair insight"; this means that the patient "recognises that the obsessive compulsive disorder beliefs are definitely or probably not true or that they may or may not be true" (APA, 2013, p. 238).[70] The second is when the patient has "poor insight: the individual thinks obsessive compulsive disorder beliefs are *probably* true" (APA, 2013, p. 238). The key thing to note here is that in the category of poor insight the patient knows that the beliefs are probably true, meaning that they think that it could be false. This is distinct from the third category where the individual does not understand (and there is no possibility of them understanding) that their beliefs are false. (They have "absent insight/ delusional beliefs: the individual is completely convinced that obsessive compulsive disorder beliefs are true" (APA, 2013, p. 238)). These categories in the DSM give insight into how the sufferer understands these obsessions and gives an idea of the spectrum of how they might understand these obsessions.

It is important to note that most individuals with OCD will have 'good or fair insight' into their obsessions, e.g., if they have an obsession that if they do not clean the door handle three times, then they will be contaminated with germs and become ill, they will know that this content is not true (APA, 2013, p. 238) (see also Jacoby and Abramowitz, 2017, p. 175; Stein et al., 2015, p. 670 for further details on how OCD sufferers with good or fair insight know that the content of their intrusive

[69] The Yale-Brown Obsessive Compulsive Scale was designed to measure OCD patients' insight into their symptoms: it was designed to particularly provide a specific measure of the severity of patient insight into their symptoms that is not based on the number or the type of obsessions and compulsions that an OCD sufferer has. For further details, see Goodman et al. (1989, p. 1006)

[70] An OCD patient with good insight may display the same obsession and compulsions as a patient with poor insight, but an OCD patient with good insight will realise the irrationality of their obsession and compulsion. One should note that there are no specific criteria as to what good or fair insight amount to in an OCD patient, but it is often recognised to be that the patient recognises their own intrusive thoughts or impulses are irrational. For further details on the fact that there are no specific criteria on OCD insight, see Marras et al., (2016, p. 325)

thought is not true when they are performing a compulsion, but they still do the compulsion anyway).[71] Few will have poor insight, e.g. if they have the same obsession with door handles and the contamination of germs, then they, probably, think that they will get ill from the contamination of germs if they do not clean the door handle three times (APA, 2013, p. 238). Few will have absent or delusional obsessions which means that the patient is convinced, in my case, that if they do not clean the door handles three times then they *will* become contaminated with germs and become ill (APA, 2013, p. 238).[72] These details are important as they help us to understand how the sufferer understands the need to perform compulsions.

As we have seen above, there are different levels of severity of OCD and there are also different stages of OCD recovery when the patient elects to have treatment. The main way in which OCD is treated is a type of cognitive behaviour therapy called Exposure and Response Prevention therapy (ERP) where an agent is gradually exposed to objects or situations which may trigger their intrusive thoughts and are encouraged not to perform their compulsions even though they may in fact be motivated to perform their compulsions. (see Reid et al., 2021, p. 2; Ferrando and Selai, 2021, p. 1 for further discussion on the effectiveness of ERP therapy for OCD patients). There are different stages that the OCD patient has to go through in order to recover, however, when the agent is first commencing treatment, they will believe that the compulsion is the only way to relieve anxiety from the intrusive thought and

[71] See Davey (2021, p. 195) for further evidence on this. On the page noted, there is a particular case study. The case study is about a man who has intrusive thoughts (which turn into obsessions) to keep his body clean. Noticeably, the case study claims that he does not believe that his compulsions are necessary to stop the intrusive thought reoccurring or happening, but he is instead driven by his urge to clean his hands incessantly because of the relief of anxiety he would feel if he did the compulsive action (I argue that if he were to undergo a course of OCD therapy, he has the potential to follow the same pattern as set P (where the agent believes that the compulsion will not relieve anxiety but still desires to relieve anxiety and therefore they are still motivated to perform the compulsion – see below in the chapter for further details)– the book specifically notes that he has not had the opportunity to go to OCD therapy). This is an example of a case where the agent believes that their compulsions are not necessary for the intrusive thought not to happen, but desires to do it anyway (in this case, he may be motivated to do the compulsion to alleviate his anxiety because he has not undergone therapy). This is also supported by the research conducted by Brakoulias and Starcevic (2010, p. 156) where they claim that patients readily acknowledge that their irrational thoughts were because of OCD. It should also be noted that believing that the compulsion is "not true" may lead to the agents believing that the OCD compulsion is false. The agents could also believe that their compulsion is "probably true". Different OCD sufferers (in this category) will have different belief insights into their OCD thoughts (see APA, 2013, p. 238 for further details).

[72] The DSM says that those patients with OCD who have absent or delusional obsessions make up 4%, or less, of OCD sufferers (APA, 2013, p. 238).

desire to relieve that anxiety and therefore will perform the OCD compulsion. However, in the middle and last stages of therapy it is hoped that the agent will understand that the compulsion will not relieve anxiety (see for example Steketee, 1999 who gives a full log of what should happen during therapy sessions for OCD, it is highlighted that the aim of therapy is for the agent to understand that the compulsions will not relieve anxiety). In fact, the OCD sufferer may understand that it provokes more anxiety, but they still have the desire to perform the action (and therefore will be motivated to perform the compulsion).

These are the cases which I discuss in this chapter – the cases where the agent believes that the compulsion will not relieve anxiety but still desires to relieve anxiety and therefore, they are motivated to perform the compulsion. These agents I will refer to as 'set P'. This is important because when the OCD patient has recovered (or can manage their OCD to the extent that it does not have a significant impact on their lives), the agent will not perform the compulsion. This is because they will believe that the anxiety will not be relieved by performing the compulsion and, at the same time, desire not to be anxious and therefore will not be motivated to perform the compulsion. In fact, they may be able to recognise that they are experiencing an intrusive thought and employ better ways of dealing with it to relieve the anxiety instead of performing a compulsion.

The agents that I am discussing are a subset of OCD patients (set P). There is in fact a time where an agent may, particularly if the agent is in therapy, desire to relieve their anxiety, but believe that the compulsion will not relieve anxiety but is still, nonetheless, motivated to do the action. This does not follow the belief-desire pattern that the Humean endorses, but rather it seems to support an Anti-Humean account, as it seems that the agent is motivated solely by the desire (rather than a relevant belief and an appropriately related desire). The intrusive thought that the OCD patients experience does not count as a relevant belief because the OCD patients do not endorse this thought (in fact, they are rather disgusted by the intrusive thought, which is why the intrusive thought is very distressing and triggers the obsession compulsion disorder pattern, see, for example, Rowa and Purdon (2003) for evidence of the OCD patients being disgusted by their intrusive thought). Moreover, an intrusive thought does not fit Owens' (2003, p. 283) definition of

belief as the OCD patients (or other individuals that experience intrusive thoughts) do not endorse the thought and thus do not find the thought rational (or irrational) or justified (an intrusive thought should be considered as a 'fleeting' thought which the agent has no control over). Furthermore, the therapy for OCD patients, which is usually applied, is a gradual process (i.e. ERP therapy, A/B therapy); there will be some point in therapy where the agent will know the anxiety will not stop if they perform a compulsion.[73] However, there will be other cases where the patients will not know how to stop the anxiety and thus will perform the compulsion. I am looking at the cases where the agent believes that the compulsion will not relieve anxiety but at the same time, they are still motivated to do the action as I argue that cases where the agent is just motivated to relieve anxiety can be described in terms of the Humean Theory. This is because cases where the agent is motivated to relieve anxiety can be described in terms of relevant belief and an appropriately related desire which leads to motivation of the compulsion which relieves the anxiety (see Steketee, 1999 for a step-by-step process of therapy).

Therefore, in the next section, I discuss my argument (and the supporting evidence) against the Humean Theory that there are OCD sufferers who are motivated with solely a desire when they are motivated to perform OCD compulsions.

3. The OCD Counterexample to the Humean Theory: How are OCD Patients Motivated to do their Compulsions?

3a. My OCD Counterexample

A usual objection to the Humean Theory is that beliefs alone are sufficient to motivate an agent. However, I am claiming in my OCD counterexample that desires

[73] A/B therapy is a therapy where the therapist offers an alternative explanation to the OCD patient's intrusive thoughts rather than the OCD patient's current explanation of their intrusive thoughts (Challacombe, 2011, pp. 135–136). Moreover, the OCD patient's explanation might be the fact that they had to deal with their intrusive thought by compulsion, whereas the therapist might offer an alternative explanation where the agent does not choose the compulsion. For example, if the agent washes their hands 10 times because they are afraid of contamination, then the therapist would offer an alternative explanation and solution (and this explanation would not feature the agent performing the compulsion) as to the outcome of the scenario in the intrusive thought.

alone are sufficient to motivate an agent to perform their OCD compulsion.[74] I argue that the reason why the Humean Theory does not explain why OCD patients (in set P) are motivated to perform compulsions can be spelt out like this:

P1 If the Humean Theory is true, then motivation occurs iff there is a relevant means-end belief and appropriately related desire present.

P2: There are some actions (OCD compulsions) in which the agent is motivated without an appropriately related means-end belief being present (but a relevant desire is present).

C: Therefore, the Humean Theory is false.

The argument against the Humean Theory relies on the premise that if the Humean Theory is true, motivation occurs iff a relevant belief and an appropriately related desire are present. If there was a case where either a relevant belief or an appropriately related desire was not present, then according to the Humean Theory, the agent would not be motivated. However, in the case of OCD, I claim agents may be motivated to perform compulsions when only a relevant desire is present. I will now discuss some case studies on how agents may be motivated to perform compulsions with the absence of any appropriately related belief.

Consider the following case where a person has an obsessive contamination intrusive thought; it could be along the lines of '*if you do not wash your hands three times then you will get seriously ill'*. Most OCD sufferers would not believe that they would get seriously ill if they did not wash their hands three times, but they would still wash their hands anyway (which would be the compulsion in this case), because they are compelled by their obsession. They are still motivated to perform the compulsion, but they may not believe that the compulsion will prevent the dreaded event from happening or relieve anxiety.[75]

[74] The claim that beliefs alone are sufficient to motivate is part of the wider discussion of metaethics. This is because metaethics discussions revolve around whether moral judgements are beliefs i.e., whether Cognitivism is true.

[75] There is a real case study of an OCD patient (who suffers from intrusive thoughts about his finances), who appears to have a relevant desire without an appropriately related belief to perform a compulsion. These compulsions were repeatedly checking with his wife how much she has spent at various points throughout the day and he needed constant reassurance from his wife that she was not

In more general terms, to perform an obsession-compulsion pattern, an OCD sufferer first has an intrusive thought, which then turns into an obsession as the intrusive thought keeps repeating. This is because when experiencing the intrusive thought, the OCD patient's first response will be anxiety (or distress).[76] Because they experience anxiety (and anxiety is an unpleasant mental state), they will try and get rid of the intrusive thought and actively try and push the thought away. Therefore, the thought will keep reoccurring as the more an agent will try and push a thought away, the more they will think (and worry) about the thought. They then are motivated to perform a compulsion, but the obsessive compulsive cycle continues, as the compulsion only relieves anxiety temporarily, and the intrusive thought will reoccur. This case is an apparent counterexample to the Humean Theory of Motivation as it seems, at least at the outset, that an OCD patient (displaying the characteristics explained above) has a desire to relieve anxiety but does not have the belief that the compulsion will relieve anxiety. This calls into question the Humean Theory of Motivation as they say all agents who are motivated must have a relevant belief and an appropriately related desire and the OCD patient does not seem to have an appropriately related belief.

3b. Objection: Redescribing the Case through the Humean Lens and my Response

The Humean Theory would claim that they could redescribe my OCD case by saying that there is a relevant belief and an appropriately related desire present, which leads to the motivation to perform the compulsion. This is because the Humean Theory claims that there is a belief that performing the compulsion will make them less anxious and that there is a desire to relieve the anxiety (or perhaps to

---

overspending, and that the family were not in debt (which is a common subtle compulsion that OCD patients sometimes exhibit). With therapy, he managed to reduce the frequency of doing his calculations to once every day, then once every week, because he realised that these compulsions were not helping him to relieve his anxiety, prevent the intrusive thought from occurring or avoiding financial debt. This is a case taken from Purdon and Chiang (2016, p. 507).

[76] I say 'or distress' here in brackets as the OCD patient may not suffer with anxiety but they may have a level of distress when they do not perform the compulsive action. There is a question about whether anxiety and distress are the same thing. I claim that they are not as anxiety is characterised as continuous worry whereas distress can be more upset as well as worry. From now on, however, to avoid confusion I only refer to the patient feeling anxiety when performing compulsive actions, but readers can take anxiety to mean distress as well, as for the purposes of my chapter, they are interchangeable.

avoid the dreaded event). Therefore, the agent is motivated to do the compulsion. The Humean Theory would therefore claim that this fits with their theory of motivation as there is both a relevant belief and an appropriately related desire present.

It should be noted that it looks likely that some OCD patients do actually believe that their compulsion will provide short term anxiety relief, but there may be other OCD patients in set P that believe that their compulsion will provide long term anxiety relief. However, it looks quite likely that there are individuals (set P) with more insight into their OCD intrusive thoughts who do not believe that the compulsion will stop the dreaded event from happening or that the compulsion will relieve anxiety in the short or long term. These individuals may be a counterexample to the Humean Theory. However, from the existing psychological evidence that I have consulted, it is hard to say for definite, that there are OCD sufferers who do not believe that their compulsion will provide them with short-term or long-term relief from their anxiety or belief that the compulsion will stop the dreaded event from happening. Moreover, further psychological evidence would help the debate progress. In other words, if psychologists could highlight whether some OCD patients lack both beliefs (that the compulsion will stop the dreaded event and the compulsion will relieve anxiety) then this may serve as a counterexample to the Humean Theory. My chapter will specifically talk about the OCD patients in set P who are referenced in this paragraph.

The Humean Theory of motivation does not accurately capture what is going on in all cases of OCD (I am referring to cases that I previously called ‘set P’) when an agent is motivated to perform a compulsion (or indeed is performing a compulsion), as the Humean Theory is making two assumptions, one of which may be incorrect, when explaining why an agent (in set P) is motivated to perform a compulsion. Firstly, I grant the Humean that there is a desire to perform the compulsive action which in this case could be based in the belief that it will relieve anxiety (or avoid the dreaded event or outcome from happening). However, I claim that we may not be able to find a belief in the motivation to perform the compulsion. This is shown by many patient transcripts with therapists when they are exploring their intrusive thoughts. It is common practice for a therapist to help a client to bring out their

beliefs and teach them to realise that these are different (and can be unrelated) to the compulsions that they are performing. For example, there was a client that had intrusive thoughts about molesting children. From the transcript, it is obvious that they were disgusted by those thoughts but they still were anxious that they might carry out their obsession anyway.[77] This is an interesting case as they did not believe that they would hurt children, nor believe that the compulsion would make the intrusive thought go away, but they were more ashamed of and disgusted by the intrusive thought (see Bream et al., 2017 for more transcripts which suggest identical psychological patterns in multiple OCD patients (set P) with differing obsession-compulsion pairs; also see the DSM-5 APA, 2013, p. 238 for further details that most OCD sufferers - and those who I am referring to have good or fair insight into their obsessions so would be disgusted by the intrusive thoughts) which led them to perform the compulsive action of avoiding children.

In the patient who is suffering from OCD behaviour due to their intrusive thoughts about molesting children; two distinct claims need to be discussed. The first is, the agent who has OCD might be motivated to perform the compulsion to avoid molesting children; the agent will feel so disgusted by the intrusive thought (as the intrusive thought will go against their morals). However, the Humean might say that the agent in this case is motivated to do the compulsion as a way of avoiding acting on the intrusive thought. In other words, by avoiding children (the compulsion in this case) there is no possible way that they could act on the thought. Moreover, the Humean might claim that the OCD patient believes that avoiding children is likely to stop the intrusive thought reoccurring, and because the intrusive thought is unwanted and unpleasant they may be motivated to avoid children. However, I propose that these two claims that the Humean defends (with regards to the OCD case) are

[77] See transcript in Bream et al. (2017, pp. 74–79). Moreover, note that the OCD sufferer also feels shame when they experience these intrusive thoughts, which turn into obsessions, and this is further evidence that they do not want the intrusive thoughts, but is not evidence that they are motivated to do the compulsion because they want to relieve the anxiety (this I will claim further down in the chapter). For further evidence that OCD sufferers do not want intrusive thoughts, see Bream et al. (2017, pp. 164–165). However, there is evidence to suggest that because agents are disgusted by their intrusive thoughts, they perform compulsions. However, I claim that this does not get the precise explanation right as to why OCD sufferers perform OCD compulsions as OCD sufferers believe that even if they perform the compulsion to try and alleviate the anxiety and disgust, they believe that these compulsions will not get actually rid them of the anxiety and disgust. For further discussion on OCD patients performing compulsions because of their disgust from their intrusive thoughts. see Starcevic et al (2011, p. 455).

plausible if we are discussing any OCD case outside set P, whereas the case that I am discussing is a sub-section of OCD patients (patients who realise that their compulsions will not stop the intrusive thought from occurring or prevent the content of the intrusive thought from happening). This is because the sub-set of OCD patients (set P) I am discussing, do not believe that the compulsion will A) will make it less likely that the content of the intrusive thought will occur, or B) make it less likely that the intrusive thought will occur. This is because, through therapy, the OCD patient (in set P) is trained to understand that neither A (that the OCD compulsion will make the content of the intrusive thought less likely to happen) nor B (that the intrusive thought will happen) will occur. Coupled with this, based on my earlier description of OCD, the majority of OCD patients will have "good or fair insight" (APA, 2013, p. 238) and thus will realise that their compulsions are not necessary to stop the intrusive thought from occurring. Furthermore, they will be taught (through therapy) ways to deal with their intrusive thoughts that do not involve performing compulsions (for example, ignoring the intrusive thought) and therefore the patient will realise that the OCD compulsion is not necessary but will still be motivated to perform the OCD compulsion anyway.

More precisely, there are two beliefs that need to be distinguished when an OCD patient (in set P) is motivated to perform a compulsion; belief A, which is that doing the compulsion will prevent the dreaded event (equivalently: if I do not do the compulsion the dreaded event will occur) and belief B: the belief that doing the compulsion will relieve the patient's anxiety. Belief A is easier to deal with, so I will address it first. Imagine an OCD sufferer having an intrusive thought that if they do not clean their hands three times, they will get ill. Most of the sufferers (and I am not discussing the sufferers with poor or absent insight and/or delusional beliefs) will understand that the intrusive thought is false and thus one does not need to clean one's hands three times a day for one not to suffer from illness, yet they are still motivated to do the action.

Belief A can be dealt with in the following way and is relatively easy and less controversial than Belief B. This is because there is much evidence to suggest that most OCD patients (in set P) do not believe that the content of their intrusive thoughts will become reality if they do not do the compulsion. Challacombe (2011)

(see also Veale and Wilson, 2005; Rachman and Silva, 2009) provides evidence by saying that although all individuals experience intrusive thoughts and they are unwanted thoughts, OCD patients (in set P) try and push intrusive thoughts away because they cannot deal with them (see Fernandez et al., 2024, p. 3; for further details see Rachman, 1981). The very nature of an OCD patient's (in set P) intrusive thought, being an intrusive thought, means that it is not a true description of what the individual desires or believes is happening in the world – whether the intrusive thought is a belief or a desire depends on the content of the intrusive thought (Veale and Wilson (2005), Rachman and Silva (2009) and Challacombe (2011)).[78] Thus, the intrusive thought is not the sufferer's description of the world, rather something other, something intruding on them. Perhaps another way to put this is that the intrusive thought in its very nature is external to the individual ((Veale and Wilson, 2005), (Rachman and Silva, 2009), (Challacombe, 2011) and (Steketee, 1999)). There has been much evidence for intrusive thoughts, which is reflected in the DSM manual (APA, 2013, p. 237) and various psychological literature (which is stated in sources above and Bream et al. (2017)). Another example is Brock et al. (2024) who claimed that some OCD patients can recognise the irrationality of their OCD obsessions and compulsions (particularly when they have good or fair insight) but these patients feel as though they cannot control their OCD obsessions and compulsions (which arrive from their intrusive thoughts). This implies that OCD patients (in set P) can recognize that their thoughts are irrational, but still perform their compulsions anyway (because they have the desire to relieve their anxiety and to stop the dreaded event from happening, but do not have the appropriately related belief that their compulsion will relieve their anxiety or stop the dreaded event).

I argue that the Humean Theory is mistaken in thinking that Belief B is present in OCD patients (in set P) when performing a compulsion. This is because the sufferer

[78] The content of an intrusive thought may vary depending on the person, as already alluded to in the text, and the thought either could be belief-like or desire-like depending on the content of the thought. For example, if I have an intrusive thought that 'if I do not clean the house 10 times I will be contaminated by germs' (which is a common intrusive thought for people with contamination OCD), this intrusive thought looks like a belief. However, if I had a different intrusive thought, which had the characteristics of an urge to perform an action, then this urge looks desire-like. For example, I may have an inappropriate sexual intrusive thought, which is an urge to sexually assault a child. This is demonstrated through the following common intrusive thought (which is desire-like), which might be 'if I do not look a certain way, I might have an urge to molest a child' (This is a common intrusive thought for people with sexual OCD).

knows that the compulsion will not result in anything they want (it will not stop the intrusive thought from occurring or relieve any anxiety caused from the intrusive thought, but they still perform the compulsion anyway). The DSM-5 highlights that most patients with OCD have "good or fair insights" into their obsessions which means that they will know that their obsessions are irrational (APA, 2013, p. 238). Moreover, Harrison (2010) claims that OCD patients (in set P) are usually aware that the compulsions are illogical and thus, usually try to hide them (which usually makes their obsessions stronger). The fact that they think that these compulsions are illogical implies that they do not believe that their compulsions will relieve anxiety, for if they did, then these compulsions would not be illogical.[79]

Furthermore, in order to explain why OCD patients (in set P) do not believe their compulsions will relieve anxiety, I now turn to explain and discuss what happens in a specific type of cognitive behaviour therapy, which is Exposure and Response Prevention (ERP) (see Bream et al., 2017; Harrison, 2010; Rachman, 2013; Rachman and Silva, 2009; Steketee, 1999). In ERP, the patients are gradually exposed to triggers and are encouraged to change the way that they respond to these triggers. This approach is also reflected in other therapeutic techniques, more specifically 'Correcting your Faulty Beliefs and Assumptions' therapy (Steketee, 1999) (although I claim that this therapy's name is not entirely precise, as I propose that OCD patients (in set P) lack relevant beliefs about the efficacy of the compulsion). As part of this therapy, OCD patients (in set P) are asked to challenge their own ideas and correct their own cognitive errors. More specifically, it is used as a therapy to prevent relapse. The aim of the therapy is that the patient must do it without the help of the therapist. This means that the patient can identify their faulty ideas (which means that they are not beliefs) and can correct the faulty ideas. The fact that the patient can identify their faulty ideas means that they do not believe these ideas, as if they did, they would believe their obsessions (which is not the case for the majority of OCD patients which is shown by symptoms characterised by the APA (2013, p. 237)). This is particularly apparent as one of the key features of Owen's account of belief that I discussed in Chapter 5, is that, for something to be a

[79] It should be noted here that I am referring to OCD compulsions only. I acknowledge that there are other compulsions that are superstition compulsions, but these are different cases and need to be dealt with differently.

belief, it must appear to be right to the believer, and therefore these faulty ideas cannot be their beliefs. This is the case because the believer can (and can be trained to) correct their faulty ideas meaning that they do not believe them. More precisely, there are particular cases of intrusive thoughts (which I established earlier are not beliefs) that the OCD patients (and other individuals) do not endorse because OCD patients are very willing to correct intrusive thoughts (and the compulsions associated with the intrusive thoughts) if taught how to do so through 'Correcting your Faulty Beliefs and Assumptions' therapy.

This is relevant because the very fact that these therapies are succeeding means that they do not believe that their compulsion will relieve their anxiety, for if they did think this, then they would not be taking measures to stop performing their compulsion. OCD patients may believe that performing the compulsion provides short-term relief from their anxiety (and this is what the Humean Theory would argue), but it seems unlikely that the OCD sufferer will see the compulsion as long-term relief to their anxiety, particularly if they are in therapy ((see for example Salkovskis and Kobori, 2015, p. 204; Steketee, 1999, p. 12 for example for evidence that compulsions such as reassurance relieves short-term anxiety). This is because they would be taught that performing their compulsions will be adding to their anxiety in the long term, and thus the OCD patient is unlikely to think that the compulsion will provide any long-term relief.

Coupled with this, another therapy which is used to treat OCD patients (in set P) is A/B therapy, which is similar to 'Correcting your Faulty Beliefs and Assumptions' therapy, where the patient is taught to re-interpret the obsessive thoughts (Challacombe, 2011, pp. 135–136). The fact that they can understand and endorse the reinterpretation means that they know that their compulsions are irrational and will not stop the anxiety. This is particularly apparent because, as they discuss their obsessions in this type of therapy, it becomes clear that they do not believe that the compulsion will stop the dreaded event from happening or relieve their anxiety as they are taught to reinterpret the obsession in a different way (see Challacombe, 2011 for further details on therapy which helps the OCD patient reinterpret intrusive thoughts). The fact that they have a thought and they do not endorse it but the fact that they are still motivated to perform the compulsion means that there is a point in

A/B therapy (and other types of therapy which have been discussed in the chapter) at which they do not endorse the thought that performing the compulsion will help relieve anxiety or stop the dreaded event from happening, and yet OCD sufferers are still motivated to perform the compulsion anyway. More importantly however, through therapy, OCD sufferers are taught that performing the compulsion will not alleviate anxiety, but they'll continue to perform the compulsion anyway, even if they are aware that the compulsion is irrational and will not relieve their anxiety (for further discussion, see Harrison, 2010)).

This is further evidenced by the following case: individuals may avoid smoking as it may cause severe health issues (such as cancer). One may think that in the short term, a cigarette may relieve stress and anxiety and feel pleasurable, but in the long run, it is incredibly dangerous (particularly as it is addictive) as one may develop a severe health condition because of smoking. Therefore, one might be motivated to avoid smoking as they do not want to risk having long-term health conditions and thus decide to forfeit their immediate pleasure to relieve anxiety and prioritise their future well-being. This is similar to the case of OCD compulsions. One will not be motivated to perform their compulsions in cases of set P, because they know that this will increase their anxiety in the long run, and though performing the compulsion may provide temporary relief, the OCD sufferer will believe that it will increase anxiety in the long term and therefore should not be motivated to do the compulsion, but is still motivated because of their desire to relieve anxiety. However, against all of this, one may argue that these cases can be described in terms of the Humean Theory. For example, if an OCD patient has a belief that the compulsion will stop the dreaded event or relieve their anxiety (in the short term) and a desire to stop the dreaded event or relieve anxiety, then these cases may be able to be described in terms of the Humean Theory (one can describe the smoking case in terms of a relevant belief that smoking will feel good and a desire to feel good which is an appropriately related desire).

Ultimately, we do not know what exactly is going on in the minds of OCD sufferers when they are motivated to perform an OCD compulsion, but at least it seems plausible for some of them to continue with their compulsion even though they do not believe it serves any of their ends. In other words, it is plausible to think that

some OCD patients (in set P) believe that they desire to relieve anxiety but do not believe that the compulsion will relieve their anxiety. I do allow for the possibility for some people who believe that this brings short-term relief, and they perform the compulsion for that reason, in these cases they may do the compulsion because they are motivated to relieve anxiety. However, there are some people (in set P) that have therapy (as discussed in the chapter) and they believe in general that the compulsion will not relieve anxiety, but they will still perform the compulsion anyway. This is because they are still motivated to do the behaviour as they desire to relieve anxiety. In other words, I claim it is possible to say that some people do believe that engaging in compulsions will provide some short-term relief from anxiety or avoid intrusive thoughts. However, it is implausible to claim that this belief/desire thought pattern is true for all OCD sufferers who have gone through therapy and have been taught that the OCD compulsion will not provide relief from their anxiety. On balance, it still appears that we have cases where OCD sufferers are motivated when there is no relevant belief associated with the compulsion (but there is a relevant desire), and therefore the Humean Theory cannot account for this motivation.

My OCD case and the Humean Theory reach a deadlock in the debate. The Humean Theory could describe the OCD case in many ways, for example, they could say that the patient desires temporary relief from their anxiety which the compulsion will provide or the Humean could say that the OCD compulsion stops the repetitive intrusive thought of the sufferer from happening (and the patient believes that the intrusive thought will be stopped by an OCD compulsion). However, my explanation of the OCD case is that the OCD case has a relevant desire but not an appropriately related belief that the compulsion will alleviate the anxiety, or the intrusive thought will stop occurring in the patient (in set P). This is because the patient (in set P) knows through therapy that neither one of these are explanations for doing the compulsion, and therefore will not be motivated by these explanations.

To conclude, there is much plausibility in my interpretation of how the OCD patients (in set P) are motivated to perform their OCD compulsions, but we need to do more psychological research to find out a definite answer as to whether my OCD counterexample works as an objection to the Humean Theory of Motivation. Finally, in this chapter, I have argued for my interpretation of how OCD patients (in set P)

are motivated to perform compulsions, but I welcome the Humean to put their case forward as to how they think OCD patients (in set P) are motivated to perform OCD compulsions using a belief-desire pattern.

4. Conclusion

The Humean Theory says that motivation occurs iff a relevant belief and an appropriately related desire are present. I have shown in the case of OCD compulsions that the agent (in set P) may be motivated to do the compulsions without an appropriately related belief being present. However, I only make a modest claim as the Humean Theory and my OCD case reach a deadlock and this is similar to other Humean and Anti-Humean debates. The reason for this is that the Humean can describe motivation in OCD cases in set P in terms of a relevant belief and appropriately related desire, whereas I have claimed there may be some cases, such as the OCD case, where an agent may just be motivated by a relevant desire. Therefore, I have argued, that OCD sufferers' motivation to perform compulsions may not be able to be described by the Humean Theory, as sufferers may lack the appropriately related belief and thus, the Humean Theory may be mistaken.

Conclusion

In this thesis, I have argued that psychological perspectives should inform metaethical theories as we can make more progress in metaethical discourse if we consider them. I have tried to show how psychological evidence bears on a range of debates in metaethics and how this can be used to help philosophers make some progress in these debates. This has been challenging as psychological and metaethical literature do not always align in terms of concepts, ideas and language, therefore my thesis has been a careful consideration of both the psychological literature of mental health conditions as well as metaethical discourse.

I have provided arguments for my thesis in the following way. In Part 1, I have argued that a metaethical theory, Moral Hermeneutic Fictionalism, may be able to provide a plausible defence to the Autism Objection because of developments in autistic understanding of morality. In Part 2, I have proposed that certain cases of depression (which meet certain conditions) may be able to provide a counterexample to Weak Motivational Internalism. In Part 3, I have reasoned that an Obsessive compulsive disorder patient who is motivated to perform a compulsion may be able to provide a counterexample to the Humean Theory of Motivation. These sections provide case studies to support my hypothesis that psychological conditions should inform metaethical theories. This is because in my thesis, I have examined case studies in depth and found that these case studies result in one changing their understanding of metaethical theories. I have shown psychological theories have made a modest impact on metaethics and thus the framework of applying psychological conditions to metaethical theories should be taken forward. Since I have found that some psychological conditions do inform some metaethical theories, we can look at other areas of metaethical discourse where this could be the case. This is where further research should take place as psychological theories can, (and in my opinion, they should), impact metaethical theories (and can have a large impact on many metaethical theories, not just the metaethical theories, which I have particularly focussed on in the thesis).

In the Introduction, I urged the reader to consider some questions while studying my thesis. The first was how knowledge of the psychological conditions may enhance the metaethical discourse, and the second was whether the psychological evidence should be used to investigate metaethical understanding. The first question I can answer simply. One should consider psychological conditions when understanding metaethical discourse because, as this thesis has shown, these psychological conditions may alter the plausibility of the metaethical theories. Furthermore, I have shown that one can motivate the metaethical theories discussed in the thesis without the psychological theories, but as soon as one considers the psychological literature with the metaethical theories, this influences the debate.

However, there are restrictions on the evidence for the psychological conditions. In particular, for my thesis, the evidence for psychological conditions that I have presented has its limitations. I have tried to highlight their limitations and provide a defeasible claim towards the metaethical theories by suggesting that they *may* influence the metaethical theory in question, rather than a strong claim that they *will* influence the metaethical theories. This is for two reasons. The first is that the psychological evidence is inconclusive. Currently, all three psychological conditions discussed in this thesis have inconclusive evidence regarding the hypothesis that I present. Unfortunately, this is due to the nature of psychological debate and experimentation; the hypothesis can only provide indicative evidence towards one argument rather than providing any conclusive answers. Having said this, the psychological literature does lean towards a certain direction, and I have interpreted and presented their findings in my thesis, and I claim that this is useful for understanding metaethical discourse. Moreover, this means that my thesis will have to be adapted in light of psychological evidence arising in the future, but at present, my conclusion is presented with the evidence available.

Finally, I wish to leave the reader with a thought to ponder over. One knows that psychological disorders are becoming increasingly prevalent in our society; there is more recognition, more diagnoses and treatments available in society. Therefore, this implies that academic disciplines, like metaethics, should take account of mental health perspectives in their discussions, particularly if they add value to a discipline's understanding of the problem it discusses. My thesis shows how

impactful mental health perspectives can be on metaethical understanding. Moreover, it is simply more inclusive and more fruitful to include mental health conditions in academic discussions. Furthermore, this inclusivity point is particularly relevant in a discourse like metaethics because the discipline is looking at how society determines what our morality should amount to, and this means that metaethics has a responsibility to look at a diverse range of perspectives to decipher this important question.

Bibliography

Alexander, B., Brewin, C.R., Vearnals, S., Wollf, G., Leff, J., 1999. An investigation of shame and guilt in a depressed sample. British Journal of Medical Psychology 72, 323–338.

Anscombe, G.E.M., 1957. Intention, 2nd ed. Harvard University Press, Cambridge, Massachusetts. London, England.

APA, 2013. Diagnostic and statistical manual of mental disorders : DSM-5., 5th ed. ed, DSM-5. American Psychiatric Association, Arlington, Va.

Armour-Garb, B., Woodbridge, J.A., 2015. Pretense and Pathology: Philosophical Fictionalism and its Applications. Cambridge University Press, Cambridge.

Ashcroft, G.A., Blackburn, I.M., Cundall, R.L., 1978. Affective disorders, in: Companion To Psychiatric Studies. Churchill Livingstone, Edinburgh, pp. 407–437.

Audi, R., 1998. Moral Knowledge and Ethical Character. Oxford University Press, New York.

Ayer, A.J., 2001. Language, Truth and Logic. Penguin Modern Classics, St. Ives.

Baron-Cohen, S., 1997. Mindblindness : an Essay on Autism and Theory of Mind. MIT Press, Cambridge, Mass.

Baron-Cohen, S., 1989. The Autistic Child's Theory of Mind: a Case of Specific Developmental Delay. Journal of Child Psychology and Psychiatry, and Allied Disciplines 30, 285–297.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a "theory of mind" ? Cognition 21, 37–46.

Beltman, M.W., Voshaar, R.C.O., Speckens, A.E., 2010. Cognitive–behavioural therapy for depression in people with a somatic disease: Meta-analysis of randomised controlled trials. The British journal of psychiatry 197, 11–19.

Berrios, G.E., Bulbena, A., Bakshi, N., Dening, T.R., Jenaway, A., Markar, H., Martin-Santos, R., Mitchell, S.L., 1992. Feelings of Guilt in Major Depression: Conceptual and Psychometric Aspects. British Journal of Psychiatry 160, 781–787.

Björklund, F., Björnsson, G., Eriksson, J., Olinder, R.F., Strandberg, C., 2012. Recent Work on Motivational Internalism. Analysis 72, 124–137.

Bjornsson, Olinder, Eriksson, Bjorklund, 2015. Motivational Internalism: Contemporary Debates, in: Motivational Internalism. Oxford University Press, pp. 1–20.

Blackburn, S., 1998. Ruling Passions: A Theory of Practical Reason. Oxford University Press, Oxford.

Blackburn, S., 1984. Spreading the Word: Groundings in the Philosophy of Language. Clarendon Press, Oxford.

Blair, R., 1995. A cognitive developmental approach to morality: investigating the psychopath. Cognition 57, 1–29.

Blair, R.J.R., 2005. Applying a cognitive neuroscience perspective to the disorder of psychopathy. Development and Psychopathology 17, 865–891.

Brakoulias, V., Starcevic, V., 2010. The Characterization of Beliefs in Obsessive-Compulsive Disorder. Psychiatr Q 82, 151–161.

Bream, V., Challacombe, F., Palmer, A., Salkovskis, P., 2017. Cognitive Behaviour Therapy for Obsessive-Compulsive Disorder. Oxford University Press.

Brink, D., 1989. Moral Realism and the Foundations of Ethics. Cambridge University Press, Cambridge.
Brink, D.O., 1986. Externalist Moral Realism. The Southern Journal of Philosophy 24, 23–41.
Brock, H., Rizvi, A., Hany, M., 2024. Obsessive-Compulsive-Disorder [WWW Document]. URL https://www.ncbi.nlm.nih.gov/books/NBK553162/ (accessed 7.8.24).
Brock, S., 2014. The Phenomenological Objection to Fictionalism. Philosophy and Phenomenological Research 88, 574–592.
Bromwich, D., 2010. Clearing Conceptual Space for Cognitivist Motivational Internalism. Philosophical Studies 148, 343–367.
Buhr, K., Dugas, M.J., 2002. The Intolerance of Uncertainty Scale: Psychometric properties of the English version. Behaviour Research and Therapy 40, 931–946.
Burgess, A.G., Burgess, J.P., 2011. Truth. Princeton University Press.
Campbell, R.J., 1996. Psychiatric Dictionary, 7th ed. Oxford University Press, New York.
Challacombe, F., 2011. Break free from OCD / Fiona Challacombe, Victoria Bream Oldfield and Paul Salkovskis. Vermilion, London.
Cholbi, M., 2011. Depression, Listlessness, and Moral Motivation. Ratio 24, 28–45.
Cholbi, M., 2007. Belief Attribution and the Falsification of Motive. Philosophical Psychology 19, 607–616.
Cholbi, M., 2006. Moral belief attribution: a reply to Roskies. Philosophical Psychology 19, 629–638.
Chrisman, M., 2007. A Dilemma for Moral Fictionalism. Philosophical Books 49, 4–13.
Copp, D., 1995. Moral obligation and moral motivation. Canadian Journal of Philosophy Supplementary 21, 187–219.
Daly, C., Liggins, D., 2010. In Defence of Error Theory. Philosophical Studies 149, 209–30.
Dancy, J., 1993. Moral reasons / Jonathan Dancy. Blackwell Publishers, Oxford.
Darwall, S., 1997. Reasons, Motives, and the Demands of Morality: An Introduction, in: Darwall, S.L. (Ed.), Moral Discourse and Practice: Some Philosophical Approaches. Oxford University Press, pp. 305–312.
Darwall, S., 1995. The British Moralists and the Internal “Ought”: 1640–1740. Cambridge University Press, Cambridge.
Darwall, S., 1983. Impartial Reason. Cornell University Press, Ithaca.
Davey, G., 2021. Psychopathology Research, Assessment, and Treatment in Clinical Psychology, Third Edition. ed. Wiley Blackwell.
Davidson, D., 1969. How is Weakness of the Will possible?, in: Moral Concepts. Oxford University Press, London, pp. 93–113.
Day, R.K., 1999. Psychomotor agitation: poorly defined and badly measured. Journal of Affective Disorders 55, 89–98.
Dorr, C., Rosen, G., 2002. Composition as Fiction, in: The Blackwell Guide to Metaphysics. Blackwell, Oxford.
Douskos, C., 2018. Learning, Acquired Dispositions and the Humean Theory of Motivation. Philosophical Papers 47, 199–233.
Dreier, J., 2000. Dispositions and Fetishes: Externalist Models of Moral Motivation. Philosophy and Phenomenological Research 61, 619–638.
Dreier, J., 1990. Internalism and Speaker Relativism. Ethics 101, 6–26.

Durkheim, E., 1953. Individual and Collective Representations, in: Adorno, T.W., Durkheim, E. (Eds.), Pocock, D.F. (Tran.), Sociology and Philosophy. Cohen and West, London.
Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H.R., Wolf, O.T., Convit, A., 2008. Dissociation of Cognitive and Emotional Empathy in Adults with Asperger Syndrome Using the Multifaceted Empathy Test (MET). Journal of autism and developmental disorders 38, 464–473.
Eklund, M., 2024. Fictionalism. The Stanford Encyclopaedia of Philosophy (Winter 2019 Edition).
Eklund, M., 2009. The Frege–Geach Problem and Kalderon's Moral Fictionalism. The Philosophical Quarterly 59, 705–712.
Fatima, M., Babu, N., 2022. Cognitive and Affective Empathy in Autism Spectrum Disorders: A Meta-analysis. Review Journal of Autism and Developmental Disorders.
Fernandez, S., Daffern, M., Moulding, R., Nedeljkovic, M., 2024. A critical comparison of aggressive intrusive thoughts in obsessive compulsive disorder and aggressive scripts in offender populations. Aggression and Violent Behaviour 76, 1–11.
Ferrando, C., Selai, C., 2021. A systematic review and meta-analysis on the effectiveness of exposure and response prevention therapy in the treatment of Obsessive-Compulsive Disorder. Journal of obsessive-compulsive and related disorders 31, 1–16.
Fileva, I., 2021. A State of Besire. Philosophia 49, 1973–1979.
Finlay, S., 2007. Four Faces of Moral Realism. Philosophy Compass 2, 820–849.
Fisher, A., 2011. Metaethics: An Introduction. Routledge, Oxon.
FitzPatrick, W., 2009. Recent Work on Ethical Realism. Analysis 69, 746–760.
Fontaine, J., Luyten, P., De Boeck, P., Corveleyn, J., 2001. The Test of Self-Concious Affect: Internal structure, differential scales and relationships with long-term affects. European Journal of Personality 15, 449–463.
Frith, U., 1989a. Autism: Explaining the enigma. Basil Blackwell, Oxford, England.
Frith, U., 1989b. Autism and "Theory of Mind," in: Gillberg, C. (Ed.), Diagnosis and Treatment of Autism. Springer, New York, NY, US, pp. 33–52.
Garrard, E., McNaughton, D., 1998. Mapping Moral Motivation. Ethical Theory and Moral Practice 1, 45–59.
Geddes, J., Price, J., McKnight, R., 2012. Psychiatry, 4th ed. Oxford University Press, Oxford.
Gendler, T.S., 2008. Alief and Belief. The Journal of Philosophy 105, 634–663.
Georgiou, G., Demetriou, C., Fanti, K., 2019. Distinct Empathy Profiles in Callous Unemotional and Autistic Traits: Investigating Unique and Interactive Associations with Affective and Cognitrive Empathy. Journal of Abnormal Child Psychology 1863–1873.
Ghatavi, K., Nicolson, R., MacDonald, C., Osher, S., Levitt, a, 2002. Defining guilt in depression: A comparison of subjects with major depression, chronic medical illness, and healthy controls. Journal of Affective disorders 68, 307–315.
Gibbard, A., 2002. Wise Choices, Apt Feelings: A Theory of Normative Judgment. Oxford: Clarendon Press.
Gillan, C., 2017. Habits and Goals in OCD, in: Pittenger, C. (Ed.), Obsessive-Compulsive Disorder - Phenomenology, Pathophysiology, and Treatment. Oxford University Press, New York, pp. 161–170.

Gilson, M., Freeman, A., Yates, M.J., Freeman, S.M., 2009. Overcoming Depression: A Cognitive Therapy Approach. Oxford University Pres, New York.
Goldman, A.H., 2005. Reason Internalism. Philosophy and Phenomenological Research 71, 505–532.
Goodman, W.K., Price, L.H., Rasmussen, S.A., Mazure, C., Fleischman, R.L., Hill, C.L., Heninger, G.R., Charney, D.S., 1989. The Yale-Brown Obsessive Compulsive Scale. The Archives of General Psychiatry 46, 1006–1011.
Green, M., 2021. Speech Acts [WWW Document]. The Stanford Encyclopedia of Philosophy. URL https://plato.stanford.edu/archives/fall2021/entries/speech-acts/ (accessed 5.8.24).
Greenspan, P.S., 1998. Moral responses and moral theory: Socially-based externalist ethics. The Journal of Ethics 2, 103–122.
Haddon, M., 2003. The Curious Incident of the Dog in the Night-time. Vintage, London.
Hale, B., 1995. Modal Fictionalism: A Simple Dilemma. Analysis 55, 63–67.
Happé, F., 1995. Understanding Minds and Metaphors: Insights from the Study of Figurative Language in Autism. Metaphor and Symbolic Activity 10, 275–295.
Happé, F., 1994. An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. J Autism Dev Disord 24, 129–154.
Hare, R.M., 1952a. The Language of Morals. Oxford, Clarendon Press.
Hare, R.M., 1952b. Meaning and Criteria, The Language of Morals. Oxford University Press, Oxford.
Harrison, P.J., 2010. Lecture notes. Psychiatry / Paul Harrison, John Geddes, Michael Sharpe., 10th ed. ed. Wiley-Blackwell, Oxford.
Hinckfuss, I., 1993. Suppositions, Presuppositions, and Ontology. Canadian Journal of Philosophy 23, 595–618.
Hirvelä, S., Helkama, K., 2011. Empathy, values, morality and Asperger's syndrome. Scandinavian Journal of Psychology 52, 560–572.
Hoffman, M.L., 2001. Empathy and Moral Development: Implications for Caring and Justice. Cambridge University Press.
Hoffman, M.L., 1987. The contribution of empathy to justice and moral judgment, in: Empathy and Its Development, Cambridge Studies in Social and Emotional Development. Cambridge University Press, New York, NY, US, pp. 47–80.
Holton, R., 2009. Willing, Wanting, Waiting. Clarendon Press, Oxford.
Hume, D., 1896. A Treatise of Human Nature by David Hume, reprinted from the Original Edition in three volumes and edited, with an analytical index. Clarendon Press, Oxford.
Hussain, N.J.Z., 2004. The Return of Moral Fictionalism. Philosophical Perspectives 18, 149–187.
Jacoby, R.J., Abramowitz, J.S., 2017. Intolerance of Uncertainty in OCD, in: Obsessive- Compulsive Disorder: Phenomenology, Pathophysiology, and Treatment. Oxford University Pres, New York, pp. 171–177.
Jacoby, R.J., Buck, B.E., Fabricant, L.E., 2014. How is the Beads Task related to intolerance of uncertainty in anxiety disorders? Journal of Anxiety Disorders 28, 495–503.

James, R., Blair, R., 1996. Brief Report: Morality in the Autistic Child. Journal of Autism and Developmental Disorders 26, 571–579.
Jarrett, R.B., Weissenburger, J.E., 1990. Guilt in Depressed Outpatients. Journal of Consulting and Clinical Psychology 58, 495–498.
Jarrold, C., Carruthers, P., Smith, P.K., Boucher, J., 1994. Pretend play: Is it metarepresentational? Mind & Language 9, 445–468.
Johnston, M., 1989. Dispositional Theories of Value. Proceedings of the Aristotelian Society Supplementary Volume, 139–74.
Jones, A.P., Gilbert, F., Happe, F.G.E., Burnett, S., Viding, E., 2010. Feeling, caring, knowing: different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. Child Psychol. Psychiatry 11, 1188–1197.
Joyce, R., 2017. Fictionalism in Metaethics, in: McPherson, T., Plunkett, D. (Eds.), The Routledge Handbook of Metaethics. Routledge, pp. 72–86.
Joyce, R., 2005a. Moral Fictionalism, in: Kalderon, M.E. (Ed.), Fictionalism in Metaphysics. Oxford University Press, pp. 287–313.
Joyce, R., 2005b. The Evolution of Morality. MIT Press.
Joyce, R., 2002. Expressivism and Motivational Internalism. Analysis 62, 336–44.
Joyce, R., 2001. The Myth of Morality [electronic resource] / Richard Joyce. Cambridge University Press, Cambridge ; New York.
Joyce, R., Kirchin, S., 2010. A World Without Values: Essays on John Mackie's Moral Error Theory. Springer, Dordrecht.
Kalderon, M.E., 2007. Moral Fictionalism, Lines of Thought. Oxford University Press, Oxford, New York.
Kant, I., 2011. Groundwork of the Metaphysics of Morals [electronic resource] : a German-English edition / Immanuel Kant. Cambridge : Cambridge University Press, Cambridge.
Karp, D., 1996. Speaking of Sadness: Depression, Disconnection and the Meaning of Illness. Oxford University Press, Oxford.
Kauppinen, A., 2017. Empathy and Moral Judgment, in: Maibom, H. (Ed.), The Routledge Handbook of Philosophy of Empathy. Taylor & Francis, Oxford, pp. 215–226.
Kelly, D., Stich, S., Haley, K.J., Eng, S.J., Fessler, D.M.T., 2007. Harm, Affect, and the Moral/Conventional Distinction. Mind & Language 22, 117–131.
Kennett, J., Fine, C., 2008. Could there be an empirical test for internalism?, in: Sinnott-Armstrong, W. (Ed.), Moral Psychology. The MIT Press, Cambridge, MA, pp. 217–226.
Kenrick, D.T., Sheets, V., 1993. Homicidal fantasies. Ethology and Sociobiology 14, 231–246.
Kim, S., 2014. A Defence of Semantic Pretence Hermeneutic Fictionalism Against the Autism Objection. Australasian Journal of Philosophy 92, 321–333.
Kim, S., 2013. Pretense and Autism. Journal of Philosophical Ideas 48, 157–173.
Kim, S., 2005. Modal Fictionalism and Analysis, in: Kalderon, M.E. (Ed.), Fictionalism in Metaphysics. Oxford University Press, p. 116.
Kim, S., Thibodeau, R., Jorgensen, R., 2011. Shame, Guilt, and Depressive Symptoms: A Meta-Analytic Review. Psychological Bulletin 137, 68–96.
Kriegel, U., 2012. Moral Motivation, Moral Phenomenology, And The Alief/Belief Distinction. Australasian Journal of Philosophy 90, 469–486.

Ladouceur, R., Talbot, F., Dugas, M.J., 1997. Behavioural Expressions of Intolerance of Uncertainty in Worry: Experimental Findings. Behaviour Modification 21, 355–371.
Layne, C., 1980. Motivational Deficit in Depression: People's expectations X outcomes' impacts. Journal of Clinical Psychology . 36, 647–652.
Lenman, J., 1999. The Externalist and the Amoralist. Philosophia 27, 441–457.
Leslie, A.M., 1987. Pretense and Representation: The origins of "theory of mind." Psychological Review 94, 412–426.
Leslie, A.M., Frith, U., 1988. Autistic children's understanding of seeing, knowing and believing. British Journal of Developmental Psychology 6, 315–324.
Leslie, A.M., Mallon, R., Dicorcia, J.A., 2006. Transgressors, victims, and cry babies: is basic moral judgment spared in autism? Social Neuroscience 1, 270–283.
Lewis, D., 1988. Desire as Belief. Mind 97, 323–332.
Liao, S., Gendler, T.S., 2011. Pretense and Imagination. WIREs Cognitive Science 2, 79–94.
Liggins, D., 2010. The Autism Objection to Pretence Theories. The Philosophical Quarterly 60, 764–782.
Lillehammer, H., 2004. Moral Error Theory. Proceedings of the Aristotelian Society 104, 93–109.
Lillehammer, H., 1997. Smith on Moral Fetishism. Analysis 57, 187–195.
Lin, S.-K., Tsai, C.-H., Li, H.-J., Huang, C.-Y., Chen, K.-L., 2017. Theory of mind predominantly associated with the quality, not quantity, of pretend play in children with autism spectrum disorder. Eur Child Adolesc Psychiatry 26, 1187–1196.
Lockwood, P.L., Bird, G., Bridge, M., Viding, E., 2013. Dissecting empathy: high levels of psychopathic and autistic traits are characterized by difficulties in different social information processing domains 7, 1–6.
Luyten, P., Fontaine, J., Corveleyn, J., 2002. Does the Test of Self-Conscious Affect (TOSCA) measure maladaptive aspects of guilt and adaptive aspects of shame? An empirical investigation. Personality and Individual Differences 33, 1373–1387.
Mabrito, R., 2013. On Some Recent Attempts To Resolve The Debate Between Internalists And Externalists. The Southern Journal of Philosophy 51, 179–205.
Mackie, J.L., 1990. Ethics Inventing Right and Wrong. Penguin, London.
Maibom, H.L., 2017. Affective empathy, in: The Routledge Handbook of Philosophy of Empathy. Routledge, Oxford.
Marras, A., Fineberg, N., Pallanti, S., 2016. Obsessive Compulsive and related disorders: comparing DSM-5 and ICD-11. CNS Spectrums 21, 324–333.
Maslen, C., 2006. Fictionalism in Metaphysics [WWW Document]. Notre Dame Philosophical Reviews. URL https://ndpr.nd.edu/reviews/fictionalism-in-metaphysics/#:~:text=One%20example%20of%20cognitivist%20fictionalism,fiction%20there%20are%20Fs. (accessed 12.9.23).
Mason, M., 2003. Contempt of a Moral Attitude. Ethics 113, 234–272.
Mazza, M., Pino, M.C., Mariano, M., Tempesta, D., Ferrara, M., Berardis, D., Masedu, F., Valenti, M., 2014. Affective and cognitive empathy in adolescents with autism spectrum disorder. Frontiers in Human Neuroscience 8, 1–6.

McDowell, J., 2005. Non-Cognitivism and Rule Following, in: Holtzman, S., Leich, C. (Eds.), Wittgenstein: To Follow a Rule. Taylor & Francis Group, Milton, United Kingdom, pp. 141–162.
McDowell, J., 1985. Values and Secondary Qualities, in: Honderich, T. (Ed.), Morality and Objectivity: A Tribute to J.L. Mackie, International Library of Philosophy. Routledge & Kegan Paul, London, pp. 110–129.
McDowell, J., 1979. Virtue and Reason. The Monist 62, 331–350.
McDowell, J., McFetridge, I.G., 1978. Are Moral Requirements Hypothetical Imperatives? Proceedings of the Aristotelian Society, Supplementary Volumes 52, 13–42.
McKenzie, K., Russell, A., Golm, D., Fairchild, G., 2022. Empathic Accuracy and Cognitive and Affective Empathy in Young Adults With and Without Autism Spectrum Disorder. J Autism Dev Disord 52, 2004–2018.
McNaughton, D., 1988. Moral Vision: An Introduction to Ethics. Basil Blackwell, Oxford.
Mele, A.R., 1996. Internalist Moral Cognitivism and Listlessness. Ethics 106, 727–753.
Mendels, J., 1970. Concepts of Depression, Approaches to Behaviour Pathology. John Wiley and Sons Inc., New York.
Milevski, V., 2017. Weakness of will and motivational internalism. Philosophical Psychology 30, 44–57.
Mill, J.S., 2014. Utilitarianism [electronic resource] / John Stuart Mill. Cambridge University Press, Cambridge.
Miller, A., 2021. Realism [WWW Document]. The Stanford Encyclopedia of Philosophy. URL <https://plato.stanford.edu/archives/win2021/entries/realism/>. (accessed 12.9.23).
Miller, A., 2013. Contemporary Metaethics: An Introduction, 2nd ed. Polity, Cambridge.
Miller, C.B., 2008. Motivational Internalism. Philos Stud 139, 233–255.
Milton, D., 2016. Disposable dispositions: reflections upon the work of Iris Marion Young in relation to the social oppression of autistic people. Disability and Society 31, 1403–1407.
Milton, D., 2012. On the ontological status of autism: the “double empathy problem.” Disability and Society 27, 883–887.
Milton, D., Gurbuz, E., Lopez, B., 2022. The “double empathy problem”: Ten years on. Autism 26, 1901–1093.
Mitchell, P., Sheppard, E., Cassidy, S., 2021. Autism and the double empathy problem: Implications for development and mental health. British Journal of Developmental Psychology 39, 1–18.
Moore, G.E., 1922. Principia Ethica. Cambridge University Press, London.
Nagel, T., 1970. The Possibility of Altruism. Clarendon Press, Oxford.
National Collaborating Centre For Mental Health, 2010. Depression In Adults With A Chronic Physical Health Problem: The Nice Guideline On Treatment And Management. The British Psychological Society and the Royal College of Psychiatrists, London.
National Institute For Health And Clinical Excellence, 2010. Depression: the Treatment and Management of Depression in Adults. The British Psychological Society, Leicester.

Nichols, S., 2002. Norms with feeling: towards a psychological account of moral judgment. Cognition 84, 221–236.
Nolan, D., Restall, G., West, C., 2005. Moral Fictionalism Versus the Rest. Australasian Journal of Philosophy 83, 307–330.
Olinder, R.F., 2012. Svavarsdottir's Burden. Philosophia 40, 577–589.
Olson, J., 2014. Moral Error Theory [electronic resource]: History, Critique, Defence, Oxford scholarship online. University Press, Oxford.
Orth, U., Berking, M., Burkhardt, S., 2006. Self-Conscious Emotions and Depression: Rumination Explains Why Shame But Not Guilt is Maladaptive. Personality and Social Psychology Bulletin 32, 1608–1619.
Owens, D.J., 2003. Does Belief Have an Aim? Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 115, 283–305.
Parfit, D., 1998. Reason and Motivation. The Aristotelian Society 71, 99–130.
Park, S., 2013. Against a Besire Theory of Moral Judgment. Organon F 20, 5–17.
Pendlebury, M., 2002. Ought judgments and motivation. American Philosophical Quarterly 39, 183–196.
Peterson, C., Schwartz, S.M., Seligman, M.E.P., 1981. Self-Blame and Depressive Symptoms. Journal of Personality and Social Psychology 41, 253–259.
Pettit, P., Smith, M., 1993. Practical Unreason. Mind 102, 53–79.
Platts, M., 1980. Moral Reality and the End of Desire, in: Platts, M. (Ed.), Reference, Truth, and Reality. Routledge and Kegan Paul.
Platts, M. de B., 1979. Ways of Meaning: An Introduction to a Philosophy of Language. Routledge and Kegan Paul, London.
Premack, D., Woodruff, G., 1978. Does the chimpanzee have a "theory of mind"? The Behavioural and Brain Sciences 4, 515–526.
Purdon, C., Chiang, B., 2016. Obsessive-compulsive disorder, in: Carr, A., McNulty, M. (Eds.), The Handbook of Adult Clinical Psychology: An Evidence Based Practice Approach. Taylor & Francis Group, Oxford, pp. 492–514.
Rachman, S., 2013. Anxiety. Taylor & Francis Group, London, United Kingdom.
Rachman, S., 1997. A Cognitive Theory of Obsessions. Behaviour Research and Therapy 35, 793–802.
Rachman, S., 1981. Part I. Unwanted intrusive cognitions. Behaviour Research and Therapy 3, 89–99.
Rachman, S., Silva, P. de, 2009. Obsessive-Compulsive Disorder, Fourth Edition. ed, The Facts Series. Oxford University Press, Oxford, New York.
Ratcliffe, M., 2018. The interpersonal structure of depression. Psychoanalytic Psychotherapy 32, 122–139.
Ratcliffe, M., 2015. Experiences of Depression: A Study in Phenomenology. Oxford University Press, Oxford.
Ratcliffe, M., 2014. The phenomenology of depression and the nature of empathy. Med Health Care and Philos 17, 269–280.
Read, H., 2019. A typology of empathy and its many moral forms. Philosophy Compass 14, e12623.
Reid, J.E., Laws, K.R., Drummond, L., Vismara, M., Grancini, B., Mpavaenda, D., Fineberg, N.A., 2021. Cognitive behavioural therapy with exposure and response prevention in the treatment of obsessive-compulsive disorder: A systematic review and meta-analysis of randomised controlled trials. Comprehensive Psychiatry 106, 152223–152223.

Reuman, L., Fabricant, L.E., Herring, B., Abramowitz, J.S., 2015. Uncertainty as an anxiety cue at high and low levels of threat. Journal of Behavior Therapy and Experimental Psychiatry 47, 111–119.
Reynolds, S., 2009. Making up the Truth. Pacific Philosophical Quarterly 90, 315–335.
Ridge, M., 2007. 3. Ecumenical Expressivism: The Best of Both Worlds?, in: Oxford Studies in Metaethics: Volume II. Oxford University Pres, Oxford, pp. 51–76.
Ridge, M., 2006. Sincerity and Expressivism. Philosophical Studies 487–510.
Rosen, G., 1990. Modal Fictionalism. Mind 99, 327–354.
Roskies, A., 2008. Internalism and the Evidence from Pathology, in: Sinnott-Armstrong, W. (Ed.), Moral Psychology. The MIT Press, Cambridge, MA, pp. 191–206.
Roskies, A., 2006. Patients With Ventromedial Frontal Damage Have Moral Beliefs. Philosophical Psychology 19, 617–627.
Rowa, K., Purdon, C., 2003. Why are certain intrusive thoughts more upsetting than others? Behavioural and Cognitive Psychotherapy 31, 1–11.
Rowe, 1978. The Experience of Depression', Chichester. John Wiley and Sons, Chichester.
Salkovskis, P.M., Campbell, P., 1994. Thought Suppression Induces Intrusion in Naturally Occurring Negative Intrusive Thoughts. Behaviour Research and Therapy 32, 1–8.
Salkovskis, P.M., Candida Richards, H., Forrester, E., 1995. The Relationship Between Obsessional Problems and Intrusive Thoughts. Behavioural and Cognitive Psychotherapy 281–299.
Salkovskis, P.M., Kobori, O., 2015. Reassuringly calm? Self-reported patterns of responses to reassurance seeking in obsessive compulsive disorder. Journal of behavior therapy and experimental psychiatry 49, 203–208.
Samson, A.C., Huber, O., Gross, J.J., 2012. Emotion regulation in Asperger's syndrome and High Functioning Autism. Emotion 12, 659–665.
Sayre-McCord, G., 1986. The Many Moral Realisms. The Southern Journal of Philosophy 24, 1–22.
Scanlon, T.M., 2000. Seligman. Harvard University Press, Cumberland, United States.
Schmidtz, D., 1994. Choosing Ends. The University of Chicago Press 104, 226–251.
Schroeder, M., 2008. What is the Frege-Geach Problem? Philosophy Compass 3, 703–720.
Schurman, J., 1894. The Conciousness of Moral Obligation. The Philosophical Review 3, 641–654.
Schwenck, C., Mergenthaler, J., Keller, K., Zech, J., Salehi, S., Taurines, R., 2012. Empathy in children with autism and conduct disorder: group -specific profiles and development aspects. J. Child Psychol. Psychiatry 53, 651–659.
Searle, J.R., 2001. Rationality in Action. MIT Press, Cambridge, MA.
Seligman, M.E.P., Railton, P., Baumeister, R.F., Sripada, C., 2016. Homo Prospectus. Oxford University Press, Incorporated, Oxford, United States.
Shafer-Landau, R., 2003. Moral Realism: A Defence. Oxford University Press, Oxford.
Shepski, L., 2008. The Vanishing Argument from Queerness. Australasian Journal of Philosophy 86, 371–387.

Shields, K., 2016. Amoralists, Inverted Commas, and the Puzzle of Moral Internalism: An Essay in Experimental Metaethics - ProQuest. University of Missouri-Columbia.
Shulman, C., Guberman, A., Shiling, N., Ballminger, N., 2012. Moral and Social Reasoning in Autism Spectrum Disorders. Journal Autism and Dev. Disord. 42, 1364–1376.
Sinclair, N., 2021. Practical Expressivism: a Metaethical Theory. Oxford University Pres, Oxford.
Sinclair, N., 2009. Recent Work in Expressivism. Analysis 69, 136–147.
Sinhababu, N., 2009. The Humean Theory of Motivation Reformulated and Defended. The Philosophical Review 118, 465–500.
Slote, M., 2007. The Ethics of Care and Empathy. Routledge.
Smith, M., 2008. The Truth About Internalism, in: Sinnott-Armstrong, W. (Ed.), Moral Psychology. The MIT Press, Cambridge, MA, pp. 207–216.
Smith, M., 1994. The Moral Problem, Philosophical theory. Blackwell, Oxford.
Smith, M., 1987. The Humean Theory of Motivation. Mind 96, 36–61.
Sobel, D., Copp, D., 2001. Against direction of fit accounts of belief and desire. Analysis 61, 44–53.
Sorley, W.R., 1935. Moral Values and the Idea of God. The Cambridge University Press, London.
Spaid, A., 2021. Reasons Internalism and the Problem of Depression. Utilitas 33, 1–16.
Spaulding, S., 2017. Cognitive Empathy, in: Maibom, H. (Ed.), The Routledge Handbook of Philosophy of Empathy. Routledge, Oxford, pp. 13–21.
Sreenivasan, G., 2020. Emotion and Virtue. Princeton University Press.
Stanley, J., 2001. Hermeneutic Fictionalism. Midwest Studies In Philosophy 25, 36–71.
Starcevic, V., Berle, D., Brakoulias, V., Sammut, P., Moses, K., Milicervic, D., Hannan, A., 2011. Functions of compulsions in obsession-compulsion disorder. Australian and New Zealand Journal of Psychiatry 45, 449–457.
Stein, D.J., Kogan, C.S., Atmaca, M., Fineberg, N.A., Fontenelle, L.F., Grant, J.E., Matsunaga, H., Reddy, Y.C.J., Simpson, H.B., Thomsen, P.H., Van den Heuvel, O.A., Veale, D., Woods, D.W., Reed, G.M., 2015. The classification of Obsessive- Compulsive and Related Disorders in ICD-11. Journal of Affective Disorders 663–674.
Steketee, G., 1999. Overcoming Obsessive-Compulsive Disorder: A Behavioral and Cognitive Protocol for the Treatment of OCD: Therapist Protocol. New Harbinger Publications, Oakland CA.
Stevenson, 1937. The Emotive Meaning of Ethical Terms Mind, 14–31.
Stevenson, C., 1944. Ethics and Language. New Haven, London.
Stocker, M., 1979. Desiring the Bad: An Essay in Moral Psychology. The Journal of Philosophy 76, 738–753.
Streumer, B., 2017. Unbelievable Errors: An Error Theory about All Normative Judgements. Oxford University Press, Oxford.
Stroud, B., 1981. Hume, Arguments of the Philosophers Series. Taylor & Francis Group, London.
Svavarsdottir, S., 1999. Moral Cognitivism and Motivation. The Philosophical Review 108, 161–219.

Swartzer, S., 2013. Appetitive besires and the fuss about fit. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 165, 975–988.
Tallant, J., 2013. Pretense, Mathematics, and Cognitive Neuroscience. The British Journal for the Philosophy of Science 64, 817–835.
Tangney, Wagner, Gramzow, 1992. Shamed into anger? The relation of shame and guilt to anger and self-reported aggression. Journal of Abnormal Psychology 101, 469–578.
Tenenbaum, S., 2006. Direction of Fit and Motivational Cognitivism, in: Oxford Studies in Metaethics. Oxford University Press, Oxford, pp. 235–264.
Tresan, J., 2009. The Challenge of Communal Internalism. Value Inquiry 43, 179.
Turiel, E., 1983. The Development of Social Knowledge: Morality and Convention. Cambridge University Press.
van Fraassen, B., 1980. The Scientific Image. Oxford University Pres, Oxford.
van Inwagen, 1990. Material Beings. Cornell University Press, Ithaca, New York.
van Roojen, M., 2015. Metaethics: A Contemporary Introduction. Taylor & Francis Group, London.
van Roojen, M., 2013. Internalism, Motivational, in: LaFollette, H. (Ed.), International Encyclopedia of Ethics. American Cancer Society, pp. 2693–2706.
van Roojen, M., 2002. Humean and Anti-Humean Internalism About Moral Judgements1. Philosophy and Phenomenological Research 65, 26–49.
van Roojen, M., 1995. Humean Motivation and Humean Rationality. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 79, 37–57.
van Straten, A., Geraedts, A., Verdonck-de Leeuw, I., Andersson, G., Cuijpers, P., 2010. Psychological treatment of depressive symptoms in patients with medical disorders: A meta-analysis. Journal of Psychosomatic Research 69, 23–32.
Veale, D., Krebs, G., Heyman, I., Salkovskis, P., 2009. Risk assessment and management in obsessive-compulsive disorder. Advances in Psychiatric Treatment 15, 332–343.
Veale, D., Wilson, R., 2005. Overcoming Obsessive Compulsive Disorder: A self-help guide using Cognitive Behavioral Techniques. Constable & Robinson.
Walters-Chapman, S.F., Price, S.J., Serovich, J.M., 1995. The Effect of Guilt in Divorce Adjustment. Journal of Divorce and Remarriage 22, 163–177.
Walton, K., 2000. Existence as Metaphor, in: Empty Names, Fiction and the Puzzles of Non-Existence. CSLI Publications, Stanford, pp. 69–94.
Walton, K., 1990. Mimesis as Make-Believe: On The Foundations Of The Representational Arts. Harvard University Press, Cambridge, Mass. ; London.
Walton, K., 1985. Fictional Entities, in: The Reasons of Art/L'Art a Ses Raisons. University of Ottawa Press, Ottawa, pp. 403–13.
Westby, C., 2022. Play in Children With Autism Spectrum Disorder. Word of Mouth 33, 6–9.
Williams, B., 1979. Internal and External Reasons, in: Harrison, R. (Ed.), Rational Action: Studies in Philosophical and Social Science. Cambridge University Press, Cambridge, pp. 17–24.
World Health Organization, 2018. The ICD-11Classification of Mental and Behavioral Disorders [WWW Document]. URL

https://faculty.ksu.edu.sa/sites/default/files/icd-_11_2018.pdf (accessed 3.20.24).
Wright, C., 1993. Truth and Objectivity, 1st ed. Harvard University Press, Cambridge.
Yablo, S., 2002. Abstract Objects: A Case Study. Philosophical Issues 220–240.
Yablo, S., 2001. Go Figure: A Path Through Fictionalism, in: Midwest Studies in Philosophy Volume XXV: Figurative Language. Blackwell, Oxford, pp. 72–102.
Yablo, S., 2000a. A Paradox of Existence, in: Empty Names, Fiction and the Puzzles of Non-Existence. CSLI Publications, Stanford, pp. 275–312.
Yablo, S., 2000b. Apriority and Existence, in: New Essays on the A Priori. Oxford University Press, Oxford, pp. 197–228.
Zaki, J., 2014. Empathy: A Motivated Account. Psychological Bulletin 140, 1608–1647.
Zangwill, N., 2015. Motivational Externalism: Formulation, Methodology, Rationality, and Indifference, in: Björnsson, G., Strandberg, C., Olinder, R.F., Eriksson, J., Björklund, F. (Eds.), Motivational Internalism. Oxford University Press, Oxford, pp. 45–60.
Zangwill, N., 2008. Besires and the Motivation Debate. Theoria 74, 50–59.
Zangwill, N., 1998. Direction of Fit and Normative Functionalism. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 91, 173–203.
Zarei, M., Golchin, M.D., Seyedi, M., Baghban, A.A., Daryabor, A., Mahmoudi, E., Mirzakhani, N., 2022. Correlation Between the Theory of Mind and Pretend Play in 5 to 7-Year-Old Children with Autism and Their Typically Developed Peers. Iran J Pediatr 32.
Zhang, X., 2021. Why de dicto desires are fetishistic. Ratio 34, 303–311.