Beliefs, Desires, & Hybrids

An Examination of the Content of Moral Judgements

Haley Brokensha

Bachelor of Behavioural Science (Honours)

Submitted in fulfillment of the degree of Doctor of Philosophy

26 April 2019

Flinders University, College of Humanities and Social Sciences,
Sturt Road, Bedford Park, South Australia, 5042.
haley.brokensha@flinders.edu.au

Contents

Abstract

Theories of moral judgement have traditionally implied that one should be either a cognitivist or a non-cognitivist about moral judgements. Each side offers unique advantages that do not come as easily for the other. In the attempt to try and capture the positive qualities of each side, several theories that are best described as hybrid theories have been developed. In this thesis I describe three significant variations of hybrid theories: *hybrid-state theory* claims that moral judgements express 'besires' which are hybrid-mental states with both cognitive and non-cognitive components; *hybrid-expressivism* claims that moral judgements express both a cognitive and a non-cognitive mental state; and a very new position, *moral thought pluralism* says that moral judgements can express more than one kind of moral thought. In this thesis I examine several central philosophical and empirical attempts to conceptualise moral judgements. I find that, at present, this evidence suggests that hybrid-state theory and hybrid-expressivism are implausible. Furthermore, there should be a presumption in favour of moral thought pluralism until such time as a more restrictive theory such as cognitivism or non-cognitivism is satisfactorily supported. If moral thought pluralism is correct, all attempts at restricting the theory will be unsuccessful.

Declaration

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Haley Brokensha

Acknowledgements

I would like to express my gratitude to my supervisors Lina Eriksson, Ian Ravenscroft, and Andrew Gleeson for their continuous support. Lina, thank you for your insightful feedback and for being such a wonderful role model. Your guidance has been invaluable to me and I am grateful for the opportunity to have had you as my supervisor. Ian, thank you for all of your advice and thoughtful observations. Your encouragement, kindness, and support means the world to me. I have been lucky to have such a wonderful mentor. Andrew, thank you for your detailed feedback and the time you have taken to critique my thesis. You have helped shape my perception in many ways including in the difficult early stages and for this I am forever grateful. Additionally, I would like to thank Greg O'Hair for his extreme generosity in mentoring my thesis. Your enormous depth of knowledge has shed light on the context of many central issues and I have thoroughly enjoyed our conversations. I am grateful for your perceptive examination of my work. Lastly, I would like to thank the Philosophy Department of Flinders University and my friends and family for their endless support during this challenging undertaking.

Introduction

This thesis investigates the nature of moral judgements. Historically, the most significant commitment one could make in this endeavour was to claim that moral judgements are either the expression of one's beliefs or the expression of one's desires. These positions, in this stark form, are known as cognitivism and non-cognitivism about moral judgements respectively. Each position has an inherent advantage that is the other's significant weakness. These advantages and disadvantages fall out of the commonly accepted distinction between belief-like mental states and desire-like mental states made by Elizabeth Anscombe. Beliefs, she claimed, have a mind-to-world direction-of-fit. They aim at representing the way the world is and can be true of false. Desires, by contrast, have a world-to-mind direction-of-fit and express our will or how we want the world to be. The implications of this for morality are significant if one holds that our moral judgements are just one kind of entity, beliefs (cognitive) or desires (non-cognitive). Significantly, our beliefs can be true or false, desires cannot. And, desires are intrinsically motivating for us, beliefs are not. The significance of this will be drawn out at a later stage. It is noteworthy that neither cognitive theorists, nor non-cognitive theorists have been content with the limits imposed on their theories by the seemingly inherent nature of the mental states that they claim moral judgements express. Advocates of both views have come quite some way towards obtaining an explanation for what their respective theories seem at face value to neglect. In overcoming such challenges, several kinds of hybrid theories have been developed. These theories aim to retain the positive features and avoid the negative features of cognitive and non-cognitive theories.

Hybrid theories of moral judgement can take one of three forms. Firstly, there are *hybrid-state theories*. Hybrid-state theories deny the commonly accepted view that mental states can have only one direction-of-fit. The proposed advantages of hybrid-state theories is that they have the potential to explain how moral judgements can be both truth-apt and motivating. Their main challenge is to defend the existence of hybrid-mental states. These states must not be reducible

to constitutive cognitive and non-cognitive parts whereby one part can alter independently of the other. Secondly, *hybrid-expressivists* claim that moral judgements are those in which an individual has both a moral belief and a non-cognitive mental state of some kind in relation to a single situation, act, or character. There are multiple possible variations of this view which depend upon what kinds of cognitive and non-cognitive mental states are thought to be involved. This issue will be further discussed at a later stage. Like hybrid-state theorists, hybrid-expressivists have the ability to explain truth-aptness and motivation. Their proposal, however, sets the bar high for what is to be counted as a moral judgement. Their main challenge is to justify setting a high bar. Lastly, *moral thought pluralists* think that moral judgements do not need to be of one particular kind, i.e. either a cognitive, non-cognitive, or hybrid mental state. Moral judgements, on this view, can be either cognitive or non-cognitive.[1] Cognitive moral judgements are truth-apt but motivationally inert. Non-cognitive ones are not truth-apt but have motivational force. Moral thought pluralism faces problems of a unique kind; it owes an explanation of the commonalities between some cognitive and some non-cognitive mental states which is capable of uniting them as moral judgements. That is, the challenge, once one has an accepted conception of cognitive moral judgements and non-cognitive moral judgements, is to argue that there are no significant differences which warrant calling one kind a moral judgement and not the other.

Various philosophical and, more recently, psychological works have attempted to adjudicate on the theories outlined above. This thesis concentrates specifically on arguments which seek to vindicate theories by appealing to the motivational nature of moral judgements.[2] Much has been said about cognitive and non-cognitive theories of moral judgement. Far less has been said about hybrid theories which is where I will focus much of my attention. The thesis will proceed as follows: Chapter 1 is largely preliminary. I outline key terminology used in discussions about moral judgements, as well as the main meta-ethical conceptions of moral judgements. This framework provides the foundations for subsequent chapters. Chapters two through five analyse literature commonly used to support and criticise various meta-ethical theories over others. Chapter 2 discusses issues relating to the relationship

[1] One could include hybrid-states, but there is no apparent advantage to doing so given that one of the main challenges to hybrid-state theorists is to show such states exist.

[2] There is further research to be done in areas unrelated to motivation in relation to the general aims of this thesis. However, this research is beyond the scope of the current project.

between moral judgements and moral motivation. Specifically, issues about whether moral judgements are necessarily motivating in virtue of the kind of mental states that they express. I find that all theories of moral judgement are able to explain the seemingly motivational nature of moral judgements. The following three chapters focus on examining the current trend of using psychological research to weigh up theories of moral judgement. Chapter 3 sets the scene for an analysis of the empirical literature on moral judgements. It explores questions about what emotions are. This is necessary as most of the empirical literature on moral judgements extensively refers to emotions. I do not aim to defend any particular theory of emotions, but I do briefly justify my reliance on appraisal theories. In Chapters 4 and 5 I discuss the empirical literature on moral judgement with an appraisal theory of emotions in mind. I find that it is not capable of supporting cognitivism over non-cognitivism (or vice versa). It is, however, useful for facilitating discussions about the kinds of content cognitive and non-cognitive mental states ought to have to be considered moral judgements. Finally, in Chapter 6, I explore the possibility that endorsement of various kinds is essential to moral judgements. Using the notion of endorsement, I argue that there should be a presumption in favour of moral thought pluralism, the view that moral judgements can be either cognitive or non-cognitive.

There are a couple of different ways that one can go about answering the question of what it is to make a moral judgement. This point applies to any related questions or reformations of the question, such as, are moral judgements necessarily motivating for us? Do moral judgements express beliefs, that is, truth-evaluable statements, or do they express a kind of desire, or both? And, what is it that distinguishes a person's moral judgements from their non-moral judgements? For any of these questions only an answer of a particular kind will be satisfying for us. That is, when we ask these kinds of questions about moral judgements we are looking for a certain kind of definition. So, before going further, it will be useful to say a little about definitions. Below are three widely recognised and commonly employed kinds of definitions.

A *descriptive definition* aims to capture the common usage of a term. The meaning of the word is *determined* by its usage. Its discovery requires the careful analysis of stereotypical examples of the thing that we are trying to label (for more on this see Gupta, 2015). When we comment that some word had a particular meaning historically which differs from its meaning today we are engaging in a descriptive task. *Stipulative definitions*, on the other hand, are those in which a specific meaning is prescribed to a word. It occurs when we, for example, name

a new born baby. A stipulative definition need not track the ordinary usage of a term. In the context of moral judgements, if someone puts forth a stipulative definition, to know if a moral judgement has been made, all we need to know is whether the judgement in question conforms to the stipulated definition; we need to know how the judgement was reached. For example, someone may claim that moral judgements *must* be non-cognitive; if someone expresses a cognitive belief in answer to the question 'is x morally wrong?', then they have not expressed a moral judgement. Empirical investigation can reveal how regularly various stipulated definitions are reached. That is, we may find that, depending on our definition, moral judgements are made anywhere from pervasively to never. This kind of empirical analysis will likely influence our attraction to possible stipulated definitions.

In philosophy, conceptual analysis avoids the extreme tasks of stipulation and description. In particular, stipulation without reason is meaningless. In the case of defining moral judgements, we may end up with some stipulated conception that does not resemble any judgements that people actually do or could make. And, description, whereby we try to capture the common usage of terms is equally unsatisfying. We tend to think that it is possible for people to be in mass error in their application of moral concepts. Instead of prescribing or describing, our goal is to scrutinise existing concepts in an attempt to improve both our understanding and our application of the concepts. This kind of definition is known as an *explicative definition*. The goal is to identify the "essence" of concepts that we use. Such a task generally involves some degree of interplay between description and stipulation. For example, some concept is applied by many different people in some way that roughly overlaps. From this foundation, philosophers attempt to identify which possible features ought to be considered essential in the application of that concept. They aim to identify what it is about a pattern of features that is important. If someone claims that some feature is essential, they are saying that without that feature we cannot have an instance of the phenomenon in question. When features are chosen arbitrarily, this is an exercise in stipulation. On the other hand, if the stipulated essential feature(s) captures the significance of the pattern that we aim to identify, it is not arbitrary. We cannot conduct a poll to settle definitional matters in this case; we aim to comment on what should be the proper use of the concept and to pick out the pattern that is worthwhile naming. While it is likely that people will converge on their recognition of the essential features of concepts, it is not a given. One way to see why the task is important is to think of the case where the popular use of a term changes so significantly

from its original purpose that it starts to be used in a way that fails to capture what some people take to be the essence of its former usage. In this case people will either aim to re-emphasise the former meaning of the term or they will introduce a new term to represent the feature of interest.

Although stipulation and description can in theory be pursued independently of one another, in the creation of an explicative definition they are not. A definition of moral judgement which is too far removed from a descriptive one is likely to be rejected. Likewise, stipulation without justification will be rejected. It is of central importance that the stipulations involved in explicative definitions are not arbitrary. Those who are seeking an explicative definition will not be satisfied with any stipulations unless there is something special about certain features that make them more deserving of being called 'moral judgements'. It is sometimes difficult to keep track of the intent of authors to either describe, stipulate, or explicate in their discussions about moral judgements. It is not uncommon for authors to slip between tasks unwittingly. Consequently, it is important to keep this issue in mind throughout this work.

CHAPTER 1

Mental states and meta-ethical positions

There are several key terms that are regularly used in discussions about moral judgements. Among those most heavily used are: belief, desire, intuition, sentiment, emotion, cognitivism, non-cognitivism, rationalism, approval, disapproval, realism, and truth. Most of us will have an intuitive grasp of the meanings of these terms. And often, in discussions about moral judgements, the meanings of these key terms are not made explicit. Unfortunately, for the kinds of discussions philosophers want to have about moral judgements, it seems that the terms are not used by different authors in sufficiently similar ways, creating unnecessary barriers to productive discourse. This variation in meaning is to be expected given the vast interest and long history that the topic of moral judgements has incited. More recently, questions about the nature of moral judgements span largely isolated disciplines, creating even greater disparities in meaning. The failure to be explicit makes it extremely difficult to evaluate theories of moral judgement. In an attempt to facilitate discussion, this work begins with a preliminary chapter which provides a taxonomy of the common terms and positions that the vast majority of meta-ethical theories of moral judgement employ in one way or another. I will begin with a discussion of the kinds of mental states often claimed to be involved in moral judgements. My aim will be to capture standard usage of the terms as far as is possible. I will then distinguish realist and anti-realist conceptions of morality, and mind-dependent from mind-independent realism. Following that, I will describe the main meta-ethical theories of moral judgement. My final comments highlight connections between common terms, suggesting that some of them are reducible to others.

1.1. Mental states and directions-of-fit

One of the oldest and most influential points of contention about moral judgements places great emphasis on the state of mind that an individual is in (and expresses) when they make a moral judgement. The many varieties of mental states were originally broken down into two, or more controversially, three different kinds: 1) cognitive mental

states, 2) non-cognitive mental states, and 3) an irreducible state with both cognitive and non-cognitive properties, sometimes referred to as besires. The distinction between the first two of these states rests upon their 'direction-of-fit' (Anscombe, 1957; Rosati, 2016; Smith, 1994; Platts, 1979; Kirchin, 2012, pp. 12–13). Elizabeth Anscombe (1957) famously stated that our mental states can be divided into two broad categories: reason and the will. Mark Platts succinctly outlines Anscombe's position as follows:

> Anscombe, in her work on intention, has drawn a broad distinction between two kinds of mental states, factual belief [reason] being the prime exemplar of one kind and desire [the will] a prime exemplar of the other... The distinction is in terms of the direction of fit of mental states with the world. 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. 256–7).

In the passage above there are already two different pairs of terms being used to refer to cognitive and non-cognitive mental states: reason and beliefs referring to cognitive ones on the one hand, and desires and the will referring to non-cognitive ones on the other. It is worth spelling these ideas out a little more. The broadest division to be made between mental states is between cognitive and non-cognitive mental states. This division rests on the mental states' direction-of-fit.

Cognitive mental states. Cognitive mental states make up the class of mental states with a mind-to-world direction-of-fit, whereby the mind represents the world as being a particular kind of way. For example, beliefs, memories, dreams (whilst asleep), hallucinations, imaginings, and visual illusions are all cognitive mental states (Kirchin, 2012, pp. 12–13; Roojen, 2016). Those who think that moral judgements are cognitive claim that moral judgements are beliefs as opposed to any other kind of cognitive mental state. Those who embrace this view are cognitivists. Whilst all of the cognitive representations mentioned are truth-apt, i.e. they can be assessed by how accurately they represent the world,

beliefs are the class of cognitive mental states which the holder takes to resemble the way the world truly is.[1] This explains why in the literature 'belief' is often used interchangeably with 'cognitive'. To summarise, beliefs are a form of cognitive mental state with a mind-to-world direction-of-fit, whereby our mind attempts to represent the way the world really is. Stereotypically, our beliefs are moulded, and are revisable by what we learn about the world. And, importantly, they can represent correctly or incorrectly; they can be true or false (Roojen, 2016).

Non-cognitive mental states. Contrasting with our cognitive mental states, are our non-cognitive mental states which have a world-to-mind direction-of-fit. Non-cognitive mental states are those which express how we *want* the world to be (see Kirchin, 2012, pp. 100–101; Roojen, 2016). It is controversial which states should belong to this category. Spanning from less to more controversial examples are desires, likes, dislikes, sentiments, intuitions, and emotions. Significant controversy surrounds the possibility that some of these mental states necessarily involve a cognitive mental state in addition to the non-cognitive mental state (or mental state component). Non-cognitive mental states do not tell us about the way the world really is. For this reason, they are not thought to be assessable in terms of truth or falsity. For example, neither my desire for coffee, nor my fear of snakes can be true or false. The following scenario illustrates this point further:

> If a bus knocks someone off her bicycle and she expresses anger, we can ask whether she was really angry, but it makes little sense to ask whether the anger itself was true or false. This is because emotions do not describe the world as being a certain way and hence cannot describe the world accurately or inaccurately. (Fisher, 2011, p. 30)[2]

Many theorists think that moral judgements depend upon our non-cognitive mental states in some way. Those who strongly embrace this view are non-cognitivists. On this topic David Hume famously wrote:

> Take any action allow'd to be vicious: Wilful murder, for instance. Examine it in all lights, and see if you

[1] This includes beliefs about the past and about the future.

[2] Some authors claim that emotions such as the anger experienced in this example also represent and so are not a purely non-cognitive mental state. For the point being made here, the status of emotions is not important. It is merely noteworthy that there seems to be some non-cognitive mental state that is incapable of being true or false involved.

> can find that matter of fact, or real existence, which you call vice. In which-ever way you take it, you find only certain passions, motives, volitions and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflexion into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action. Here is a matter of fact; but 'tis the object of feeling, not of reason. It lies in yourself, not in the object. So that when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it. (1738–40, T3.1.1.25)[3]

According to Hume, non-cognitive mental states are essential to morality. Cognitive beliefs alone cannot tell us what we ought to do as is clearly expressed in the passage: "[r]eason is, and ought only to be the slave of the passions and can never pretend to any other office than to serve and obey them" (1738–40, T 2.3.3.4). This basic idea has permeated through time and disciplines. Modern day psychologist, Jonathan Haidt portrays the idea as follows:

> Reason can let us infer that a particular action will lead to the death of many innocent people, but unless we care about those people, unless we have some sentiment that values human life, reason alone cannot advise against taking the action. Hume argued that a person in full possession of reason yet lacking moral sentiment would have difficulty choosing any ends or goals to pursue... (2001, p. 816)

Once again, the presence of a multitude of undefined terms threatens to confuse the issue. As Platts (1979) noted, desires are the prime exemplar of non-cognitive mental states; as such, the term 'desire' is used by some as a catch-all phrase referring to any mental states which express the will, such as, emotions, wishes, sentiments, feelings, etc. Non-cognitive mental states are also sometimes referred to as conative mental states. Non-cognitivists disagree about which kind(s) of non-cognitive mental states are central to our moral judgements.

[3]The precise nature of the mental states involved in Hume's position stated here is not clear. It does seem clear, though, that he does not think moral judgements are beliefs.

In the coming chapters I will become more specific about which non-cognitive mental states various authors have in mind because the various non-cognitive mental states differ from each other in ways which may turn out to be significant for theories of moral judgement. Specifically, different non-cognitive mental states have differing abilities to rise to the challenges levelled at non-cognitivists. For now, the general distinction between cognitive and non-cognitive mental states will suffice.

Distinct entities. In virtue of their directions-of-fit, cognitive and non-cognitive mental states are widely taken to be fundamentally distinct entities (Rosati, 2016; Hume, 1738–40; Cohon, 2010; Smith, 1994). The Humean theory of motivation states that for an individual to be motivated to act, both kinds of mental states (e.g. beliefs and desires) must be in play. Beliefs, on Hume's view, are causally inert. Further to this, no belief entails a desire i.e. nothing about the way the world is can tell us how it ought to be (Hume, 1738–40, T3.1.1.27; Sayre-McCord, 2014; Cohon, 2010; Rosati, 2016). Michael Smith describes the relationship as follows:

> . . . what Humeans must deny and do deny is simply that agents who are in belief-like states and desire-like states are ever in a single, unitary, kind of state. This is the cash value of the Humean doctrine that belief and desire are distinct existences. And their argument for this claim is really quite simple. It is that it is always at least possible for agents who are in some particular belief-like state not to be in some particular desire-like state; that the two can always be pulled apart, at least modally. This, according to Humeans, is why they are distinct existences. (1994, p. 119)

According to this view the coming together of *distinct* cognitive and non-cognitive mental states is what drives directed action. Although not without its challengers, the Humean theory of motivation is a dominant and widely held position (see Rosati, 2016; Smith, 1994; Kirchin, 2012, p. 161; Miller, 2003, p. 271). It has important implications for any moral theory which claims that moral judgements are the result of either (and only) cognitive or non-cognitive mental states. These implications will be discussed as the thesis progresses. For now, however, the main ideas to remember are the distinction between the directions-of-fit of mental states and the proposed distinctiveness of these mental states in that one does not entail the other.

Hybrid mental states: besires. The division of mental states into cognitive and non-cognitive kinds has been criticised as not doing

justice to the full range of mental states that we possess. It is argued by some that there exists a third kind of mental state: the state of being motivated and representing the world at one and the same time, namely, besires (see Altham, 1986; Kirchin, 2012, p. 160; Rosati, 2016). This state is proposed to have both directions-of-fit. One can think of besires as having properties in common with beliefs and desires, but, importantly, they are not merely the sum of a belief and a desire. These can be thought of as beliefs (cognitive mental states) which have motivational force, or, as desires (non-cognitive mental states) which also represent how things are, were, or could be. These states *must* be unitary, non-decomposable ones (Kirchin, 2012, p. 166). To adopt the notion of a besire is to reject the Humean story. As Smith argues, anti-Humeans are committed to claiming that "...it is impossible for agents who are in a belief-like state to the effect that their ϕ-ing is right not to be in a desire-like state to the effect that they ϕ; that the two cannot be pulled apart, not even modally" (Smith, 1994, pp. 119–20). Uriah Kriegel succinctly depicts the essence of a besire as follows:

> ...the ban is not on any propositional attitude with a cognitive component and a conative [non-cognitive] component that can be denoted with a simple term. For example, being disappointed that p involves believing that p and desiring that q (where $q = \sim p$), but disappointment is not a besire. What distinguishes disappointment from besire is precisely the modal separability of its components, which casts it as a mere sum of two elements rather than an organic whole. Essentially, a disappointment involves not only two different contents, but at bottom also two different attitudes, whereas a besire involves a single, non-decomposable attitude. (2012, p. 472)

The appeal of positing such a state ought to become apparent in the following chapter on moral motivation. The term 'besire' captures the dual direction-of-fit in the mental state, with the 'b' signifying a belief quality and the remainder of the word representing the desire quality. More recently, a newer term 'alief' has been used to describe the same kind of mental state with the 'a' representing an affective component in a belief-like state (Gendler, 2008).[4] Emotions and intuitions are examples of other mental phenomena which have been (or could be) conceptualised as potential mental states of this kind (e.g. Prinz, 2004).

[4]Not all who use the term 'alief' think of them as besires. For some, an alief is just a robust connection of some sort between two distinct kinds of mental state.

The existence of this third kind of state with both directions-of-fit is controversial. (For a detailed discussion of the problems with besires see Rosati, 2016; Mackie's Argument from Queerness in Miller, 2003, p. 111; Sayre-McCord, 1988, pp. 111, 116; Smith, 1994). Anscombe's distinction between the direction-of-fit of mental states and Hume's theory of motivation are very intuitively appealing. The positing of hybrid mental states complicates the story and should be avoided if it is possible to explain what is going on without it. As I will explain, mental states proposed to have both directions-of-fit are employed in support of a particular kind of hybrid account of morality: hybrid-state theory.

Much of the conflict between modern meta-ethical positions centres on the various conceptions of the roles of cognitive, non-cognitive, and hybrid mental states in moral judgement, as will become apparent in Section 1.3.

1.2. Realism and anti-realism

One of the most fundamental divides in ethical thinking is between moral realists, who think that there is moral truth (that moral properties exist), and moral anti-realists, who think that there is no moral truth (that moral properties do not exist) (Joyce, 2015; Sayre-McCord, 2014). As James Dreier notes, 'in the good old days' realism and anti-realism were easily distinguishable: realism went hand in hand with cognitivism, and anti-realism went hand in hand with non-cognitivism (2004; also see Sayre-McCord, 2014; Kirchin, 2012, p. 99). There are intuitively appealing reasons to associate realism and anti-realism with mental states in this way. Realists think that there are moral truths and are inclined to think that our moral judgements express our beliefs about what those truths are. Conversely, moral anti-realists think that there are no moral truths and are inclined to think that our moral judgements express our views about how we want the world to be. However, there is no longer any straightforward correspondence between one's stance on moral truth and one's stance on the mental states that are expressed when one makes a moral judgement (Dreier, 2004; Kirchin, 2012, p. 99; Miller, 2003). I begin with a brief description of the most obvious strengths and weaknesses of the straightforward correspondence before describing the complex relationships that have evolved from them.

The moral domain is sometimes said to be notoriously riddled with intractable moral disagreement that is dissimilar from the disagreement we see in other areas of inquiry such as science, mathematics, or history. The degree of disagreement has led many people to conclude that

moral disagreements are fundamentally irresolvable (Stevenson, 1944; Stevenson, 1963). This kind of disagreement, it is claimed, makes more sense if we give up the idea that people are expressing beliefs about moral properties when they make moral judgements. Instead we should embrace the view that moral judgements are expressions of non-cognitive mental states. Relatedly, realists have struggled to deny the claim that moral disagreements are irresolvable, notoriously facing difficulties in identifying moral properties or defining 'goodness' (Moore, 1903). Anti-realists think that the task is doomed to failure as there are no moral properties to identify (Sayre-McCord, 1988, p. 12; Sayre-McCord, 2014; Kirchin, 2012, p. 108).[5] The final major attraction of non-cognitivism (and so anti-realism) that I will mention is that moral judgements appear to be very intimately connected with moral motivation whereby a change in moral judgement reliably leads to a change in motivation. Non-cognitivism (traditionally an anti-realist position) provides a straightforward explanation of the connection as non-cognitive mental states are inherently motivating. Cognitive mental states, on the Humean view are not inherently motivating and so the connection remains to be accounted for (Smith, 1994; Hume, 1738–40; Rosati, 2016; Sayre-McCord, 1988, p. 12). The issue of motivation is the focus of Chapter 2. Cognitive (and so realist) positions do have the tools to account for these issues although they do so in a less straightforward manner. They do, however, have significant advantages of their own.

Despite the cognitivists challenge to identify moral properties and to explain the connection of moral judgements to motivation, cognitivists (and so realists) are seemingly better equipped with the ability to retain notions of moral truth and preserve many of our intuitions about the function and purpose of moral discourse (Kirchin, 2012, p. 20). It seems as though there are objective answers to our moral questions. We adamantly stand by our moral viewpoints as though there were some objective truth to them and we strongly oppose what we consider to be bad moral decisions. We recall times when we have made decisions which we no longer endorse (although we are craftily good at forgetting). We

[5]Further, if one is a realist who is not persuaded that moral properties are reducible to natural properties it is very difficult to explain supervenience. Supervenience is the intuitively appealing idea that "... if two things have exactly the same natural properties, then they also have exactly the same moral properties. If you find that two things have different moral properties, you must also find that they differ in some way in respect of their natural properties" (Miller, 2003, p. 31; see also McLaughlin and Bennett, 2014; Sayre-McCord, 1988). Natural reductive realists and non-cognitivists more intuitively explain supervenience, though there are problems for both.

also have a sense of moral progress and decline, and we fight for change with the apparent conviction that the change we fight for is morally superior. For cognitivists, these intuitions are easy to explain because there are right and wrong answers to our moral questions and when we think about morality we are trying to reach this truth.[6] Non-cognitivists (anti-realists) have a harder time accounting for these intuitions. For them, there is nothing about the way the world is to which we can appeal in order to claim that one person's moral view is superior to another's, or even to claim that one's current moral conviction is superior to one's past conviction. But of course, our realist intuitions could be an illusion. Non-cognitivists have developed some very sophisticated and compelling explanations of many of the aforementioned intuitions. In doing so, non-cognitivism brings into question its traditional anti-realist commitments, as will become apparent.

Moral realists are committed to the claim that there is moral truth and moral anti-realists, that there is no moral truth (Kirchin, 2012, p. 5). This picture is made complicated by disagreements about what one might mean by moral truth. Dreier discusses the complexities of the various commitments one might have in relation to this issue. He identifies two factors that he claims have led to a break down in distinguishing moral realism from moral anti-realism. Firstly, he notes that modern non-cognitivists have the goal of vindicating moral language insofar as is possible (Dreier, 2004, p. 25). Consequently, they have made concessions to their position which allow them to speak of moral properties and moral facts. Secondly, there is the problem of what he calls 'creeping minimalism' whereby many modern non-cognitivists have begun to adopt "'minimalist' theories of facts and truth" (Dreier, 2004, p. 25). These factors are at the heart of modern realism and anti-realism rifts.

Realists and anti-realists disagree about what it takes for something to be a moral fact in part because of their commitments to what they believe moral properties are. Here are two possible definitions of objectivity in relation to moral properties:

(1) "Something (a property or object say) is objective if its existence is independent of what anyone does or could believe, desire, be committed to, and the like" (Kirchin, 2012, p. 25).
(2) "Something (a certain subject matter and discourse such as ethics or aesthetics, say), is objective if it admits of correct and incorrect judgements, answers, and the like" (Kirchin, 2012, p. 25).

[6]Error theory being an exception.

Some examples of propositions generally thought to satisfy the former definition include mathematical and scientific propositions such as '1+1 = 2', 'water is H_2O', and 'the earth is flat'. All of these propositions are either true or false and have an answer that is independent of what anybody desires. Examples which would not satisfy the former, but that would satisfy the latter criterion for objectivity include: 'ripe tomatoes are red', 'fire is hot', 'grass is green', and 'chocolate is delicious'. According to definition (2) these claims can be true or false, but their truth or falsity depends upon our minds somehow. It is objectively true that ripe tomatoes look red *to me*. This is an objective fact about my subjective conscious experience. One which depends upon my mind in a way that water is H_2O does not. It is possible for us to make objective claims about the subjective experiences of our own minds, of the minds of others, and of minds in general. For example, it might be objectively true that ripe tomatoes look red to *all* humans, or that fire feels hot to *all* humans. These propositions (if they are true) are objective facts about the subjective experiences of human beings. They are mind-dependent truths. Properties such as taste, colour, smell, and sound are all mind-dependent properties, not existing in the world independently of our minds. The terms 'primary' and 'secondary' properties are often used to denote this difference in objective facts, whereby secondary properties are mind-dependent ones (Locke, (1690) 1975; McDowell cited in Sayre-McCord, 1988, p. 166).

The specific distinction between primary and secondary properties, although pointing towards something significant, does not directly identify what is meant by mind-independent and mind-dependent *moral properties*. Mind-dependence in the moral realm refers to dependence on a specific kind of mental phenomenon: mind-dependent moral properties are those which come to exist once some *non-cognitive* mental state has been accepted or appointed as a goal. For example, *given the goal* of promoting health, we can identify certain truths and falsities about what is good and bad. If the goal is rejected, so too are the claims about what is good and bad. To see the contrast, an attempt at stating a mind-independent truth may look something like 'consciousness is good, and would be good even if no mind ever had a non-cognitive world-to-mind attitude towards it'. To see this, imagine there were only two planets in the universe, neither of which has any conscious life. One has the potential to develop life and the other does not. One planet will be destroyed by an asteroid. There are no non-cognitive mental states to which truth about whether it is better that one planet survive over the other can be dependent. In this case, if one wishes to say that it is better that the planet incapable of forming life is destroyed, the

only basis for doing so is by appeal to some mind-independent property of goodness. The truth of the claim is akin to the truth of claims that, 'Earth is round', or 'grass appears green to human beings'. Notice that the second example refers to a secondary property (mind-dependent in the primary/secondary property distinction), yet it does not have the kind of mind-dependence discussed by non-cognitivists.

As noted, historically, non-cognitivists were anti-realists who denied that there is moral truth. Today, some anti-realists, sometimes referred to as quasi-realists, deny any kind of moral truth, but claim that there are moral 'truths' which we create, construct, or project onto the world (Blackburn, 1998). Here, in maintaining the anti-realist label, lies an insistence that moral truth, if it exists, must be mind-independent. Specifically, independent moral realism (IMR) maintains that moral truth exists independently of our non-cognitive attitudes which are not ultimately truth-apt. For independent moral realists "... [t]he existence of moral properties and moral reasons is a mind-independent matter. That is, the existence of such properties and reasons is not dependent on what human beings, either individually or collectively, think, desire, are committed to, wish for etc." (Kirchin, 2012, p. 22). Quasi-realists maintain this standard of truth, yet still speak of moral 'truth'. Dependent moral realists, by contrast, find ways of grounding truth in some kind of non-cognitive mental state (see Kirchin, 2012, p. 22 and Chapter 6; Miller, 2003, Chapter 7; Sayre-McCord, 1988, p. 15). It seems that the difference between the quasi-realist and the dependent moral realist lies in their willingness or unwillingness to adopt a notion of truth about moral properties which takes a non-cognitive mental state as a foundation. However, in the end, they are speaking of the same phenomenon despite their terminological commitments. This is the worry voiced by Dreier (2004) in what he calls 'the threat of creeping minimalism'.

The introduction of mind-dependent truth raises some very serious problems, many of which are avoided by independent moral realists. Dependent moral realists claim that moral truth and moral facts are born from some kind of non-cognitive mental state or perhaps some state which is both cognitive and non-cognitive at the same time (a besire). This line of thought leads quite obviously to the following questions concerning moral relativism: if moral truth derives from human non-cognitive attitudes and those attitudes differ (as they seem to), how can a non-cognitive attitude come to provide a standard of moral truth? Can we claim that there are some attitudes that are superior to others? Whose perception of moral circumstances matter? Addressing these worries is any defender of dependent moral realism's biggest challenge

and a number of very different answers have been proposed. Truth has generally been anchored to the non-cognitive responses of some individual (perhaps an idealised individual), or to the non-cognitive attitudes of some group of individuals, such as a culture. Below are three influential attempts at providing a standard of moral truth and their main problems. The problem of moral relativism can be more or less intense depending on the proposed standard of truth and the similarities in non-cognitive mental states between the minds that the truth is dependent upon. All of the proposed answers to this question have problems.

Subjectivists claim that the individual provides the standard of truth; that my judgements are true for me and yours, true for you. There is no objective truth beyond this. This can be likened to our claims about aesthetics or tastes; for example, it is true for me that sun-dried tomatoes taste good. This move in the moral domain seems unsatisfactory to many. On this view it can be true for me that helping sick people is good, and equally true of somebody else that hurting sick people is good. This is not what we tend to mean when we say that there are moral truths. We tend to think that truth in the moral domain is more substantial than this. A famous problem with this view is that it cannot explain moral disagreement. If we are both making a claim about ourselves and we are both right then there can be no disagreement between us.

Variations of ideal-observer theories where some idealised individual's judgements set the standard of truth overcome many of the problems that the subjectivist faces. Ideal observer(s) are usually thought to have the qualities of being perfectly knowledgeable, rational, and impartial. The appeal to an ideal observer avoids the problem of relativism, but it faces problems of another kind. For example, what qualities exactly is one required to have in order to be considered an ideal observer and on what grounds can one claim that the qualities of an ideal observer are important (Jollimore, 2017)? Furthermore, if our moral judgements are judgements about what we believe a perfectly rational, knowledgeable, and impartial spectator would want for us, why should we be motivated by such judgements? Moves towards idealisation or away from the judgements of actual individuals gives substance to our moral disagreements: disagreements are no longer about personal preferences. However, such moves lead us back to the problem of moral judgements being disconnected from motivation such that the

non-cognitivist position looses its natural advantage of explaining moral motivation (Kauppinen, 2016).[7]

A third hugely influential view says that standards of truth can be grounded in the non-cognitive attitudes of groups of people. The group might be all of human kind, or, particularly influential, is the idea that culture provides a standard of truth. An obvious problem with the first suggestion is that humans appear to widely disagree about moral issues. This leads us back to the same problems mentioned above: is whatever a human thinks is correct correct? In which case there is no moral disagreement. Or, do we prioritise some individual(s) judgements? In which case we are led back to the problems created by appealing to ideal observers and missing motivation. The idea that culture provides the standard of moral truth faces equally challenging problems. On this view, if a culture thinks that a particular action is right, then that action is right for them. This move leaves open the possibility of widespread relativisation of moral truth. If one culture of people share the view that murder is wrong, it is true for them that murder is wrong; if another culture thinks that murder is not wrong, it is true for them that murder is not wrong. A culture's view on the permissibility of murder does not apply to cultures who do not share their moral views. This can be compared to secondary property standards of truth in general, for example, if a group of people perceive bananas as red, it is true, for them, that bananas are red. If this view is correct, then there is no objective position from which to criticise another groups value system, providing it is internally consistent. And, without consensus in non-cognitive attitudes we are left with relativism. Another problem is that the moral convictions within any group, including cultures, are seldom unanimous. It is possible that such worries about variation in judgements within a group can be overcome by claiming that the view of the majority sets the standard of truth. That is, if most people within a group think that x is wrong, then it is true for them that x is wrong. But we tend to think that there are occasions where the majority can be wrong in their moral convictions. There are further problems concerning the issue that group membership is not easily identifiable; people generally see themselves as belonging to many groups. What is special about culture? Why shouldn't one's membership to a religion, friendship circle, or psychological kind (such as psychopaths) be classed as a legitimate group?

[7]See Smith, 1994 for a position which claims to avoid this.

To summarise, I wish to re-emphasise the important possible source of confusion which may result from failing to distinguish between *considering* mental states (non-cognitive ones or conscious sensations) when making moral decisions from *using* non-cognitive mental states to make moral decisions. In other words, as part of one's moral judgement, one may *consider* mental states in determining, for example, the degree of wrongness of an action. They may, for example, appeal to states of consciousness such as pain or suffering. This is not the same as claiming that a moral judgement *consists*, at least in part, of a non-cognitive moral attitude *of their own*, such as a desire. Moral judgements are mind-dependent only in the latter case. This is what it means for moral judgements to be non-cognitive. For dependent moral realists, truth and justification can be found by, for example, appealing to people who share moral frameworks or by reference to an ideal observer. In this thesis I will at times refer to the ability of various theories to address, identify, or explain moral truth. I do not commit myself to any particular view on this matter. So, for now I wish to make no point other than to highlight that realists, although they all agree that there are moral truths, disagree on the very important and contentious question of whether moral properties are mind-dependent or mind-independent. Definition (2) of objectivity says that something can be true or false even if it depends on our non-cognitive attitudes in an important way, definition (1) rejects this (for overviews of this distinction see Miller, 2003, p. 4; Prinz, 2008, Chapter 1; also see the 'Moral Objectivity and Moral Relativism' supplementation in Joyce, 2015). Both independent and dependent moral realism have unique strengths and weaknesses, some of which have already been touched upon, and will come up again throughout the thesis. However, this basic distinction along with the basic dichotomy of mental states in Section 1.1 will be adequate for the purpose of outlining the major meta-ethical conceptions of moral judgements in the following section.

1.3. Meta-ethical positions

In this section I will set out five possible meta-ethical conceptions of moral judgement in both the realist and anti-realist camps and briefly describe the main challenges faced by each position. The positions are distinguished from one another centrally by their commitments to the kinds of mental states that they claim are involved in moral judgements.[8] The realist and anti-realist sections mirror one other. Not all of the positions within these camps have known adherents. I provide

[8]This set of meta-ethical positions is not exhaustive.

greater detail where the meta-ethical position is generally taken to be most plausible.

Moral anti-realist positions.

(1) Cognitivism. Moral judgements aim at reporting truth evaluable facts about moral properties (they are beliefs). This is a mistake, as there are no moral facts or properties. Consequently, all moral beliefs are false (Mackie, 1977; Miller, 2003, p. 112; Kirchin, 2012, p. 78; Joyce, 2015). This view is appropriately labelled Error Theory (Mackie, 1977; Miller, 2003, p. 111; Joyce, 2015). Key challenges for this view are to explain why we have such beliefs in the first place and why, if our moral judgements express beliefs, we are generally motivated to act by them.

(2) Non-cognitivism. Moral judgements express a non-cognitive attitude of some kind, such as a desire, emotion, intuition, or sentiment about how we want the world to be (Kirchin, 2012, p. 99; Joyce, 2015; Miller, 2003).[9] Whichever non-cognitive mental state(s) the non-cognitivist claims are essential to morality will have a world-to-mind direction-of-fit. Such non-cognitive states are not truth-apt; there are no moral facts or properties (Blackburn, 1998).[10] Non-cognitivists deny that moral judgements express beliefs. Beliefs, they say, can tell us facts about the world such as 'the person is cutting the dog's skin off'. 'The dog is still alive and appears to be in pain'. But, they claim that the moral judgement (the claim that the action is wrong) depends upon a non-cognitive mental state such as a desire which is not truth-apt. It is possible to remain indifferent to factual information if it has no relevance to our non-cognitive attitudes.

Typically, moral anti-realists have taken one of these first two stances when defining moral judgements. Non-cognitivism, however, remains the dominant anti-realist position. In addition to these two popular positions, anti-realists may also say that moral judgements express a hybrid mental state, or that they express more than one kind of mental state. However, these latter positions (described below) are unlikely to be held by an anti-realist as anti-realists deny that there is moral truth and the main attraction of holding any of the latter views is their potential for vindicating a sense of moral truth.

(3) Hybrid-state theory. Moral judgements express a mental state with both directions-of-fit: a besire. There are no moral facts or

[9]The non-cognitive status of some of these views is questionable and further discussed in the views and chapters that follow.

[10]As noted, some modern non-cognitivists adopt a quasi-realist position which allows for talk of moral 'truth'. This is not meant to be understood in a primary or secondary objectivity sense of truth.

properties. It is a cost to this view that it gives up a simple intuitive dichotomy of mental states. Adopting this view means you must reject the Humean theory of motivation or at least deny that it tells the full story about motivation (this problem is discussed in depth in Chapter 2). The motivation to posit besires is most appealing to the realist who sees it as a way to explain both moral truth and motivation in moral judgements (Again, this point will be taken up in Chapter 2).

(4) Hybrid-expressivism. Hybrid expressivism shares with hybrid-state theory the idea that moral judgements are simultaneously cognitive and non-cognitive. But, unlike a besire, the cognitive and non-cognitive elements of the moral judgement are separable. Both elements must be present for a judgement to be a moral one. Theorists may disagree about exactly what the content of the cognitive and non-cognitive mental states must be. For example, a moral judgement that x is wrong may express a cognitive belief that x is wrong or it may express a belief that some non-cognitive responses are merited; and the non-cognitive mental state may express some kind of desire that not-x (see the section entitled 'Hybrid Expressivism' in Kauppinen, 2016). These differences can lead to the development of significant variations in hybrid-expressivist positions. For example, perhaps a moral judgement consists of a non-cognitive mental state such as a 'like' or 'dislike' of ours which is then cognitively endorsed as being morally merited, or something similar. Or, a stronger position may claim that both the cognitive and non-cognitive elements must have moral content, and for this to be the case, the belief must explicitly entertain moral concepts such as x is 'bad' or 'wrong' and the non-cognitive mental state must be of a kind adequate for the label 'moral judgement', for example, perhaps moral non-cognitive mental states are those which we demand others share (and not mere likes or dislikes). The significance of the differences in these positions will come into focus at a later stage. For an anti-realist, there are no moral facts or properties to which any of the moral beliefs correspond. The moral beliefs are formed in error as was the case for the anti-realist cognitivist. Likewise, the non-cognitive judgements do not allow for the possibility of grounding moral truth.

(5) Moral thought pluralism (MTP). Like hybrid-expressivism, moral thought pluralism is the idea that moral judgements can be expressions of more than one kind of mental state. For example, moral judgements may be either cognitive or non-cognitive. It differs from hybrid-expressivism by claiming that only one kind of mental state needs to be present in a moral judgement. An anti-realist moral thought pluralist would claim that all moral beliefs are mistaken and non-cognitive

moral judgements cannot provide any possible grounding for moral truth.

In this overview we can plainly see divisions within the anti-realist camp about what kind of mental states various theories take moral judgements to express. All three hybrid positions find greater adherence in the realist camp that follows and will consequently be explored in greater depth there.

Moral realist positions.

(1) Cognitivism (IMR). Moral realists think that there are moral facts. Historically, cognitivism was the dominant realist position. It says that moral judgements are beliefs which aim at reporting truth evaluable facts about moral properties (Roojen, 2015). Moral beliefs represent the world as being a certain way and these beliefs are made true or false by the way the world is. According to independent moral realists there are objective facts about morality to be discovered and the truth of moral facts does not depend on our non-cognitive attitudes. Independent moral realists may think that human mental states need to be considered when identifying good and bad. They do not however, endorse the stronger claim that our beliefs, or non-cognitive attitudes such as desires, sentiments, emotions etc. play a role in *determining* moral truth. As discussed in the previous section, this difference has similarities to the distinction made between primary and secondary properties, though it should not be confused with it. Primary properties are those which do not depend upon minds, for example, the claim that Earth is the third planet from the Sun refers to primary properties, and secondary properties are those that do depend upon our minds, such as the property of grass looking green. The distinction between mind-dependent and mind-independent moral properties differs from the primary/secondary distinction because mind-dependent moral properties depend upon *non-cognitive* mental states. That is, dependent moral truths depend upon world-to-mind mental states. So, although the truth of the claim that grass is green depends in part upon how our minds perceive grass (it is a secondary property truth), it does not depend on any of our world-to-mind mental states. That is, it does not depend upon our will. Independent moral realists think that we can identify moral truths without referring to any non-cognitive mental states. Saying that something is bad is like saying that x object is a bicycle. For dependent moral realists, moral truth is dependent on non-cognitive mental states such as desires. Kirchin illustrates the independent moral realist position as follows:

> Imagine we have a deer dying in an earthquake in some place remote to humans. Imagine that humans had no part in causing the earthquake, they did not cause the deer to be there, and so on. Now imagine that for whatever reason we want to say that this is a bad thing, rather than just a thing that just happened. (I mean here just that it has some general value of being bad; it does not necessarily have to be morally bad.) We can now ask the question of what it is that grounds this badness. The deer's death is bad, but is it bad because of humans – if any were there to judge it, or us considering it now – judge it to be bad? Or, is it bad no matter what humans say or could say about it? If you take the latter option, you are saying that this badness is a mind-independent matter. This is clearly so because we have specified that human judgement plays no part in why the action is bad. (Kirchin, 2012, pp. 22–3)

There are two main problems which face the independent moral realist. Firstly, there are difficulties with identifying moral properties (Moore, 1903). Secondly, cognitivists have difficulty explaining the connection between moral judgements and being motivated to act on those judgements. According to the Humean theory of motivation, beliefs disconnected from non-cognitive mental states are not motivating. On this view it is possible for an individual to judge, for example, that murder is bad without being motivated to refrain from murder (the implications of this will be further discussed in Chapter 2).

(2) Non-cognitivism. According to non-cognitivists, moral judgements express a non-cognitive attitude of some kind such as a desire or sentiment (see Roojen, 2016; Kauppinen, 2016; Kauppinen, 2015; Kirchin, 2012, p. 99). We may form beliefs about, for example, the natural facts which are relevant to our non-cognitive states. But, the non-cognitive states (such as our desires) are the locus of our moral judgements. Non-cognitivists must formulate a view on exactly which of our non-cognitive mental states are and which are not *moral judgements*. As will become apparent, this particular task is central to many meta-ethical disagreements. The following passage by Mackie illustrates the significance of non-cognitive mental states to moral judgements for the non-cognitivist:

> Another way of trying to clarify this issue is to refer to moral reasoning or moral arguments. In practice, of

> course, such reasoning is seldom fully explicit: but let us suppose that we could make explicit the reasoning that supports some evaluative conclusion, where this conclusion has some action-guiding force that is not contingent upon desires or purposes or chosen ends. Then what I am saying is that somewhere in the input to this argument – perhaps in one or more of the premises, perhaps in some part of the form of the argument – there will be something which cannot be objectively validated – some premise which is not capable of being simply true... but is constituted by our choosing or deciding to think in a certain way. (Mackie, 1977, pp. 29–30)

Historically, non-cognitivism about moral judgements was held in conjunction with anti-realism. Non-cognitivists can, however, be realists. A realist who is a non-cognitivist may hold either a mind-dependent or a mind-independent conception of moral truth, as follows.

(2a) IMR. Non-cognitivists drawn to independent moral realism will maintain that non-cognitive moral judgements are not truth-apt. They may argue that there are moral facts or properties, but that when we make moral judgements we do not aim at identifying moral truth. Instead we express non-cognitive attitudes which are not truth evaluable. This view involves rejecting the tendency to anti-realism that seems to easily follow from the belief that moral judgements express non-cognitive mental states. Instead, one may maintain the view that there are, in fact, mind-independent answers to our moral questions, but that we do not attempt to identify them, as all of our actual moral judgements express non-cognitive mental states. This version of non-cognitivism draws a very strict line between what is real and what is not, holding that mind-dependent properties are not suitable as a foundation of moral truth. This view is the mirror image of an error theory which says that there are no moral truths although we do attempt to identify them. To reiterate, there are moral truths, but our moral claims do not attempt to describe them (we do not express truth-apt propositions).

(2b) DMR. Like non-cognitivism IMR, non-cognitive dependent moral realists think that moral judgements are an expression of a non-cognitive mental state. They differ in their view on the truth-aptness of non-cognitive mental states. For the non-cognitive dependent moral realist, non-cognitive attitudes are mind-dependently truth-apt. For example, just as it can be true *for me* that strawberries are red, it can be true *for me* that I desire well-being or no unwanted physical contact etc.

In this way, non-cognitive mental states are mind-dependently truth-apt. From here, one may claim that there are some mind-dependent mental states which provide us with a moral foundation upon which we can claim that certain actions are right or wrong. For example, if it is true that a certain group of minds all desire a particular thing (such as good health), then it is true *for them* that disease is bad. As discussed in Section 1.2, one of the biggest challenges to non-cognitive mind-dependent realism is the threat of relativism. If truth is mind-dependent, and minds differ, objectivity is threatened.

The three hybrid positions that follow aim to satisfy our realist intuitions whilst retaining some of the advantages that non-cognitivism seems to have in explaining moral motivation. The first two (hybrid-state theory and hybrid-expressivism) take non-cognitive mental states or mental states with a non-cognitive component to be essential to moral judgements. The third (moral thought pluralism) takes non-cognitive judgements to be sufficient for moral judgements, but it does not take them to be necessary. Common to all of the realist positions presented, except cognitivism, is a commitment to the claim that non-cognitive mental states feature causally in moral judgements.

(3) Hybrid-state theory. Hybrid-state theorists claim that moral judgements express besires. Besires are states of mind with both directions-of-fit; they are at once representational of how the world is and of how one wants it to be. This view rejects the Anscombean dichotomy of cognitive and non-cognitive mental states. Besires are not decomposable into cognitive and non-cognitive components (Smith, 1994; Kirchin, 2012). If they were decomposable, we should be able to see that the moral judgement is confined to either the cognitive or the non-cognitive part (Smith, 1994). As mentioned above, realists claim that there are moral facts or properties. Advocates of hybrid-state theory are dependent moral realists. Two significant challenges for hybrid-state theorists are to prove that these mental states exist, and to address concerns about moral relativism. (For a comprehensive defence of this position see Prinz (2004) and Prinz (2008)).

(4) Hybrid-expressivism. According to hybrid-expressivists, moral judgements express two separable kinds of moral thought: cognitive and non-cognitive. For example, a moral judgement that x is wrong may express both a cognitive belief about the wrongness of x, as well as a non-cognitive mental state, such as a desire that not-x (Kauppinen, 2016). Unlike a besire, the moral judgement is decomposable into its cognitive and non-cognitive components.

An advocate of hybrid-expressivism can claim that moral facts or properties are either mind-dependent or mind-independent. The way

that they lean on this issue will depend upon exactly which kind of cognitive and non-cognitive mental states they claim are involved in moral judgements. For example, a hybrid-expressivist may claim that the moral beliefs expressed aim to identify mind-independent properties which do exist (endorsing IMR). But they may think that this belief alone is insufficient for a moral judgement, requiring the judge to also have a related non-cognitive mental state. This non-cognitive mental state does not contribute to determining truth. Alternatively, one may claim that one's non-cognitive mental states are necessary for determining truth conditions (endorsing DMR), but for the judgement to be a moral one, an individual must also express some kind of belief about those non-cognitive mental states. There are several possible variations of hybrid-expressivism which can be developed by centralising particular kinds of cognitive and non-cognitive mental states. As mentioned, one may think that any non-cognitive 'like' or 'dislike' is moral so long as it is cognitively endorsed. For example, perhaps some of our non-cognitive mental states such as our 'likes' or 'dislikes' are cognitively endorsed as being morally merited, or something similar. Or, perhaps only non-cognitive mental states of a certain kind ought to be considered moral, for example, those of our non-cognitive mental states which we demand others share.

(5) Moral thought pluralism (MTP). Moral thought pluralism is a very new meta-ethical position which essentially denies the assumption shared by cognitivists, non-cognitivists, and hybrid-state theorists that moral judgements are an expression of just one kind of mental state. The assumption that moral judgements are of one kind is reasonably widespread in meta-ethics. Theories emphasising one kind of mental state are born out of the commitments that theorists make to the importance of one or another feature of cognitive or non-cognitive mental states. For example, if one thinks that moral judgements are inherently motivating, one might be attracted to the idea that this is because they express non-cognitive mental states such as desires (beliefs, on the Humean view, being causally inert). On the other hand, if one thinks that moral judgements are truth-evaluable, they will sympathise with the idea that moral judgements express beliefs (desires, on the Humean view, are not truth-apt). For moral though pluralists, moral judgements can be expressions of more than one kind of mental state.

Less than a handful of philosophers have entertained the idea of moral thought pluralism. In 2006, Elizabeth Radcliffe argued that the idea that there are two distinct kinds of moral thought can be seen in Hume's work. In 2012, Uriah Kriegel proposed that moral judgements are expressions of either moral beliefs, or of moral aliefs which have

cognitive and non-cognitive components. Antti Kauppinen (2015) offers a theory of moral thought pluralism which says that moral judgements express both, moral beliefs and moral intuitions. For him, moral beliefs are cognitive, and intuitions consist of 'moral appearances' which have a non-cognitive sentimental aspect. Similar to both Kaupinnen's and Kriegel's positions is the idea that one kind of moral judgement is internally motivating and the other is not. Significantly, neither kind of judgement depends on the other for its instantiation. This point will be discussed further in Chapter 2.

Although moral thought pluralists can postulate the involvement of many kinds of mental states, in this thesis I will work with the assumption that moral judgements can be either cognitive or non-cognitive. That is, moral judgements may express a cognitive mental state, such as a belief that x is morally wrong, which has no relation to non-cognitive mental states, or they may express a non-cognitive mental state such as a desire that x. The moral belief must explicitly make reference to moral concepts such as 'bad', 'good', 'right', and 'wrong'. And, as I will later argue, the moral non-cognitive mental state must be adequately endorsed. There are several ways in which a non-cognitive mental state may be endorsed. The requirement of endorsement, including the possible different ways a mental state can be endorsed will be discussed in depth in Chapter 6. For now I will invoke the intuitive notion of second order desires. A second order desire arises when we have a desire about our desires. For example, one has a second order desire when one approves of, or disapproves of, one of their desires; or, when one has a preference for one desire over another upon recognising that the fulfilment of both of desires is not possible because the fulfilment of one necessarily interferes with the fulfilment of the other. On this view, cognitive judgements are truth-evaluable and non-cognitive judgements are not. This version of moral thought pluralism can retain the Humean theory of motivation and avoid postulating the existence of besires.

According to Hume we need both a cognitive and an appropriately related non-cognitive mental state to be motivated to act. Although some of the non-cognitive mental states that we experience in conjunction with cognitive mental states are very intimately linked, it is possible that the components of such cognitive and non-cognitive pairs have the ability to be pulled apart. It is also possible that only one or the other (but perhaps both) of these states has moral content. Moral thought pluralists claim that only one cognitive or non-cognitive element in a pairing that leads to motivation needs to have moral content to be considered a moral judgement (although both might be present). For

example, an individual may have a cognitive moral belief that murder is morally wrong. They may use this belief to advocate for the death penalty but they may be motivated to advocate by a non-moral desire to be elected into parliament. Or, an individual may have a non-cognitive moral desire to make the world liveable for future generations and be motivated to act when their beliefs indicate that this goal is being threatened. It is also possible that individuals regularly have both moral beliefs and moral desires which are relevant to various propositional objects. For example, one might have a (cognitive) belief that peace is good (A moral belief because it explicitly contains moral content), and they might also have a non-cognitive desire for peace (a desire that contains moral content by virtue of how it is endorsed).

For moral thought pluralists, a judgement's status as moral does not depend upon it being of one particular direction-of-fit, i.e. that it be either cognitive or non-cognitive (or both as the hybrid-state theorist claims). Cognitive moral judgements are truth-apt and non-cognitive ones are not. Moral thought pluralism faces problems of a kind not yet discussed. It owes an account of the commonalities between some cognitive and some non-cognitive mental states which unite them as moral judgements. This challenge is related to the challenges that pure cognitive and pure non-cognitive theories face in explaining exactly which beliefs are moral ones and which non-cognitive mental states are moral ones. For the pluralist, there must be some similarities between the two independent answers to these individual questions. In meeting this challenge, moral though pluralists will have to defend themselves against those who argue for restricting the class of moral judgements to certain kinds of mental states. This challenge will be further discussed as the thesis progresses.

In the remainder of this chapter I highlight some ways in which the terminology discussed in the chapter thus far relates to theories of moral judgement that emphasise intuitions, sentiments, and rational judgements.

1.4. Other related terminology

'Rationalism', 'sentimentalism' and 'intuitionism' are three common terms that often come up in the context of discussions about moral judgements. Specifically, moral judgements are regularly claimed to be rational judgements (rationalism), the expressions of our sentiments (sentimentalism), or the expression of intuitions (intuitionism). These terms have been in use for perhaps as long as 'cognitivism' and 'non-cognitivism'. However, it is not intuitively obvious what is meant

by them. It is also doubtful that the meanings of these terms have remained stable over time and across disciplines. The ambiguity in relation to the meanings of these terms may be alleviated by discussing their relationship to cognitive and non-cognitive mental states, that is, with Anscombe's emphasis on direction-of-fit in mind. Great care must be taken here as the terms can be employed in a wide range of ways in relation to the concepts introduced thus far. It appears that both 'sentimentalism', and more controversially, 'intuitionism' have ties to feeling states, specifically ones that have a world-to-mind component. This makes it likely that sentimentalists and possible that intuitionists are certain kinds of non-cognitivists or some kind of hybrid theorists. Having said that, it seems that we can also have cognitive intuitions that something is true or false which involve no non-cognitive component. These cognitive intuitions do not have any world-to-mind direction-of-fit. There is an even wider range of possible conceptualisations of rationalism available to us. Some even advocate for versions of rationalism which say that non-cognitive mental states have a causal role in determining moral judgements. It can become very difficult to identify what is meant by these terms at times as many authors fail to explicitly state their position.

There is a way in which the terms 'rationalism', 'sentimentalism', and 'intuitionism' can be mapped onto 'cognitivism' and 'non-cognitivsm' without any complications. Given the wide ranging possible uses of these terms I think that the most useful place to start will be to describe that mapping before moving on to describe more complex rationalist, sentimentalist, and intuitionist positions.

Rationalism and cognitivism: A rational moral judgement may be a pure belief involving no non-cognitive element. This makes it analogous to a cognitive moral judgement (a pure belief involving no non-cognitive element).

Sentimentalism and non-cognitivism: A sentiment can be characterised as a kind of (pure) non-cognitive mental state. On this view, there is no cognitive representation of how things are, were, or could be.

Intuitionism and cognitivism and non-cognitivism: A moral intuition can be characterised as a pure cognitive belief, or as a pure non-cognitive mental state about what is desirable or undesirable.[11]

[11]A moral intuition can be characterised as a pure cognitive belief about what is true or false. However, in the context of moral judgements, intuitions are not generally taken to be purely cognitive beliefs. This is indicated by the observation that those who advocate for them seem to take them to be motivating mental states

Any departure from this usage of the terms moves away from characterising rational judgements, sentiments, or intuitions in purely cognitive or purely non-cognitive forms. Many modern rationalists, sentimentalists, and intuitionists endorse more complicated stories. I will now map out these positions in terms of the mental states they may employ (assuming realism):

Sentimentalism and intuitionism. There are three ways to characterise the claim that moral judgements express sentiments or intuitions:

(1) They are a kind of non-cognitive mental state which does not aim to identify moral truth (this view is consistent with IMR).
(2) They are a besire which is a unitary mental state with both directions-of-fit. They are mind-dependently truth-apt (DMR).
(3) They are the union of two kinds of mental state. The states are modally separable.

Sentimentalists and intuitionists can easily map themselves onto the cognitive/non-cognitive/besire debate by committing themselves to one of these three views. If you think that intuitions necessarily involve some kind of cognitive component (for example, a representation of how things are, were, or could be), then you will be inclined to reject option (1). Option (2) will be attractive to those who think that intuitions involve a feeling of approval or disapproval which is irreducible into component parts. Option (3) will be most attractive to those who think that intuitions involve some kind of appraisal which consists of separable cognitive and non-cognitive mental states coming together to form a feeling (intuition). One can ask the same questions about sentiments. Sentiments are generally described as being consistent with non-cognitive mental states. But not just any non-cognitive mental state; they are generally taken to have a robust nature that persists over time. The exact stories told here are not important. The key point is that it is possible to discuss the terms 'sentiment' and 'intuition' more explicitly than they generally are.

Rationalism and intuitionism. The lack of clarity in the definition of moral reasoning in discussions about rationalism has not gone unremarked:

---

of some kind. That is, moral intuitions are presumed here to have a world-to-mind component. For those who claim that intuitions are purely cognitive, the motivation to call a cognitive (mind-to-world) mental state an intuition rather than a belief may arise as a way to distinguish beliefs for which we have, or could acquire, evidence for, from those which don't have or require evidence for, intuitions being the latter.

> The broad variability in definitions and characterizations of moral reasoning in the literature leads to two problems. First, it is not clear that people are all talking about the same thing when they argue that moral reasoning does or does not play a causal role in the production of moral judgements. Because there are many different theorists working with many different definitions or characterizations of the object of study, it is quite possible for theorists to mean entirely different things when they conclude that moral reasoning does or does not play a causal role in the production of moral judgements. (Saunders, 2015, p. 4)

As Saunders (2015) notes, the definition of rationalism one holds will determine whether or not one finds that moral reasoning is essential to making moral judgements. The definition, he warns for fear of begging the question, cannot be mere stipulation:

> ... the typical strategy being pursued is to simply designate a particular process as the one that constitutes moral reasoning, and then ask weather that designated process plays any causal role in moral judgement by looking at the empirical literature. Call this the stipulation without argument strategy... But in order for any of these arguments to be successful, it must be the case that they have designated as moral reasoning the process that really is what moral reasoning consists in; and there is good reason to be sceptical that any of them have done so. (Saunders, 2015, p. 6)

One broad claim that can be made of proponents of rationalism is that moral reasoning is required to decide what is good or bad, or to decide what ought to be done. The details of how moral reasoning may be required seem to range from the view that reasoning alone gives us moral principles, to the vague claim that moral judgements require some kind of reasoning processes. In what follows, I will outline three ways that one may think that cognitive mental states (or mental state components) are necessary for rational moral judgements or moral intuitions. I will then discuss the idea that rational moral judgements are those which involve conscious, multi-step processing. The latter kind of reasoning is consistent with non-cognitivism.

(1) A rational moral judgement or moral intuition is a purely cognitive one (a belief) (IMR)[12].
(2) A rational moral judgement or moral intuition is a besire (DMR).
(3) A rational moral judgement or moral intuition occurs when both a cognitive (belief) and a non-cognitive mental state occurs in relation to a propositional object. The cognitive (rational) component is essential.

One can determine their stance on position (1) by thinking about their opinion on the following statement:

> . . . if moral reasoning just is domain-general reasoning applied to moral questions, then there is not really anything special or significant about our capacity for it, and the capacity does not require any sort of special explanation; it is just like scientific reasoning, or mathematical reasoning, or any other domain of reasoning. If, however, moral reasoning cannot be fully explained by domain-general reasoning, but requires the addition of other psychological processes, such as the emotions or a moral faculty. . . then there is something genuinely different and special about the causal mechanisms that is not fully explained by domain-general reasoning. (Saunders, 2015, p. 12)

As should now be apparent, options (2) and (3) in the rationalist/intuitionist and the sentimentalist/intuitionist sections above make the same claims when it comes to the kinds of mental states that they claim moral judgements express (or consist of). Many modern rationalists allow a central place for some kind of non-cognitive mental state (or mental state component) in moral judgement. These kinds of rationalists seem to distinguish themselves from non-cognitive intuitionists and sentimentalists by emphasising the importance of the role that cognitive mental states (or components) play. Such rationalists argue that the non-cognitive world-to-mind mental state (or component) alone is insufficient for moral judgement; that, although such states might be in play, or even be a necessary pre-requisite for moral judgements, it is the addition of cognitive processes (or properties) which transform them into (potential) moral judgements. In contrast, sentimentalists who endorse (2) or (3) tend to think that, *despite* any cognitive elements which occur in moral judgements there is a fundamental non-cognitive element to them which is why they should best be viewed as arational

[12]This view is famously associated with Immanuel Kant (Hanna, 2017)

non-cognitive intuitions or sentiments. Those who take this view must claim that rational judgements are insufficient for moral judgements:

> What those who claim that moral reasoning cannot produce a moral judgement by itself typically mean is that the psychological processes that subserve domain-general reasoning, by themselves, cannot produce a moral judgement – that there must be some other (typically emotive) process that also plays a role. (Saunders, 2015, p. 11)

Saunders has attempted to identify some of the ways in which various authors seem to be using the terms 'rationalism', 'sentimentalism' and 'intuitionism'. For example, he thinks that prominent sentimentalist Jesse Prinz and prominent intuitionist Jonathan Haidt both hold some kind of hybrid view of moral judgements. And, that both avoid describing their position as rationalist despite there being cognitive components to their positions. Saunders notes that they both seem to view moral reasoning as domain-general reasoning, defining rational judgements as those which are 'free from affect' (this is option (1) in the list of possible rationalist positions above) (Saunders, 2015, p. 3).[13] Contrary to Haidt and Prinz, there are others who endorse rationalism whilst also claiming that both cognitive and non-cognitive mental states or mental state components are involved in moral judgements. They do not endorse rationalist position (1). Rather, they claim that rational processes are an essential causal component of the process of forming moral judgements (specifically, it seems as though these authors endorse rationalist option (3)). (Saunders, 2015, p. 4; Kennett and Fine, 2009, p. 77; Reichlin, 2014). As noted, advocates of rationalist options (2) or (3), importantly, reject the claim that non-cognitive mental states are sufficient for moral judgements. These examples demonstrate the worry that the distinctions made between rationalists and sentimentalists can be purely semantical when the underlying mental states are made explicit. Having said that, there is another possible explanation for their disagreement which arises from their being important differences in the kinds of cognitive and non-cognitive mental states that may be thought to be in play.

[13]Haidt thinks that rational processes only occasionally lead to a moral judgements within an individual and that rational processes are more likely to work *between* individuals (see Levy, 2006 in support of this view). Prinz ascribes to the view that reasoning involves conscious processing in steps and moral judgements are besires which do not require this.

Identifying as a cognitivist or a non-cognitivist does not tell a complete story about what ones theory of moral judgement is. For example, there are moral and non-moral beliefs, moral and non-moral desires, and plausibly moral and non-moral besires. Theories of moral judgement need to be able to distinguish moral from non-moral judgements. It is not enough to say that moral judgements are beliefs or moral judgements are desires. If one believes that moral judgements are non-cognitive, one is likely to also believe that the class of moral judgements is smaller than that of non-cognitive judgements i.e. there are non-cognitive judgements which we tend not to think of as moral judgements. This same issue arises for intuitionists and for rationalists, but perhaps not for sentimentalists if 'sentiment' is the name given to the subclass of non-cognitive moral judgements. To reiterate, in addition to identifying a class of mental state that moral judgements occupy, one needs to also identify a way to discriminate between moral and non-moral cases of mental states within that class. The appeal to 'sentiments' rather than non-cognitive mental states in general seems to be an attempt to isolate moral non-cognitive mental states from non-moral ones. However, more needs to be said about what a sentiment is in relation to other non-cognitive mental states. Further to this, if one holds a hybrid-state theory or a hybrid-expressivist theory, one needs to be able to specify the relationship of the cognitive and non-cognitive mental states (or mental state components). This is particularly important for the hybrid-expressivist as the mental states involved are separable so there is more opportunity for defending a rationalist position. Whichever way an individual leans in their conceptualisations of the mental states involved in moral judgements, the level of description will need to be expanded upon.

It is possible that the means by which the moral and non-moral are distinguished depend upon some kind of reasoning processes. For example, there seems to be a certain level of conscious processing that requires multiple steps in order to recognise that one has multiple desires, and to recognise that there can be conflicts in the fulfilment of desires such that the fulfilment of one interferes with the fulfilment of another. To prefer one over the other is a form of endorsement. The level of endorsement can be more or less rational. For example, moral non-cognitive mental states may be generated by a process of committing oneself to one non-cognitive desire over another not for any reason, but as a matter of personal preference. Or, they may be endorsed via highly rational methods, without which they would not qualify as moral judgements. As Saunders (2015) pointed out, it is clear that one's definition of a rational judgement is of central importance

to how one interprets these issues. All of the issues raised here will be expanded upon in due course.

In this chapter my aim has been to describe popular meta-ethical conceptions of moral judgements and the mental states which they rely upon or invoke. It is apparent that modern cognitive or non-cognitive positions attempt to incorporate into their view some of the positive qualities of the opposing view. Consequently, the maintenance of the distinction comes into question at times. The thesis proceeds to explore these theories, particularly the hybrid positions as they have received significantly less attention in the literature.

CHAPTER 2

Moral judgements and moral motivation

Observation suggests a close connection between moral judgements and moral motivation. This chapter explores the nature of that connection. The meta-ethical positions described in 1.3 vary in their ability to account for the relationship of moral judgements to motivation. Specifically, any view which claims an essential role for non-cognitive mental states (or mental state components) can easily explain why moral judgements tend to be motivating. Proponents of non-cognitive theories tend to use this fact as evidence for their preferred theory. Much of this issue has come to focus on the possibility or the actuality of the amoralist: an individual who can make moral judgements in the absence of feeling any motivation to act on those judgements.

2.1. Moral judgements and moral motivation

Michael Smith calls it a 'striking fact' that knowing a person's moral judgement about an issue reliably tells us something about how they are motivated to act (Smith, 1994; Rosati, 2016; Kirchin, 2012). The following passage illustrates our expectations that if someone thinks, or comes to think, that some action is morally wrong or bad they will in general have, or form, at least some motivation to avoid performing that action:

> Suppose we are sitting together one afternoon. World Vision is out collecting money for famine relief, so we are waiting to hear a knock on the door. I am wondering if I should give to this particular appeal. We debate the pros and cons of contributing and, let's suppose, after some discussion, you convince me that I should contribute. What would you expect? I take it that you would expect me to answer the door and give the collector my donation. But suppose I say instead 'But wait! I know I should give to famine relief. But what I haven't been convinced of is that I have any reason to do so!' And let's suppose that I therefore refuse to donate... But absent some such explanation,

> the puzzlement will be such as to cast serious doubt on the sincerity of my claim to have been convinced that it is right to give to famine relief at all. (Smith, 1994, pp. 6–7, 71)

The ability to explain this connection between moral judgements and moral motivation varies between the different theories of moral judgement. This chapter explores those variations.

Internalism and externalism. Individuals more formally commit themselves to a view on the relationship between moral judgements and moral motivation by identifying as either motivational *externalists* or motivational *internalists*. Motivational externalists claim that there is no necessary connection between moral judgements and moral motivation. For externalists, it is possible to make moral judgements and to fail to be motivated to act on those judgements. Motivational internalists, on the other hand, think that moral judgements are necessarily motivating (Rosati, 2016; Miller, 2003, pp. 7, 217; Kirchin, 2012, pp. 151, 150, 107):

> . . . moral claims are essentially bound up with motivation in a way that nonmoral claims are not (Ayer, (1936) 1971; Stevenson, 1937; Gibbard, 1990; Blackburn, 1993). Exactly what the connection to motivation is supposed to be is itself controversial, but one common proposal (motivation internalism) is that a person counts as sincerely making a moral claim only if she is motivated appropriately. To think of something that it is good, for instance, goes with being, other things equal, in favor of it in ways that would provide some motivation (not necessarily decisive) to promote, produce, preserve or in other ways support it. If someone utterly lacks such motivations and yet claims nonetheless that she thinks the thing in question is good, there is reason, people note, to suspect either that she is being disingenuous or that she does not understand what she is saying. This marks a real contrast with nonmoral claims since the fact that a person makes some such claim sincerely seems never to entail anything in particular about her motivations. Whether she is attracted by, repelled by, or simply indifferent to some color is irrelevant to whether her claim that things have that color are sincere and well understood by her. (Sayre-McCord, 2014)

The passage above demonstrates a move from the observation that moral judgements and moral motivation ordinarily co-occur to a stipulation: it is a moral judgement 'only if she is motivated appropriately'. Motivational internalism comes in unconditional and conditional forms, whereby, *unconditional* internalists think that the connection between judgements and motivation always holds and *conditional* internalists think that the connection is dependent on the judge's mental condition. The internalist/externalist divide forms one of the major rifts in ethical thought and will be spelt out in greater detail in Section 2.4 (see Björnsson et al., 2015b, p. 7).

2.2. Humean theory of motivation

David Hume's theory of motivation says that to be motivated to act we need both a belief (a cognitive mental state) and a suitably related desire (a non-cognitive mental state) (Hume, 1738–40, t.2.3.3, t.3.1; Rosati, 2016; Kirchin, 2012, p. 162; Miller, 2003, p. 7; Smith, 1994, Chapter 4). According to Hume (1738–40, t.2.3.3), beliefs on their own, do not cause behaviour. We can (and do) have beliefs about all kinds of things, none of which will cause us to act be it not for our desires. The Humean theory of motivation has great intuitive appeal. For example, to be motivated to pursue weight loss, one needs to have a desire to lose weight and beliefs about how to achieve that goal; neither desires or beliefs are, on their own, adequate for explaining directed action. Even the belief that one has a brain tumour would not prompt one to receive treatment unless one has some suitably related desire, such as a desire to continue living.[1] John McDowell expresses the appeal of the Humean theory of motivation as follows:

> . . . to cite a cognitive propositional attitude - an attitude whose content is expressed by the sort of proposition for which acceptability consists in truth [a belief] - is to give at most a partial specification of a reason for acting; to be fully explicit, one would need to add a mention of something non-cognitive, a state of the will or a volitional event. (1998, p. 213)

The Humean theory of motivation has historically been taken to provide strong support for non-cognitivism about moral judgements

[1]The causally inert nature of beliefs is supported by empirical findings, for example, damage to emotional capacities in the brain has had the consequence of leaving people unable to make decisions about what actions to take, despite their apparent wealth of cognitive knowledge (beliefs) about their situation (see Damasio, 1994, Chapter 3).

(over cognitivism) as it is the meta-ethical view which most straightforwardly explains the intuitive connection between moral judgements and moral motivation. Beliefs on their own are not motivating; if our moral judgements are expressions of cognitive moral beliefs, why should we see such a reliable connection between moral judgements and motivation? Non-cognitive theories of moral judgement are able to straightforwardly explain the connection as non-cognitive mental states are inherently motivating: if moral judgements are expressions of something like desires, it is clear why we are motivated by them.[2] As I will explain the picture has become far more complicated than this.

2.3. The Moral Problem

Michael Smith famously captures the essence of the difficulties of relating moral judgements to moral motivation. He brings to light the inconsistency of holding the following three commonly held and intuitively plausible beliefs about moral judgements. He calls this 'The Moral Problem':

> (1) Moral judgements express beliefs. This is the standard cognitivist position.
> (2) Moral judgements have a necessary connection to motivation. This is motivational internalism. It says that when we judge that something is wrong or right we will necessarily have some level of motivation to act accordingly.
> (3) Motivation requires a non-cognitive mental state. According to Hume, when desires combine with a means-end belief an individual will be motivated to act. Beliefs and desires are distinct existences. (1994, p. 126)

Smith spells out three popular ways of responding to this problem which each involve denying one of the three propositions. Firstly, one can deny (1): that moral judgements express beliefs. According to Hume, beliefs alone cannot motivate, yet it seems as though our moral judgements are motivating. The claim here is that if we accept that moral judgements are necessarily motivating, it can only be so because moral judgements have an essential non-cognitive component. Proponents of

[2]The Humean theory of motivation supports any theory of morality which allows that non-cognitive mental states have an essential causal role in the constitution of moral judgements (Rosati, 2016; Miller, 2003, pp. 7, 217; Kirchin, 2012, pp. 150, 107).

this view are attracted to non-cognitivism. The second way of responding to The Moral Problem is to deny (2): that moral judgements have a necessary connection to moral motivation. This position can maintain that moral judgements express beliefs and that distinct non-cognitive mental states are necessary for motivation. Proponents of this view are motivational externalists as they deny that there is a necessary connection between moral judgements and moral motivation. Whether or not our moral judgements are motivating depends on whether a relevant (external) desire also exists. The third possible response to The Moral Problem is to hold that moral judgements express beliefs and hold that this mental state is in fact motivating. Proponents of this view are forced to deny (3): the Humean theory of motivation. They cannot claim that beliefs and desires are distinct entities (Smith, 1994, pp. 127–8; also see Miller, 2003, p. 219).

2.4. Possible responses to The Moral Problem

The possible range of solutions is more nuanced than described by Smith. The range of responses roughly corresponds with the meta-ethical positions outlined in Section 1.3. Located among them are the 3 standard solutions as canvassed by Smith, as well as an option not yet mentioned, developed by Smith himself, who found the standard solutions unsatisfying. In addition, there are two other hybrid solutions. I have divided the responses into two sections, 'the historical divide' and 'the divide today' to trace the progression of thought over time.

The historical divide. As noted in Section 1.2, 'in the good old days' one was either a cognitivist or a non-cognitivist (Dreier, 2004). The non-cognitivist has a simple explanation of why we seem to be motivated by our moral judgements: our judgements are expressions of non-cognitive mental states such as desires, emotions, sentiments or attitudes which are intrinsically motivating. Non-cognitivists are motivational internalists. On the other hand, cognitivists thought that moral judgements express beliefs. Those cognitivists who accepted the Humean theory of motivation were motivational externalists and they struggled to explain the close connection of moral judgements to moral motivation. Supposing we allow that moral judgements are expressions of beliefs, if we adopt the Humean theory of motivation, it looks as if we can make moral judgements and be completely unmotivated by those judgements. Think of other cognitive beliefs such as 'insects have six legs', 'water is H_2O', and '$1 + 1 = 2$'. Such beliefs in no way compel us to act. If moral judgements are like these, the close connection between moral judgements and moral motivation is puzzling. One might have

the cognitive moral belief that murder is wrong because, for example, it is against God's commands, or because it causes suffering, but these beliefs are not sufficient for motivating behaviour without some kind of suitably related desire. The motivation one feels in relation to one's moral beliefs hinges on one's personal non-cognitive mental states such as desires which are distinct from beliefs.

Cognitivism. As noted in Section 1.3, cognitivism aligns with many of our intuitions about what we think our moral judgements and discourse seeks to achieve; that is, reporting our beliefs about objective moral facts. It is for this reason that many people are attracted to the idea that moral judgements are expressions of beliefs. In rising to the challenge to explain the close connection between moral judgements and moral motivation, it is open to cognitivists to deny that there is any strong connection. The majority of cognitivists, however, take the observational link between moral judgements and moral motivation seriously and seek to provide an explanation that is consistent with their claim that moral judgements express beliefs. They embrace one of three lines of thought in response to the challenge: externalism, unconditional internalism, and conditional internalism.

1. Cognitive externalism. Many cognitivists embrace motivational externalism. They think that moral judgements do not necessarily motivate and that when they do it is because of the judge's non-cognitive mental states. Externalists reject (2) in Smith's trifecta of incompatible statements. Cognitive externalists who accept the Humean theory of motivation need to explain either why, despite appearances, moral judgements are in fact not closely connected to moral motivation, or, they can tell a story about how a robust connection between moral judgements and moral motivation can exist despite there being no necessary connection. Defending the first option seems futile. An influential attempt at the second option says that people generally have the psychological motivation to 'be good', or to 'do good', or something similar, and this is why our moral judgements are reliably motivating. If the connection between moral judgement and moral motivation depends on the external desires of individuals, and there is reliable co-occurrence of moral judgements and motivation (as there seems to be) such external desires to 'be good' (or something similar), must be widespread. This particular way of explaining the connection has been accused by Michael Smith of being fetishistic:

> ... though externalists admit that there is a reliable connection between the moral judgements that a morally good and strong-willed person makes and her

> motivations, they can only explain why this is so by assuming, implausibly, that what makes a person morally good is the fact that she is motivated to do what she believes to be right, where this is read *de dicto* and not *de re*. Externalism thus enshrines a form of moral fetishism. (Smith, 1994, pp. 127–8)

According to this view, the moral person acts out of some abstract desire to be a good person, or to do what is right, or good. Smith thinks that this is a problem as the moral person is one who acts out of their immediate concerns for others as opposed to a commitment to some abstract desire that they wish to fulfil, thus the charge of fetishism (Smith, 1994, pp. 71–76; Kirchin, 2012, p. 157). The claim that acting to promote the good (or something similar) is morally inferior or fetishistic is far from intuitively obvious (Kirchin, 2012, pp. 154–8). In fact, it is commonly thought that attending to immediate concerns can sometimes be the wrong thing to do when doing so conflicts with the more abstract goal of 'doing good'. It may also explain our inclination to break general rules in particular cases when the rule fails to serve to increase good.

The second and third common cognitivist responses embrace motivational internalism and deny that beliefs and desires are distinct entities. There are two versions of this approach: conditional and unconditional.

2. Cognitivism: Unconditional motivational internalism. Unconditional motivational internalists are persuaded by the claim that moral judgements are necessarily motivating. That is, "... [n]ecessarily, if a person judges that it is morally wrong to ϕ, then she is, at least to some extent, motivated to refrain from ϕ-ing" (Eggers, 2015, p. 85). They nevertheless maintain that moral judgements express only beliefs. They claim either that beliefs motivate or that moral beliefs are necessarily connected to desires somehow. They deny the Humean theory of motivation, in particular, the claim that beliefs and desires are distinct entities.

3. Cognitivism: Conditional motivational internalism. Conditional motivational internalists claim that "... necessarily, if a person judges that it is morally wrong to ϕ, and if she is psychologically normal/ practically rational/virtuous, then she is, at least to some extent, motivated to refrain from ϕ-ing" (Eggers, 2015, p. 86; also see Kirchin, 2012, p. 154). Unlike unconditional motivational internalists, conditional motivational internalists think that observational evidence supports the view that moral judgements and moral motivation can come apart. However, they maintain that there is a necessary connection between moral judgements and moral motivation when certain conditions are

met. They differ in their views about which conditions need to be met for the link to manifest. The link has been described quite broadly, for example, some people claim that the link will obtain if the judge is 'virtuous', or 'psychologically normal', or "... in the absence of any malady such as depression, psychopathy, exhaustion etc." (Björnsson et al., 2015b, p. 8). It is not clear that these explanations are specific enough to adequately explain what is going on. What does it mean to be virtuous? Is it to have moral goals? If so, why should such a position be internalist rather than externalist? And, what exactly is it that leads psychologically non-normal individuals to not be motivated by their moral judgements? A hugely influential version of the conditional nature of the link is owed to Michael Smith who claims that moral motivation will follow moral judgements if the judge is *practically rational*. Smith argues that values are beliefs about what we would desire if we were fully rational (Smith, 1994, p. 151; 2015, p. 27). Having arrived at such a belief, he says that it would be irrational not to be motivated by it. This view seems to provide an adequate specification of how the link between moral judgements and moral motivation can break down; for example, it is possible that psychopathy, depression, or fatigue can lead people to be practically irrational; to fail to have the desires that they believe they would have were they fully rational. Absent practical irrationality, the connection will hold. This position retains the Humean theory of motivation for motivating reasons - we need both beliefs and desires to be motivated - but says that normative reasons are beliefs which rationally commit us to relevant desires. It thus denies that normative beliefs and desires are distinct entities in the way that Hume described. There are important questions about Smith's position, for example, how do we come to determine what we would desire were we fully rational? Does such a process always (or usually) depend upon the person making the judgement to deliberate upon their actual desires? And, by what method can a practically rational individual preference desires?

It seems that cognitivists do have ways of explaining the connection of moral judgements to motivation, however, their explanations are all more convoluted than the ones offered by non-cognitivists. If one is not persuaded by any of the above attempts to explain the link of moral judgement to moral motivation then they may be forced to give up the idea that moral judgements express beliefs (absent some other explanation). The alternative option of denying the connection seems intuitively unappealing to many. The major cost of giving up the idea that moral judgements express beliefs (Option (1)) is that moral judgements lose their ability to be mind-independently true or false.

Non-cognitivism. Those unpersuaded by cognitivist attempts to explain the connection between moral judgements and moral motivation have commonly found themselves drawn to non-cognitivism. The non-cognitivist rejects option (1) in Smith's triad: the claim that moral judgements express beliefs. If the Humean theory of motivation is true, and if our moral judgements are necessarily motivating, it appears that non-cognitivism has the upper hand in explaining the connection. It provides a simple explanation of why it is that our moral judgements are motivating: moral judgements express our non-cognitive mental states which are intrinsically motivating. For the independent moral realist (non-cognitivist), this has the consequence of rendering moral judgements incapable of being true or false, as they require truths to be mind independent. So, although non-cognitivism easily explains motivation, it challenges our realist intuitions.

In sum, the debate between the cognitive and non-cognitive camps has largely focused on the connection between moral judgements and moral motivation by addressing the question of whether it is possible for someone to judge that something is wrong without feeling motivated to act in relevant ways, or the question, as highlighted in Smith's example: could someone who says that giving money to charity is morally right, and who feels no motivation whatsoever to give to charity, be making a moral judgement? At face value, externalists say yes and internalists say no.[3] The general question has developed into a debate between internalists and externalists about the possibility of the amoralist, an individual who can make moral judgements without being motivated by those judgements. This is discussed further in Section 2.5. Non-cognitivism's ability to explain the connection between moral judgements and moral motivation is where the appeal of the theory lies and it has led some people to flatly reject cognitivism (and with it, realism). Simply put, many people have a strong intuition that moral judgements are necessarily motivating and so must involve some non-cognitive mental state. For them, the amoralist is impossible. Historically, moral judgements were taken to be pure incarnations of either cognitive beliefs or non-cognitive desires. And, if you accept that there is a robust connection between moral judgements and moral motivation then the idea that moral judgements are non-cognitive best explains that connection. Today the story is more complicated.

The divide today: Three hybrid accounts. Aside from plausible attempts by cognitivists to explain the connection between moral

[3]As Smith notes, internalists may say yes under certain conditions, making amoralists irrational rather than impossible.

judgements and moral motivation, the question about whether moral judgements are internally or externally motivating is made more complicated by the introduction of hybrid theories of moral judgement. It is no longer only non-cognitivists who emphasise the importance of non-cognitive mental states in moral judgements; there are also three hybrid accounts which aim to tell a story which satisfies cognitivist as well as non-cognitivist intuitions about morality. All three hybrid accounts presented in Section 1.3 state that non-cognitive mental states are important to moral judgements, though they take different views about whether such non-cognitive states are necessary or sufficient. I will now examine how they address The Moral Problem, and highlight their differing commitments to the Humean theory of motivation.

Hybrid-state theory. The Humean theory of motivation has been hugely influential. It is intuitively compelling. However, as discussed in Chapter 1, advocates of hybrid-states think that it does not tell the full story. According to some, the Humean theory of motivation leaves out a third kind of mental state, the state of being motivated and representing the world at one and the same time: besires (see Kirchin, 2012, Chapter 6). Importantly, the cognitive and non-cognitive components cannot be separated or pulled apart. The existence of besires is highly controversial. And, even if they exist, there is a further question of whether at least some of them should be considered *moral* judgements. But if they exist and should be considered moral judgements, they open an avenue for explaining why moral judgements are motivating, and they create the potential for mind-dependent moral truth, though as discussed in Section 1.2, they face challenges in specifying a standard of truth (Kirchin, 2012, p. 172). Advocates of besires are internalists. They deny option (3) in Smith's inconsistent triad: moral judgements are a single mental state with both cognitive and non-cognitive content. Moral motivation is not dependent on the union of distinct entities according to advocates of besires (Smith, 1994, p. 128).

The four positions discussed in this section thus far cover the three responses to The Moral Problem that Michael Smith describes (external cognitivism, non-cognitivism, and hybrid-state theory) as well as Smith's own view which is a cognitive internalist position (see Smith, 1994; McDowell, 1998; Wiggins cited in Miller, 2003, p. 7). Smith's position argues for a necessary connection between moral beliefs and moral motivation without positing besires, or claiming that the link between judgement and motivation is unconditional. He takes both of those options to be problematic. The following two positions illustrate two further ways in which one might respond to The Moral Problem.

Hybrid-expressivism. As noted in Section 1.3, hybrid-expressivists reject the common assumption that moral judgements are only one kind of mental state. They think that moral judgements consist of both cognitive and non-cognitive mental states. They are motivational internalists in the sense that they think that all moral judgements are motivating because they all have a non-cognitive component. However, they also require that a cognitive moral component is present. So although a non-cognitive mental state is necessary, it is not sufficient.

Hybrid-expressivists can retain the Humean theory of motivation by claiming that the connection between the cognitive moral judgement and non-cognitive moral judgement is incidental: that moral judgements occur when cognitive judgements and non-cognitive judgements happen to co-occur in relation to the same statement, for example, if I say killing is wrong, according to the hybrid-expressivist, I am expressing a belief, as well as a desire about how I wish things to be. If either one is missing, the judgement does not reach 'moral judgement' status. It is an advantage of hybrid-expressivism that beliefs and desires can remain distinct entities as Hume proposed. Kauppinin describes one plausible way in which beliefs and non-cognitive states such as desires may be related, though not necessarily so:

> Importantly, like other appearances, moral intuitions [non-cognitive states] also attract us to assent to the corresponding proposition, to form the belief or judgement. Suppose I do form the belief that I ought to ϕ on the basis of the intuition that I ought to ϕ. Since the intuition both defeasibly motivates me to ϕ and attracts the belief that I ought to ϕ, my belief that I ought to ϕ is non-accidentally linked to defeasible motivation to ϕ – they have a common cause. Of course, on this view moral belief itself doesn't motivate. But it reliably co-occurs with motivation. It is nevertheless not pointless to form the belief, since it amounts to the kind of endorsement that makes a difference to further reasoning. (Kauppinen, 2015, p. 246)

In the moral case, an action may appear wrong to us because we recognise that it is in tension with our non-cognitive goals (which form part of our moral judgements). The salience of the situation for our non-cognitive moral desires may attract us to form beliefs about the wrongness of certain acts which we do not take to depend on our non-cognitive mental states which were active originally. For example, perhaps we come to form the belief that x natural property is bad

and would be bad even if we did not have any desires that would be frustrated by the bad things happening. Non-cognitive mental states can open up new avenues for belief formation although there is no necessary connection. Similarly, beliefs can open up new avenues for desires, though there is no necessary connection. Take the example by Kirchin that the desire to travel to the moon is facilitated by the belief that the moon is a place that one could travel to (Kirchin, 2012). In the moral case, the desire that the practice of female genital mutilation is ended may be facilitated by the discovery and belief that it is a thing that occurs. That is, perhaps one will go on to form a moral desire for the practice to end once learning of it, but perhaps not. The beliefs and desires remain distinct. Although this kind of hybrid-expressivist says that there is no necessary connection between cognitive and non-cognitive mental states, both kinds of mental state must eventuate for there to be a moral judgement. This process described whereby one kind of judgement directs ones attention such that one is poised to make a judgement of a different kind can also occur for the moral thought pluralist.

Hybrid-expressivists may retain the Humean theory of motivation if they tell the above story, though they may choose not to do so. They may instead claim that there is a necessary connection between cognitive moral judgements and non-cognitive ones. And if they think that there is a necessary connection between the judgements, they may specify that it goes in either direction. However, depending on the specifications of the proposed connection, they may be led back to a position which is indistinguishable from cognitivism. Hybrid-expressivists may say that moral judgements begin with a non-cognitive moral judgement which causes a cognitive moral judgement. However, there is little reason to think that there is a necessary connection in play here. The relationship is more likely to be as described above by Kauppinen: there is assent to belief, rather than a necessary connection. That is, the non-cognitive judgement directs attention in a way that increases the likelihood that one will form a moral belief. They may also claim that cognitive judgements cause non-cognitive ones. But, this position collapses into cognitive motivational internalism: one makes a cognitive moral judgement which necessitates a related moral desire. Hybrid-expressivists can also claim that the necessary causal connection goes both ways: cognitive judgements cause non-cognitive ones and vice versa. But again, there is little reason to think that the connection is necessary. To summarise, for the hybrid-expressivist the presence of one mental state is not sufficient for moral judgements. The most plausible

version of this view says that the connection between cognitive and non-cognitive mental states is incidental.[4]

Moral thought pluralism. As discussed in Section 1.3, moral thought pluralists, like hybrid-expressivists, reject the common assumption that moral judgements consist of just one kind (cognitive or non-cognitive) of mental state. For the moral thought pluralist, they can be either one or the other; they may even, at times, be both, as the hybrid-expressivist claims. In what follows, I will discuss how moral thought pluralism deals with The Moral Problem as explained by Uriah Kriegel (2012). Kriegel notes that if moral judgements can be expressions of different kinds of mental states, it may be the case that both internalism and externalism are true; internalism true of any moral judgement which arises from a non-cognitive state, and externalism true of our moral beliefs. He argues that if this is the case, then "... there is no single notion for which all three theses in the [Smith's] triad are compelling... The triad can now be dissolved into two perfectly valid (and plausibly sound) arguments about the nature of different kinds of moral judgement" (Kriegel, 2012, p. 481). Kriegel refers to one kind of moral judgement as moral aliefs. For now, aliefs can be thought of as a kind of non-cognitive mental state. Later on when I address reasons for and against restricting the class of moral judgements, differences between the kinds of non-cognitive mental states will become important, i.e. it is likely the case that not just any non-cognitive mental state should qualify as a moral judgement. The other kind of moral judgement, he claims, are moral beliefs (cognitive moral judgements). He describes the nature of the relationship of these judgements to motivation as follows:

Non-cognitive moral judgements:

(1) Moral aliefs are inherently motivating
(2) Mental states in general, including moral aliefs, cannot both have objective purport and be inherently motivating. Therefore,
(3) Moral aliefs do not purport to be about objective matters of fact. (Kriegel, 2012, p. 481)

Cognitive moral judgements:

(1) Moral beliefs purport to be about objective matters of fact.
(2) Mental states in general, including moral beliefs, cannot both have objective purport and be inherently motivating. Therefore,

[4]For a discussion of a major problem with this view see Street, 2006.

(3) Moral beliefs are not inherently motivating.
(Kriegel, 2012, p. 481)

Kriegel declares that moral thought pluralists can retain the Humean theory of motivation in its entirety. For them, cognitive moral judgements are externally motivating and non-cognitive moral judgements are internally motivating: "[t]he emerging picture combines cognitivism and externalism about moral beliefs with internalism and non-cognitivism about moral aliefs – while holding on to Humean psychology" (Kriegel, 2012, p. 481).

Moral thought pluralists may disagree about the proportion of cognitive judgements that are made relative to non-cognitive ones. For example, they may think that moral judgements are more often non-cognitive, though sometimes people make cognitive moral judgements; they may think that the proportion of cognitive to non-cognitive moral judgements is roughly equal; or, they may think that cognitive moral judgements are more common than non-cognitive ones. The commitments that a moral thought pluralist makes to this range of possibilities will influence how capable they are of explaining the link between moral judgements and motivation. For example, pluralists who claim that moral judgements are generally non-cognitive have the advantages that non-cognitivists have in explaining motivation (they also have the difficulty of explaining truth-aptness unless they claim that non-cognitive moral judgements are regularly accompanied by moral beliefs as the hybrid-expressivist claims). On the other hand, pluralists who think that most of our moral judgements express beliefs will have to defend an externalist account of motivation and defend the view that motivations to 'do good' are common among the general populous. Those in the middle can postulate some combinational story about the connection of moral judgement to motivation.

2.5. Amoralism and normal variation in moral judgements

The connection between moral judgements and moral motivation has its origins in our observations about ordinary moral judgements. Debates about the motivational status of moral judgements have come to focus on the possibility and the actuality of the amoralist: an individual who can make moral judgements and be entirely unmotivated to act on those judgements. For motivational externalists, the amoralist is possible and for motivational internalists, impossible. Empirically, there have been efforts to find cases where the connection between moral judgement and motivation does not hold. In these cases, individuals appear to make moral claims and use moral concepts but are not

motivated by those judgements. In response to these cases, theorists can commit themselves to conceptual claims about moral judgements. That is, they can stipulate that moral judgements *require* motivation. In this case, a judgement does not qualify as a moral one unless it is motivating to the person who made it.

Of the positions described in Section 1.3, unconditional cognitive internalists, non-cognitivists, hybrid-state theorists, and hybrid-expressivists are internalists, all claiming that amoralists are impossible. External cognitivists, conditional cognitive internalists, and moral thought pluralists, on the other hand, think that the amoralist is possible. Defending cognitivism by appeal to the amoralist is not an option for unconditional external cognitivists. Such findings would discredit their claim that the connection between moral judgements and motivation is unconditional. For those who think that there are moral facts, the denial of the possibility of the amoralist is particularly puzzling. It rules out the possibility that, say, some individual, or alien race could somehow become aware of the moral facts, which they, as realists, think exist, and be completely unmotivated by them.

The actuality of the amoralist. Many theorists have sought to find real life examples of amoralists whose moral judgements dissociate from motivation. Psychopaths and individuals with damage to the ventromedial prefrontal cortex are considered the closest real world examples of amoralists. Adina Roskies (2003) famously defended the actuality of the amoralist, presenting examples of individuals who use 'moral' sentences without experiencing motivation. These individuals employed moral concepts such as 'good', 'bad', 'right', and 'wrong' and use them to describe certain acts, states of affairs and so on. These cases are challenged by bringing into question whether these individuals are competent with the moral concepts that they employ (Roskies, 2003; Rosati, 2016). For example, both Cholbi (2006) and Kennett and Fine (2008) criticize Roskie's argument by questioning whether individuals with ventromedial damage have proper moral beliefs. A famous experiment by Blair (1995), found that psychopaths were unable to distinguish between moral and conventional rules. However, the results of Blair's study have failed to be replicated (Aharoni, Sinnott-Armstrong and Kiehl, 2012; Dolan and Fullam, 2010; Cima, Tonnaer and Hauser, 2010). One study by Cima, Tonnaer and Hauser (2010) suggests that psychopaths are just as capable of distinguishing right from wrong as healthy individuals. These discussions have as yet failed to determine whether psychopaths are competent with moral terms. Even if the individuals described are competent with moral concepts,

the empirical literature cannot tell us whether such cognitive judgements should be classified as moral ones or not.

One problem is that it is difficult to establish the actuality of the amoralist empirically because it is a challenge to determine whether an individual who makes a moral judgement, and who appears to lack motivation to act on that judgement, nevertheless experiences a very slight motivation, perhaps motivation which is very quickly overridden: "... one of the problems of the internalism/externalism debate is the notorious difficulty of proving the absence of (relevant) motivation. (See Finlay, 2006, p. 209)" (Eggers, 2015, p. 90). Eggers notes that to empirically test the motivation of an amoralist we would need a situation of the following kind: "... the situation should be one in which all that is asked of the agent is, literally, to make only one move with his finger, or to only speak one word, so that the motivation to act coincides almost necessarily with the action itself, and no dissociation of motivation and action is to be expected" (Eggers, 2015, p. 91). With this serious obstacle in play, discussion about the amoralist may be unproductive aside from its being a useful tool for discovering, or helping people to think about what they believe about the motivational nature of moral judgements.[5]

Nick Zangwill claims that discussion about the possibility of amoralists is counter-productive for making the case against internalism. He argues that the phenomenon of variation in the strength of our moral motivation (as opposed to looking for cases of absent motivation) may provide support for motivational externalism.

Of the possibility of the amoralist, he claims:

> ... it is a possibility that is of little dialectical significance by comparison with the actuality of variation. The possible amoralist should not have a central role in the dialectic as it has done, both because it is merely an extreme case of a spectrum that is interesting, and also because being an extreme case, its

[5]Some people have attempted to use aggregate opinions to conceptually define moral judgements, but this process has yielded conflicting results and there is also the possibility that people are bad at identifying the features of moral judgements consciously, but nevertheless are able to use the term accurately. Shaun Nichols (2002; 2004, Chapter 3) found that laypeople maintained that psychopaths, although lacking in relevant motivation, had the ability to understand cognitive moral requirements. He took his study to challenge motivational internalism. Contrary to Nichols findings, other research suggests that laypersons intuitions about morality support motivational internalism when conditions were set (Björnsson et al., 2015a; Strandberg and Björklund, 2012; Björnsson et al., 2015b).

> actuality is doubtful, and its alleged mere possibility of little consequence. Actual variation is of much greater dialectical significance. (Zangwill, 2015, p. 51)

Zangwill argues that instead of thinking about the amoralist, we should be focusing on which position, internalism or externalism, best explains the normal everyday variation in moral motivation that normal individuals experience. And he thinks that externalism provides a more plausible explanation of variation than does internalism. Amoralists, he says, are at the extreme end of variation in moral motivation and it is better to focus on explaining the variation of moral motivation itself. Zangwill thinks that there is much to learn from investigating the motivational differences between people who share moral opinions, yet who seem to be motivated quite differently, or by looking at the variation in moral motivation that a single individual can experience just in the course of a single day. He argues that it is plausible to think that our moral judgements can remain constant whilst our degree of moral motivation can fluctuate quite dramatically. He begins the case for externalism by reiterating the distinctness of cognitive and non-cognitive mental states:

> If the motivating desire and belief are two distinct states, then they are independent in this sense: people can have the same beliefs but different desires (in the sense of different strengths of desire); people can retain their beliefs while changing their desires; and people who actually have certain beliefs and desires might have had those beliefs with different desires. The motivating desires are not determined by the beliefs because they are distinct. (Zangwill, 2015, p. 52)

Zangwill's Indifference Argument claims that our moral judgements often remain constant while our motivation to act on those judgements varies (Zangwill, 2015, p. 51). Faced with three possible explanations of ordinary variance, "(1) the presence or absence or strength of a distinct desire; (2) a difference in rationality; (3) just a brain difference", Zangwill surmises that the most plausible explanation of the variance is due to fluctuations in one's desires (Zangwill, 2015, pp. 50–51). Further to this, he claims that this issue is beholden to empirical investigation:

> The form of the indifference argument for externalism is this: here are two cases which are similar in respect of their moral judgements, but which differ in motivational upshot. So there must be some other difference between the cases. The question is: What

> best explains the difference? The internalist argues that a difference in non-instrumental rationality is the best explanation. The externalist says that a distinct desire is the best explanation. The present point is that it looks like an empirical question. (Zangwill, 2015, p. 55)

The phenomenon of variation in moral motivation, he thinks, is best explained by positing that we have stable moral beliefs and variation in our desires, a position which supports the externalist over the internalist. However, the possible explanation that we experience a difference in rationality (2), is also very plausible. As already mentioned, most people do not take just any non-cognitive mental state to be a moral one. Among the possible features which separate my moral from my non-moral non-cognitive mental states may be something like my insistence that others share some of my desires and my indifference to their sharing others (the former being my 'moral' judgements). Or, perhaps judgements become moral ones when I prioritise them upon realising that their fulfilment contradicts the fulfilment of my other desires. In either case, the 'moral' judgements are the more stable, non-fluctuating members of my set of non-cognitive judgements. To explain variation in motivation in this case, one could maintain that one's moral non-cognitive judgements remain stable, but that variation in motivation to act on those moral non-cognitive judgements occurs when ones other competing non-moral desires temporarily overcome 'moral' ones. Directed action in those cases is not the outcome of any cognitive or non-cognitive moral judgements. When this happens to an individual they experiences a failure in rationality as they compromise the fulfilment of desires that they themselves take to be more important.

It seems that both internalists and externalists can explain normal variation in motivation. In the case of the externalist, people can share moral beliefs, yet act differently because they have different desires. In the case of internalism, people may share non-cognitive moral judgements, yet fail to be motivated by them when their directed action is the outcome of other non-moral non-cognitive mental states.

The possibility of the amoralist. Discussion about the amoralist can be viewed as a tool to test one's intuitions and commitments to what they think a moral judgement consists of. In discussing the possibility of the amoralist, there is the worry that one's intuitive theoretical commitments are likely to influence one's interpretation of the possibility of the amoralist:

> A general worry about the discussion of such cases is that conceptual commitments (internalist or externalist) will govern our interpretation of experimental data, undermining the role of actual cases as providers of independent evidence for or against internalism (Kauppinen, 2008). Almost the opposite worry is that the intuitions in question are not in fact a priori intuitions, but rather are based on assumptions about the nature of actual moral judgements: Perhaps when we think that amoralists are impossible, we do so because we think that normal actual moral judgements motivate (Björnsson and Francën Olinder, 2012; for criticism, see Kauppinen, 2013; Wood, 2014). (Björnsson et al., 2015b, p. 15)

The following hypothetical illustrates one of many attempts to show that moral judgements and moral motivation can come apart:

> Imagine Sam is suffering from terrible depression. Much or all that he values (or perhaps, only seemingly 'values') appears colourless to him, bleak, and at a distance. He can see and judge that he should give to charity, but feels no motivation to do so. Further, he is being sincere and not acting, and unlike the parrot he understands exactly what he is saying. . . we have a judgement and no motivation, And, the judgement appears to be a legitimate one. (Kirchin, 2012, p. 153)

Discussion about the possibility of the amoralist via the development and consideration of hypothetical scenarios may be more useful than looking for real life amoralists for helping us to decide what to think about whether people can make moral judgements without experiencing emotion. Importantly, in considering hypothetical cases we are not giving *authority* to what our intuitions say about them. Rather, we are attempting to aid the development of our theories, by identifying the reasons behind our intuitions. It is possible to stipulate arbitrarily or to decide what to think based on ones intuitions about hypothetical cases. To avoid making theoretical commitments which are arbitrary in this way, those who take a stance on the moral status of hypotheticals need to justify their position. When we demand justification, the practice of developing hypothetical situations is merely a tool used to gain explicit access to underlying theories. As an example, consider how one might justify their answers to questions about the following case: Consider Sam. Sam does not have any desires one would describe as moral in the

sense that he thinks others should share those desires. He has however, over the span of his life developed a theory about which things are morally bad and morally good. He, himself, is not motivated to act on those bad or good things except when doing so fulfils one of his desires. Should we say that Sam makes moral judgements and hold him morally accountable for his beliefs? If we are using intuitions as a tool, we cannot take one's feelings one way or another about the answer to this question as evidence in support of a view. I will discuss these issues further in Chapter 6.

Prinz (2015, p. 65) claims that philosophical debates about the possibility of the amoralist have reached an impasse and are not capable of providing support for either internalism or externalism, claiming that such discussion has failed to yield any verdicts despite the intense focus it has attracted over the years. He turns to other empirical methods of investigating the mental states involved in moral judgements which are the focus of Chapters 4 and 5. The impasse in relation to the possibility or actuality of the amoralist is to be expected if moral thought pluralism provides an accurate depiction of moral judgements.

2.6. Concluding remarks

Historically, moral judgements were thought to be either cognitive or non-cognitive. The Humean theory of motivation provided strong support for non-cognitivism as it was significantly better able to explain the robust connection between moral judgements and moral motivation than its rival, cognitivism. Both internal and external cognitivists have since provided explanations which go some way towards explaining how moral judgements are so reliably connected to moral motivation. However, non-cognitivism still seems to better explain the relationship when pure cognitivists are compared to pure-non-cognitivists. Today, with the advent of hybrid theories of moral judgements the landscape has become convoluted.

The Humean theory of motivation, whereby our beliefs are causally inert and require the presence of a related non-cognitive mental state to motivate us to act remains intuitively appealing and is still widely accepted. Of the three possible hybrid positions (hybrid-state theory, hybrid-expressivism, and moral thought pluralism), advocates of hybrid-state theories reject the Humean theory of motivation, claiming that there are mental states which are both cognitive and non-cognitive at the same time (besires). For advocates of besires moral judgements are necessarily motivating states with cognitive content. They endorse motivational internalism and deny that beliefs and desires are distinct

entities. Hybrid-expressivists agree with the Humean theory of motivation that motivation requires both a cognitive and a non-cognitive mental state. They deny the claim that moral judgements can be instantiated by just one kind of mental state. Moral judgements occur when a cognitive judgement with moral content and a non-cognitive judgement (with or without moral content, depending on the theory) apply to a single moral statement. For hybrid-expressivists, moral judgements are necessarily motivating; you cannot have a moral judgement which is not at all motivating. Hybrid-expressivists may or may not think that the connection between the two mental states is necessary. Externalists think that it is not necessary and internalists think that it is. They may think that the causal pathway goes one way or the other, or both ways. However, hybrid-expressivists who think that cognitive judgements with moral content cause non-cognitive states with moral content are led back to an internal cognitivist position. Moral thought pluralists can hold on to the Humean theory of motivation. For moral thought pluralists, externalism is true of our cognitive moral judgements and internalism is true of non-cognitive ones. Their ability to explain the connection of moral judgements to motivation depends upon the proportion of cognitive relative to non-cognitive moral judgements and their stance on externalist motivation.

Debates about the actuality and the possibility of the amoralist have featured heavily in debates between internalists and externalists. Both, evidence of real world examples, and the appeal to the theoretical possibility of amoralists have been used to show that it is plausible for moral judgements and moral motivation to come apart. There are practical obstacles to empirically investigating the amoralist. It is incredibly difficult to tell whether, for example, psychopathic or depressed individuals who seem to lack moral motivation whilst retaining the ability to make moral judgements experience *zero* motivation or just very weak motivation which is quickly overturned. And of such cases, it is always an option for internalists to retreat to stipulation in defining moral judgements, and claim that motivation just is an essential part of what it means to have made a moral judgement. Externalists may be thought to fare better when they concentrate on explaining normal everyday variation in moral motivation. It seems plausible that variation in my motivation to help people is contingent on my external desires as opposed to variation in my moral belief that it is good to help people in need, for example. However, internalists are able to explain variation in motivation by appealing to competition between moral and non-moral non-cognitive mental states. In sum, appealing to the phenomenon of motivation does not obviously help us decide

what moral judgements are. Even if one is willing to stipulate, without reason, that motivation is essential, we are left with plausible cognitive, non-cognitive, and hybrid views of moral judgement which are all able to explain why moral judgements are motivating.

The following three chapters discuss the empirical literature which purports to show that ordinary moral judgements contain a non-cognitive component. Many of the findings from this literature claim to show that ordinary moral judgements depend upon feelings and conclude from this that moral judgements contain an essential non-cognitive component. That is, they claim to discredit cognitive theories. I examine the force and potential of this literature for contributing to meta-ethics. I find that the literature does not have the ability to support non-cognitive theories over cognitive ones. However, the literature has a complex ability to compare hybrid theories of moral judgement as the various hybrid theories have differing views about the specific kind of content that is involved in moral judgements. As I will explain, some experimental studies raise important questions about the kinds of content that some judgements that are ordinarily taken to be moral ones have.

CHAPTER 3

Emotions, appraisals, and moral judgements

The following three chapters concentrate on the empirical literature from psychology that has been used by both philosophers and psychologists to investigate the nature of our moral judgements. Much of the work from psychology looks at the relationship of emotions to moral judgements. A commonly held conclusion of the empirical literature is that people make moral judgements on the basis of their emotional gut-feelings. This is generally considered to cast doubt on cognitive theories of moral judgement as emotions are presumed by many theorists to have an essential non-cognitive nature.

There are two important problems inherent in the psychological research on moral judgements. Firstly, the studies generally set up some stereotypical moral situation and ask people their views on the situation. Researchers take answers to such questions to reveal people's *moral* judgements and proceed to examine the mental states involved in their decision making process. As discussed previously, it is always an option for theorists to stipulate a definition of moral judgement, such that the answers people provide to experimental moral questions do not necessarily reveal their moral judgements at all. The problem with stipulating a definition though, is that, depending on the stipulated definition, the research may reveal that people rarely, if ever, make moral judgements. There needs to be a good reason to discount the answers people give to stereotypically moral questions as being their moral judgements.

The second fundamental issue relating to the conclusions of the psychological studies employed is the question of what emotions are. Many theorists take emotions to have an essential non-cognitive component based on their close connection with feelings and motivation. And, based on the empirical literature which shows that emotions are active in moral judgements, they have concluded that moral judgements are not purely cognitive mental states. This conclusion may be premature. Perhaps moral judgements can still be purely cognitive even if emotions are reliably correlated with moral judgements. Interpretation of the empirical research will depend upon the answers to many questions

about emotions. For example, how do emotions come to arise? Are they purely non-cognitive mental states? If not, are non-cognitive mental states necessary in order for them to manifest? Do they have cognitive components which are also essential to their manifestation, and if they do, what is the relationship between the cognitive and the non-cognitive components? As I will discuss, the debate is further complicated by the fact that various theorists define 'cognitive' in different ways.

Given the strong emphasis on emotions in the psychological literature, this chapter explores what they are. I will discuss two highly plausible claims about the nature of emotions. Firstly, I will discuss the claim that non-cognitive mental states (or mental states with that property) are a necessary component of emotions. In other words, emotions have a world-to-mind component. Secondly, I will discuss the idea that emotions depend upon appraisals. Appraisals are the means by which we assess the significance of external states of affairs for things that matter to us. Commonly, appraisal theorists claim that, in addition to requiring a non-cognitive, world-to-mind mental state (or a mental state with that property), appraisals also involve a cognitive mind-to-world mental state (or a mental state with that property). The suggested distinction between mental states and states with certain properties reflects two common ways in which this appraisal process can be conceptualised which I will address later on. Finally, I will discuss how the meta-ethical theories of moral judgements described in Section 1.3 can explain the empirical and psychological literature regarding the place of emotions in moral judgements.

This chapter lays the foundation for Chapters 4 and 5 where I discuss three broad empirical methods that have been employed to investigate the relationship between emotions and moral judgements. Firstly, in Section 4.1 I will discuss the evidence suggesting that moral judgements and emotions regularly co-occur. I find the phenomenon of co-occurrence to be of minor significance for illuminating the concept of moral judgement. All of the meta-ethical theories outlined are able to explain the why moral judgements and emotions co-occur, some more straightforwardly than others. Nevertheless, there is something useful that studies of co-occurrence can tell us about the prevalence of emotions and moral judgements. Secondly, in the remainder of Chapter 4 I look at studies which induce emotions or some other feeling state, such as a sensation and measure the effect that the induced non-cognitive states or feelings can have on our moral judgements. The authors of these studies claim to empirically demonstrate that we use our feelings as information to decide whether something is morally wrong or right. As I will explain, some make a further claim that

these kinds of studies provide strong evidence for a theory of moral judgements which emphasises non-cognitive mental states or properties to moral judgements. These kinds of studies are taken to provide some of the most insightful empirical research on moral judgements. I argue that many conclusions drawn from these kinds of studies are unfounded. In Chapter 5 I turn my attention to the third broad empirical method which I take to be most useful from the point of view of adjudicating between theories of moral judgement. These studies have the potential to demonstrate that judgements that we ordinarily take to be moral ones can be made despite such judgements missing *moral content* of either a cognitive or non-cognitive kind. For example, moral dumb-founding studies claim to demonstrate that cognitive moral beliefs are (at least sometimes) absent despite apparent moral judgements. They emphasise the apparent authority afforded to gut-feelings in making moral decisions. These studies have sought to show that moral judgements have an essential feeling-based nature by demonstrating the lack of cognitive moral beliefs in some apparent cases of moral judgements. The studies also demonstrate missing non-cognitive content of a certain kind. I find that these studies do not contribute to debates between cognitivists and non-cognitivists, but they do have important implications for hybrid theories of moral judgement.

3.1. Emotions

The emotions of happiness, sadness, fear, anger, surprise and disgust exemplify a set of emotions widely taken to be universal (Ekman and Friesen, 1989).[1] There are many theories about what these emotions are and no general consensus on which, if any, theory is correct. This makes it extremely difficult to make sense of the empirical literature on moral judgements which focuses heavily on the role of emotions. For now I will leave questions about the relationship between emotions and moral judgements aside. I will return to this issue in Section 3.7. Now I will concentrate on discussing the nature of emotions. Emotion researcher Jesse Prinz (2004) has well formulated the issues and problems facing emotion researchers. He starts his inquiry by noting the many components that emotional experiences tend to have, including, thoughts, feelings, desires, attention, and action tendencies (Prinz, 2004, p. 3). Differing theories of emotion emerge from the commitments that various theorists make to the importance of the many components of

[1]Some authors question whether surprise and disgust should be considered emotions. The worry is that they differ significantly from the others in their degree of simplicity and should be thought of as sensations as opposed to emotions.

emotional experience, that is, from which components they take to be necessary and sufficient for an emotion. Some theories of emotion attempt to incorporate numerous components of emotional experience, whereas others focus on particular components which they consider to be essential (Prinz, 2004, p. 4). Prinz notes that those theories claiming that many features are essential must show how all of the features function as a whole, and theories which focus on particular features must show why those features hold a privileged status in explaining emotional experience (Prinz, 2004, p. 19). Despite the lack of a strong consensus in emotion research, there is some agreement about some aspects of emotional experiences. I will follow Prinz's lead in exploring these components in terms of asking whether certain parts of an emotion can be removed without 'losing the emotion' (2004, p. 4).

I will begin by highlighting some definitional issues. I will then discuss the role of feelings and appraisals in emotions and their relation to cognitive mental states, non-cognitive mental states, and hybrid mental states. In doing so, I will discuss the widely accepted idea that feelings (or dispositions to feelings) are an essential component of an emotion. Many feelings are taken to be, or to require, non-cognitive mental states.[2] If non-cognitive feelings are essential, this will provide some support for the strong claim that emotions are purely non-cognitive mental states; but it will provide conclusive support for the weaker claim that emotions at least have a necessary non-cognitive component, even if it is not sufficient. I will then discuss some reasons to think that emotions also have a cognitive mind-to-world component and involve appraisals. This sheds doubt on the strong claim that emotions are purely non-cognitive mental states. If emotions have both mind-to-world and world-to-mind components the question becomes one about whether the emotion is a single mental state with both directions-of-fit: a besire, or the product of a meeting of independent cognitive and non-cognitive mental states. If emotions require appraisals, the case for purely cognitive or purely non-cognitive theories of emotion is undermined.

3.2. Definitional discrepancies

Broadly speaking, theories of emotion tend to fall into one of three categories: there are non-cognitive theories, cognitive theories, and appraisal theories. It is often very difficult to label theories of emotion as there seems to be a great deal of inconsistency in the ways in which

[2]Keep in mind that sensations, such as itches, feeling hot or cold etc. are not world-to-mind mental states.

the labels are used in the literature. In particular, there seems to be no widely used and accepted definition of what is meant by 'cognitive'. Jesse Prinz describes the issue as being "frightfully unresolved" (Prinz, 2004, p. 45). How one conceives of 'cognitive' mental states will determine whether a theory of emotions is cognitive or not. Many theorists, embracing Anscombe's distinction between the directions-of-fit that mental states can have, think that a cognitive mental state is just one which has a mind-to-world direction-of-fit (such as beliefs or direct visual perceptions) and non-cognitive ones are those which have a world-to-mind direction-of-fit (such as desires or wishes). This way of distinguishing cognitive from non-cognitive mental states is common in the philosophical literature and is well defended by Michael Smith in *The Moral Problem* (Smith, 1994, p. 112; see also Schroeder, 2015; Rosati, 2016; Jacob, 2014).

I refer to theories of emotion which claim that emotions are either mind-to-world or world-to-mind mental states as pure cognitive and pure non-cognitive theories respectively. For pure cognitive theories, cognitive mental states are sufficient for emotion. Likewise, for pure non-cognitive theories, non-cognitive mental states are sufficient for emotion. Most modern philosophical theories of emotion are, however, appraisal theories (Sousa, 2017).

Appraisal theories generally take emotions to have both cognitive and non-cognitive properties in the way that I have been using the terms. Yet, there are cognitive and non-cognitive versions of appraisal theories. This reflects a shift in emotion research to disagreement over the cognitive status of the *appraisal process*. Appraisal theorists generally take one of two stances in conceptualising the appraisal process; they either claim that the appraisal is a single mental state with both directions-of-fit (a besire), or that the appraisal involves a cognitive process of representing the significance of distinct cognitive mental states for our non-cognitive ones, such as when one recognises the implications of a threat to their safety for their desire to survive. As noted, the former involves postulating a third kind of mental state: a besire. It is anti-Humean (and anti-Anscombian) as it denies the Humean theory of motivation which takes Anscombe's view that mental states have just one direction-of-fit. The besire view is generally taken to be a non-cognitive perceptual theory. The latter kind of appraisal theorist position maintains the Humean view, that the beliefs and desires are distinct entities which combine to provide motivation. It is common for appraisal theorists who think that emotions involve independent cognitive and non-cognitive mental states to describe their position as cognitive to emphasise the representational nature of appraisals.

That is, the appraisal can be said to be cognitive, despite involving a non-cognitive mental state, as an appraisal of the significance of independent cognitive mental states for non-cognitive ones has a mind-to-world direction-of-fit. To illustrate this, note that the appraisal can be made theoretically, that is, without actually holding the particular cognitive and non-cognitive mental states. Despite the emphasis on the appraisal, it is important to keep in mind that these appraisals depend upon typical non-cognitive mental states such as desires. Much of the debate between appraisal theorists is about the cognitive status of the appraisal process and whether the appraisals need be consciously processed. This will be further discussed in Chapters 5 and 6. This is a very basic outline of the ways in which major theories of emotion have conceptualised emotions in terms of other widely known kinds of mental states, and although it captures many theories, it does not capture them all.

The definitional discrepancies within the emotion literature reflect the discussion of possible mental states portrayed in Section 1.1. In summary, emotions may be cognitive (a mental state with a mind-to-world direction-of-fit), non-cognitive (a mental state with a world-to-mind direction-of-fit), or some kind of appraisal which involves both a cognitive and a non-cognitive mental state, or, a mental state with both directions-of-fit. The former appraisal position can retain the Humean theory of motivation, maintaining that emotions are the product of distinct cognitive and non-cognitive mental states (I sometimes refer to it as a Humean appraisal theory); the latter appraisal position is anti-Humean. This debate marks a shift in the emotion research to the question about the cognitive status of the appraisal process. It is important to keep in mind that any alterations to these conceptions of cognitive and non-cognitive mental states will potentially alter the label of one's theory of emotions.

3.3. Essential non-cognitive components

It is a commonly held view that feelings are an essential component of emotions. Feeling theories of emotion equate emotions with feelings (Sousa, 2017; Prinz, 2004, p. 4). They claim that even if emotions sometimes have other elements, for example, cognitive causes, they do not require any cognitive mental states or cognitive appraisals for their manifestation; feelings or non-cognitive mental states are sufficient. (Prinz, 2004, p. 4). Feeling theorists William James and Karl Lange independently proposed that our emotions are the perception of our physiological sensations (Prinz, 2004, pp. 4–5). According to James

(1884, pp. 189–190), "... bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion". These physiological states may arise from a number of bodily systems, such as 'the respiratory system, circulatory system, digestive system, musculoskeletal system, and endocrine system' (Prinz, 2004, p. 5). In his famous paper, 'What is an Emotion?', William James made the claim that common sense gets it backward: he said that ordinarily, we presume that we tremble because we are fearful when in reality, we are fearful because we tremble (1884, p. 190; also see Sousa, 2017). The importance of feelings to emotions is clearly demonstrated by James and Lange's mental subtraction argument which asks us to imagine an emotion, say fear, and then to systematically subtract any bodily feelings associated with that emotion, for example, we would subtract the elevated heart rate, high blood pressure and so on (Prinz, 2004, p. 4). Once we have done this, James notes, "we find we have nothing left behind, no "mind-stuff" out of which the emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains" (James, 1884, p. 193; Prinz, 2004, p. 5). The mental subtraction exercise aims to demonstrate that feelings are an essential component of emotions. Without them, the emotion ceases to exist.

Unsurprisingly, theories of emotion that emphasise the role of feelings are often favoured in virtue of their agreement with the intuitively appealing idea that feelings are essential to emotions. There are, however, important differences between the kinds of feelings that we experience. At face value, not all feelings are thought to be emotions. One essential important difference is between the feelings that we experience which involve no world-to-mind direction-of-fit and ones which do. The former are bodily sensations such as headaches, feeling hot or cold, itches, and tickles; the latter are expressions of our will such as desires, goals, attractions or aversions etc. The non-cognitive mental states such as desires and goals (amongst others) differ significantly from sensations because they have propositional content; we desire *that* something be the case. The non-cognitive mental states involved in emotions are expressions of the will; they are world-to-mind mental states, unlike sensations which are passive (sensations 'just are'). We can have non-cognitive attitudes about our sensations, for example, we can desire some sensations and desire the absence of others, but sensations themselves are not world-to-mind mental states. Sensations are relatively uncontroversially not taken to be emotions. I say 'relatively', as it is not obvious how a feeling theorist can distinguish emotions from sensations.

Standard non-cognitive mental states, such as attractions, aversions, desires, and goals all involve non-cognitive feelings. But, the claim that

these non-cognitive mental states *are* emotions is controversial. Many emotion theorists think that although non-cognitive mental states are essential to emotions, they are not sufficient. For appraisal theorists, an emotion such as fear involves a non-cognitive desire or drive such as one to survive, as well as a cognitive mental state or (a mental state which has a cognitive component) such as a belief or percept which, for example, represents states of affairs. The appraisal is of the implications of states of affairs for non-cognitive desires etc. Essentially, appraisal theorists think that both non-cognitive (world-to-mind) and cognitive (mind-to-world) components are needed to produce emotions. For besire advocates, the cognitive and non-cognitive components are inextricably linked in one mental state, and, for Humean appraisal theorists there is a cognitive representation of the significance of one kind of mental state for the other. The Humean theory of motivation says that motivation (emotions being motivating states) requires both cognitive and non-cognitive mental states. The anti-Humean conception of emotions as besires sees them as essentially motivating states, as besires are mental states with both directions-of-fit.

Both non-cognitive 'feeling' theories and appraisal theories of emotion claim that feelings are essential to emotions. Specifically, they both posit a world-to-mind mental state or state with that component which is intrinsically motivating. We can isolate the non-cognitive component (at least in theory). For example, someone can have a desire that we live in a democracy without any kind of cognitive representation of the kind of society that we do live in, or they may desire a drink of water without any cognitive beliefs about how to fulfil that desire, or the likelihood of being able to fulfil that desire. Desires have propositional content, but not content with a mind-to-world direction-of-fit understood as depicting the world as being a certain way. They are not truth-apt. If appraisal theories are true, we can always demonstrate the necessity of non-cognitive mental states (or components) in generating emotions. For example, if we take the emotion of fear, we can analyse any potential cognitive components out of the equation, such as the belief that one is in danger or the visual perception that one is in danger (see Smith, 1994, p. 117). Once we have done this we may be left with a non-cognitive mental state such as a desire or drive for survival (such talk is purely artificial in the case of besires as the components are inextricably bound). Without such non-cognitive desires or drives the cognitive belief or percept that one's life is in danger would not cause one to be motivated, fearful or similar. For Humean appraisal theorists, the emotions that arise as a result of the appraisal process necessarily depend upon the individual having non-cognitive goals, or desires, or

drives. In other words, the emotion of fear in response to some cognitive representation cannot occur unless the individual has non-cognitive goals, desires, or drives etc. The individual need not be conscious of the non-cognitive goals driving their behaviour, they are nevertheless there (Smith, 1994, p. 107). Appraisal theorists deny that emotions are purely non-cognitive, yet they, like non-cognitive theorists, think that a non-cognitive mental state (or state with that quality) is essential. Without it, the emotion ceases to exist.[3]

To summarise, for appraisal theorists, non-cognitive mental states (or mental state components) are a critical part of the appraisal process. Non-cognitive mental states such as desires, drives, and goals all express the will of an individual, as do besires which have an essential world-to-mind component. I will now move on to discuss the claim that emotions also have a mind-to-world component as both Humean and anti-Humean appraisal theorists claim.

3.4. Emotions and appraisals

According to appraisal theories of emotion, non-cognitive feelings are not sufficient for an emotion. Although non-cognitive feelings may have a phenomenology to them, there must also be a related cognitive component to create the kind of phenomenology distinct to emotions. Appraisals are generally taken to be "...representations of organism-environment relations with respect to well-being" or something similar (Prinz, 2004, p. 52). For Prinz, "...to appraise something is to see it as affecting oneself in some way that matters" (Prinz, 2004, p. 14). Prinz says that appraisal theories appeal to us because "...there is a deep intuition that emotions are meaningful. They are not simply arbitrary feelings. Instead they inform us about our relationship to the world, they embody our convictions, and they factor intelligibly into our decisions in life" (Prinz, 2004, p. 16).

As I have already highlighted, there are two ways of explaining the way in which the world-to-mind and mind-to-world components are involved in an emotion. Firstly, we could conceive of the relationship in Humean terms as some sort of pairing of a cognitive mental state, such as a belief or percept, and some sort of non-cognitive mental

[3]Pure cognitive theories of emotion claim that emotions are identical to thoughts, such as the thought that x is good. If we accept that non-cognitive feelings are essential for emotions this undermines pure cognitive theories of emotions which deny that non-cognitive states are necessary. The lack of a state with a world-to-mind direction-of-fit has led cognitive theories of emotion to be criticised for divorcing emotion from feeling, as, on the Humean view, cognitive mind-to-world mental states, on their own, are not motivating.

state, such as a desire or drive. Or, secondly, we can invoke the anti-Humean idea that emotions are besires, which are unitary mental states with both directions-of-fit that are not decomposable into cognitive and non-cognitive parts. Although we can speak of the world-to-mind and mind-to-world elements involved in a besire, the components are inseparable and unalterable in their expression. Some proponents of this view argue that emotions are a form of perception (Prinz, 2004; Kirchin, 2012). This difference marks a distinction between those who claim that emotions *are* appraisals, and those who claim that emotions *follow* appraisals. The former view is extolled by besire theorists; the latter is often embraced by Humean appraisal theorists. Prominent appraisal theorists, Magda Arnold and, following on from Arnold's work, Richard Lazarus (1991) argue that emotions are the result of recognising that states of affairs are congruent or incongruent with our goals. For them, "something would not count as an emotion if it were not the result of a dimensional appraisal process. But that does not mean that emotions are appraisals... Arnold and Lazarus do not say that emotions should be identified with the cognitive appraisals they postulate. They imply that those appraisals are causes, not components, of emotions. At the same time, they regard appraisals as necessary preconditions for emotions" (Prinz, 2004, p. 17).

Why should we prefer appraisal theories to non-cognitive theories? The equating of emotions with non-cognitive bodily feelings has been heavily criticised as not doing justice to the phenomenon of emotions (Sousa, 2017). It is common to evaluate emotions in terms of whether they are appropriate or inappropriate, justified or unjustified. Feeling theories have been criticised as not having the tools to explain such practices. I will, in part, defend feeling theories on this front: non-cognitive mental states can be evaluated as appropriate or inappropriate in some sense. However, I agree that there are other key practices that feeling theories are not able to account for, in particular, the practice of speaking of emotions as justified or unjustified.[4]

As I have already mentioned, there is an essential difference between non-cognitive desires or goals etc. and sensations, such as itches and tickles. The former have propositional content; we desire *that* something be the case. Non-cognitive mental states are expressions of the will, sensations are not. I do not think that this particular distinction between non-cognitive desires and sensations does anything to legitimise the way we talk about emotions in comparison to sensations. We can, as an

[4]Please take the meanings of the terms 'appropriate', 'inappropriate', 'justified', and 'unjustified' used here as a suggestion.

observer, appraise the desires or the sensations of others in terms of our own desires; in terms of desires that we know the sensation experiencer to have; or in terms of ones that we think the individual ought to have. For example, we can say that a pleasure response to bodily injury that an individual experiences is inappropriate or non-conducive to some goal, such as a goal of mine, a goal of the individuals, or a goal I think that they ought to have. We can criticise somebody's distaste for bugs in a food shortage, or their tendency to be sexually attracted to children in terms of our own goals for survival or to see children flourish. Importantly, in these cases we are not assessing the truth of either the sensations or the desires that the individual has (they are not truth-apt). We are assessing the implications of some phenomenological sensations and non-cognitive desires for their congruency with other non-cognitive desires. In other words, both feeling theories and appraisal theories can be criticised in this way as they both involve mental states which are not truth evaluable (sensations and non-cognitive mental states or components). They may only be evaluated in terms of their implications for one's own besires, non-cognitive goals or desires, or the goals and desires of others.[5] One might be tempted to criticise this view by pointing out that, for example, someone's emotion, such as anger, can be justified and that what I have said so far doesn't allow for that.[6] Keep in mind that I am strictly talking about *non-cognitive* mental states here, and it is questionable if such emotions are purely non-cognitive. If emotions involve appraisals (and so cognitive mental states), emotions have a greater potential for criticism than I have described here; criticism that non-cognitive theories do not have the potential to account for.

Our emotions typically have causes which extend outside of us; we refer to external objects or states of affairs to justify our emotions (Sousa, 2017). It is not clear upon what basis such talk could be justified if emotions were (usually) purely non-cognitive (Sousa, 2017). Appraisal theories recognise a role for cognitive mind-to-world mental states (or components) in the formation of an emotion. An individual's emotional responses can, on this view, be criticised in a way that non-cognitive theories of emotion cannot be. In particular, the cognitive

[5]Although some non-cognitive theories of moral truth will privilege the non-cognitive states of some individual or culture etc., letting them set a standard of truth to which others can be measured, this is not the sense of truth which I am talking about here.

[6]The limits of criticism described here are incorrect if beliefs and desires are not distinct mental states as the internal cognitivist claims (that is, if some of our desires are necessarily connected to beliefs).

mental state (or component) is truth evaluable. We can ask whether the cognitive mental state (or component) accurately represents the world. For example, the belief or the percept that is involved in the appraisal may be false. Both Humean and anti-Humean appraisal theorists can concede this point. That is, the truth of the cognitive aspect can be questioned on both accounts.

3.5. The nature of appraisals

For Humean appraisal theorists there is one more potential avenue for evaluating emotions that is not available to anti-Humean theories of emotion. Humean appraisal theorists claim that emotions are the outcome of an individual recognising the significance of one of their mind-to-world, cognitive mental states for one of their world-to-mind, non-cognitive mental states. For example, the belief that x, and the desire that not x, will result in an emotion, should one see the implications of some state of affairs for their non-cognitive desires or goals. The significance of one kind of mental state for the other must be comprehended or 'appropriately related': "To say that the desire and belief must be 'appropriately related' is merely to acknowledge that in order for a desire and belief to constitute a motivating reason the agent must, as it were, put the relevant desire and belief together" (Smith, 1994, p. 92). To take an example from Smith (1994), to be motivated to buy a Picasso painting in front of me I require both the desire to own a Picasso as well as the belief that the painting in front of me is a Picasso. Should I lack one (belief or desire) I will not be motivated to buy the Picasso in front of me. In the moral case, one may have the desire for well-being and the belief that there is presently a threat to well-being which may generate, for example, the motivating emotion of fear. An individual can have a cognitive mental state with implications for one of their non-cognitive mental states without 'putting the two together'.

The appraisal process itself, recognising the implications of cognitive mental states for non-cognitive ones, is also a cognitive, mind-to-world one, on this view. It is a truth-evaluable mind-to-world representation about the implications of states of affairs for goals. If they are goals that we have, they will be motivating, if they are not, they will not be. For Humean appraisal theorists, the question of whether such cognitive mental states are relevant to our non-cognitive mental states can be evaluated in terms of whether the linking is justified or not: that is, are the desire and belief 'appropriately related', or is somebody mistaken in their inference? Phobic responses are obvious examples of

cases that we take to involve mistaken inferences, i.e. at some level, we mistakenly represent something as being a threat to one of our goals, or we mistakenly represent the severity of a threat. Advocates of besires claim that the appraisal is a unitary mental state with both directions-of-fit. There is no combining of two independent mental states: the appraisal *is* the emotion. Consequently, this second avenue of criticising emotions is not available to advocates of besires. There are not two independent mental states which are put together, rather, there is one, which is unalterable in its expression; there is no room for error in making an inference as there is no inference. Besires can be appropriate or inappropriate; they can be assessed by how well the cognitive component represents the world; but, as there is no cognitive appraisal connecting beliefs to desires, the appraisal cannot be justified or unjustified.[7]

In this section my aim is to discuss some potential ways in which the debate between appraisal theorists, about the kind of thing an appraisal is, may proceed. Above, I highlighted one significant consequence of falling one side over the other on this matter; that is cognitive (Humean) appraisal theorists have, at their disposal, an additional avenue of evaluating emotions which is not available to the besire theorist. This section is not supposed to be a defence of Humean appraisal theories. I intend only to highlight some of the issues that need to be considered in relation to besires. I can see two possible reasons one might wish to invoke besires. Firstly, the notion of a besire might seem to be supported based on the speed at which appraisals take place; and secondly, there may be mental states which contain world-to-mind and mind-to-world components which are not cognitively penetrable: the parts do not ever come apart from one another. Even if besires do exist, it is a further question whether they should be considered *moral* judgements. This further question is taken up in later chapters.

The speed of emotions. Emotions, and so appraisals, can and often do take place extremely quickly and with little (or no) higher conscious thought. Think of the fear one experiences when confronted with a snake, or with extreme heights. The very simple cognitive representation (or perception) of a snake can effortlessly elicit fear which may be mediated by a very simple biological drive towards survival. If these emotions depend upon appraisals, introspection would suggest that those appraisals are, at the very least, only minimally consciously taxing. Phenomenologically, it even seems possible that the appraisal,

[7]Again, the distinctions drawn here between appropriateness, inappropriateness, justified and unjustified are a suggestion only.

if it happened, occurred entirely unconsciously, with only the emotion presenting itself in consciousness. The high speed at which the appraisal appears to take place is what attracts some people to a hybrid-state theory of emotion. Hybrid-state theories are sometimes referred to as perceptual theories, emphasising the immediacy of the emotion, whereby perception and feeling appear to occur simultaneously, unmediated by any cognitive appraisal. But, although the process is fast and possibly largely unconscious, it is possible that the emotion is still best explained in terms of a combination of separable cognitive and non-cognitive processes rather than as a besire. Humean appraisal theorists think that this connection is always mediated by a cognitive appraisal of the significance of independent cognitive mental states for non-cognitive ones. The term 'alief' is sometimes used to describe an emotion resulting from a strong, fast connection that occurs between some cognitive and non-cognitive mental states. Humean appraisal theorists maintain that although some appraisals are strong and relatively reflexive, they are still a product of a combinational process, the parts of which are independently alterable (see Gendler, 2008; Zabzebski, 2003; Kriegel, 2012).

The process of a cognitive and a non-cognitive mental state combining can be quite easily demonstrated when we examine more abstract elicitors of emotions. Think of experiencing fear when confronted with an unopened letter, the contents of which will in some way majorly impact upon ones future. There are clearly quite abstract beliefs involved in this case, as well as rather abstract goals. For example, the letter may contain an anticipated response to a job application. There is no automatic connection between perception and feeling in this case. In theory, either part of the process can alter while the other part remains constant, for example, maybe your beliefs about the contents of the letter remain constant while your desire for the job alters, or perhaps your desire for the job remains constant while your beliefs about the content of the letter alters. There is little reason to think that these kinds of appraisals do not take place at great speed. For simplicity's sake, if we can explain motivation and emotional experience without positing besires, we should. Defending besires via an appeal to speed is very difficult. There is another, more fruitful, method of investigating the existence of besires which examines the cognitive penetrability of the mind-to-world and world-to-mind components of emotions.

Cognitively impenetrable mental states which have both directions-of-fit. The case for besires best rests on the possibility of cognitively impenetrable mental states with both directions-of-fit.

That is, doubt would be cast upon the basic cognitive/non-cognitive dichotomy of mental states if we were to find cognitively impenetrable mental states, whereby the cognitive and non-cognitive components of the mental state are not independently alterable. This possibility seems to be behind Jesse Prinz's favoured conception of cognition. Prinz argues that "...cognitive states and processes are those that exploit representations that are under the control of an organism rather than under the control of the environment" (2004, p. 45). On this view, for emotions to be cognitive, the appraisal process must involve manipulation and control by executive systems in the brain. Prinz thinks that emotions can arise without such cognition, so although appraisals may involve perceptions, mind-to-world representations with the potential to inform us about our relationship with the world, they are not cognitive (2004, p. 45). The lack of this cognitive process is what prompts Prinz (and others) to describe emotions as a form of perception. And, it is related to (perhaps even created to highlight) the idea that at least some emotions are cognitively impenetrable. This debate is between Humeans and anti-Humeans, whereby anti-Humeans (advocates of besires) claim that some emotions are single mental states; it is impossible for the cognitive and non-cognitive elements to ever be pulled apart and Humeans deny that such unitary mental states exist (Smith, 1994, p. 119).

There are, no doubt very strong connections between some cognitive representations and some non-cognitive mental states. Phobias are promising candidate besires. As are some cognitive representations linked to biologically based drives that have been evolutionarily instilled within our species. The desire or drive to survive produces some of the most robust cognitive/non-cognitive connections that people experience. Many forms of entertainment, such as theme parks, glass bridges, and horror movies, capitalise on these robust appraisals. For Humean appraisal theorists, the connection between representations and drives can come apart. It seems that aversions to threats to life can come apart in cases of suicide, or for individuals who engage in extreme thrill-seeking behaviour. And, in the case of phobias, the connections between the stimulus and response appear to be capable of being successfully pulled apart using cognitive behavioural therapy. The fact that, at least often, such connections can be broken, or can change from one time to the next suggests that there are two independent and variable mental states in play. Cordelia Fine has noted that emotions which arise relatively reflexively within us plausibly have their origins in slower, more effortful cognitive appraisal processes. The process of learning to drive a car, or to play the piano, is often used to demonstrate how many

tasks or problems that we face are originally very cognitively taxing, and become automatic only through cognitive effort. She speculates that in the case of relatively reflexive moral judgements it is plausible to think that conscious reflection or reasoning has taken place at a previous time in an individual's life (Fine, 2006). The process of atomisation supports the possibility that emotions which seem like besires are actually the result of a cognitive appraisal process. Besire advocates would need to show that there never was such a process and that the relatively strong connections are unalterable; that we have world-to-mind and mind-to-world connections which never come apart which are better described as besires.

The phenomenon of variation in motivation despite seemingly unaltered cognitive mental states also suggests penetrability, or the working of two independent mental states in creating an emotion. It appears that in cases of variation in motivation that one mental state can change whilst the other remains constant. Smith claims that the notion of besires is inconsistent with such variation as besire theorists would be forced to say of individuals whose motivation has altered (other things being equal) that they must have forgotten something that they used to know (1994, p. 123).

Disanalogy. The cognitive impenetrability of besires has been likened to the cognitive impenetrability of some perpetual illusions, such as the Muller-Lyer illusion. This analogy is, in my view, mistaken. The cognitive percept (viewing of the Muller-Lyre image) is cognitively impenetrable; one cannot help but see the lines in the image as differing lengths. Yet one can hold the cognitive belief that the lines are the same length. There is no non-cognitive, world-to-mind mental state (or component) involved in this example. What we have is two representations of the world, one of which we disregard despite appearances. Both are on the cognitive mental state spectrum. The same can be said of hallucinations and dreams, for example one can believe them to be true of the world until one forms other cognitive representations which indicate otherwise. Believing that one is experiencing a hallucination does not make the hallucination go away. An analogous case in the moral realm may be the cognitive impenetrability of one method of forming cognitive beliefs. For example, one may have the cognitive moral belief that killing an innocent child is the right thing to do in a particular circumstance. Yet, killing an innocent child may always *appear visually* wrong to them, even in cases where killing an innocent child might be the morally right thing to do. That is, one may *never* be able to visually perceive the killing of an innocent child as right, however, the cognitive belief that it is right in a particular case to kill

a child may arise, should one, for example, believe that 1000 children will die should the one child not be killed. In both this example and the Muller-Lyer illusion the *cognitive percepts* are cognitively impenetrable, but there are other cognitive representations indicating that the percepts are inaccurate leading to the percepts being disregarded. This demonstrates the limits of some of our means of forming beliefs about the world, in this case, the limits of perception. Fortunately we have several means of forming cognitive beliefs. In the above examples, one method is taken to provide a more accurate representation of the world. This is not the kind of support that a theory of besires requires. Besires require that some cognitive property and some non-cognitive property are inseparably bound. It is their connection that must be cognitively impenetrable.

3.6. Emotions summary

James and Lange's mental subtraction argument provides good reason to believe that emotions have non-cognitive feelings as essential components. Both non-cognitive theories and appraisal theories of emotion allow for this as desires and goals are quintessential non-cognitive mental states. Whether such non-cognitive states are sufficient for emotions is contested. The class of emotions may also require a cognitive, mind-to-world component. It seems very plausible that emotions are the result of an appraisal, be it a besire or Humean combinational appraisal process. Imagine a parent thinking of their child graduating and that this thought generates positive affect in the parent in the form of a desire or wish. Such a state can be paired with numerous cognitive mind-to-world mental states, such as, the cognitive percept of the graduation invitation on the fridge, the belief that one's child passed away in a car accident, or the belief that the child may not complete their studies due to their drug addiction. The various cognitive mental states all have the effect of influencing the emotion that the parent experiences, be it pride, sadness, or fear. Non-cognitive theories of emotion have been criticised as not having the resources to explain the ways in which we think and talk about emotions, particularly if we think that emotions can be justified or unjustified. Appraisal theories can meet this challenge to varying degrees, with Humean appraisal theories of emotions allowing for the greatest scope for evaluating emotions. With these descriptions of the nature of emotions and the problems associated with the various accounts in mind, I will now move on to describe the ways in which moral judgements may be related to emotions.

3.7. Theories of moral judgement and emotions

In this Section I will explain the ways in which the meta-ethical theories of moral judgements discussed in Section 1.3 can explain their relationship to emotions. This will better place us to evaluate the experimental literature. As I will explain, when thinking about conceptions of moral judgements and their relationship to emotions it is essential to take note of where each theory says that the *moral content* constituting the moral judgement is located.

External cognitivism. The external cognitivist claims that moral judgements are cognitive. They usually claim that moral judgements are beliefs, such as beliefs about what is right or wrong, good or bad and so on. For external cognitivists who subscribe to the Humean theory of motivation, moral beliefs are not motivating on their own. But they may be motivating should they be relevant to the believer's external non-cognitive desires. The external cognitivist does not think that there are any necessary connections between any cognitive mental states and any non-cognitive mental states. On this view, emotions that are relevant to morality can occur when an appraisal takes place which pairs a cognitive moral judgement with an external non-cognitive mental state such as a desire, wish, or goal. The mental states must be appropriately related; should the holder of a moral belief and a desire put the two mental states together (make an appraisal), the cognitive moral judgement may contribute to the individual experiencing an emotion. Importantly, the locus of the moral judgement occurs within the cognitive domain. If moral beliefs regularly combine with relevant desires, moral judgements may be cognitive, yet still highly correlated with emotions. The non-cognitive desire or goal need not be in any way related to morality. In terms of explaining an individual's moral decision, the interesting point of analysis is their cognitive belief.

Internal cognitivism. Like external cognitivists, internal cognitivists also think that our moral judgements are cognitive. Unlike external cognitivists, internal cognitivists think either that beliefs motivate or that there is a necessary connection between some cognitive and some non-cognitive mental states. In particular, they think that some cognitive moral beliefs necessarily lead people to develop related non-cognitive desires. As discussed in Section 2.4, the internal cognitivist can think that the pathway from cognitive moral judgements to non-cognitive mental states is either conditional or unconditional. The presence of both cognitive and non-cognitive mental states has the potential to generate emotions. Like external cognitivism, the locus of the moral

judgement is within the cognitive belief; that is, the moral content is cognitive. The internal connection of moral beliefs to desires generates the possibility for individuals to regularly experience emotions in regards to their cognitive moral judgements. This view is similar to the above externalist position, apart from the denial that beliefs and desires are distinct entities. To illustrate, cognitive moral beliefs, on this view, such as the belief that killing is wrong will either conditionally or unconditionally lead the belief former to form related desires, the combination of which will produce some kind of emotion.

Non-cognitivism. For non-cognitivists, moral judgements are expressions of our non-cognitive mental states, such as our goals or desires. Such non-cognitive mental states can be sufficient for moral judgements. It is possible to hold a non-cognitive theory of emotions. That is, one may claim that non-cognitive desires are sufficient for an emotion. But, if one holds an appraisal theory of emotions, the non-cognitive moral judgement will not itself be an emotion. Appraisals involve recognising how mind-to-world beliefs (say factual beliefs) bear on our non-cognitive goals and desires. When this happens we experience an emotion. For example, an individual may have a cognitive belief that some policy will result in people being harmed. The recognition of the implications of this belief for that individual's desires may generate an emotion. Despite any appraisals that may occur, the moral judgement (mental state with moral content) is contained within the non-cognitive mental state. That is, although appraisals and so emotions may regularly occur, the moral judgement is reducible to the individual's non-cognitive mental states. This claim is the basis of the common assertion that moral judgement bottoms out in affect. That is, at the base of a moral appraisal is a non-cognitive moral judgement, such as a desire, which is not truth-apt.

Hybrid-state theory. For hybrid-state theorists, moral judgements are unitary mental states which have both cognitive and non-cognitive content. This state, according to some, is an emotion. As discussed, the relationship between the mind-to-world and world-to-mind components in this mental state must be cognitively impenetrable. Such theories of moral judgements are sometimes described as perceptual theories based on the idea that some cognitive mental states are inextricably linked to some non-cognitive ones in a way that, like perception, presents itself to us effortlessly. There is no *cognitive* appraisal that takes place. If this theory of moral judgement is correct, then moral judgements *are* emotions.

Hybrid-expressivism. According to hybrid-expressivists, moral judgements express both cognitive and non-cognitive mental states. Such combinations of mental states do have the ability to produce emotions if they are appropriately related. They may be somehow appropriately related on their own; or one of the cognitive or non-cognitive components comprising the moral judgement may be appropriately related to a third belief or desire which generates an emotion. Hybrid-expressivists vary in the kinds of content that they claim is involved in contributing to a moral judgement. For example, an individual may believe that killing is wrong and they may have a negative attitude towards killing. These cognitive and non-cognitive mental states may not produce an emotion on their own, but may do so when a third belief or desire enters into the picture, for example, the belief that someone is about to be killed may motivate in conjunction with the moral desire. On this view, individuals may experience an emotion in a sense which is adequate for a cognitivist or a non-cognitivist, but which do not reach the standards of what it is to be a moral judgement according to hybrid-expressivists.

Moral thought pluralism. According to moral thought pluralists, moral judgements can be either cognitive or non-cognitive (or both, as the hybrid-expressivist described). And, emotions can occur as either the cognitivist (external) or non-cognitivist described (or as the hybrid expressivist described). That is, emotions may be the outcome of appraisals that contain either cognitive or non-cognitive moral judgements.

Other related terminology. Appraisal theories of emotion state that emotions are assessments of the significance of states of affairs for things that matter to us, and can be conceptualised as a non-decomposable besire, or as a decomposable cognitive representation of the relevance of cognitive beliefs for our non-cognitive mental states. In the latter case, theorists may think that we can locate *moral content* in one of the mental states comprising the emotion. Individuals who identify as rationalists sometimes claim that the moral judgement is rational because the appraisal contains cognitive moral content. They also sometimes identify as rationalists by claiming that the *appraisal process* is rational. They endorse a Humean theory of motivation.

Sentimentalists, on the other hand, emphasise the role of non-cognitive moral judgements in appraisals. Sentiments are generally construed as a complex emotional disposition, a feeling of approbation or disapprobation, or a desire that we desire. Importantly, on the sentimentalists' view a sentiment is more akin to what we, *in fact*,

most desire, not what we have *most reason* to desire. Sentiments are sometimes referred to as grounding values based on the idea that they are not amenable to reason; they are cognitively impenetrable. If sentiments are non-cognitive, then they are not emotions, if an appraisal theory of emotions is correct.[8]

Another key term heavily related to emotions is 'intuitions'. I will make a distinction here between cognitive moral intuitions and appraisal intuitions. *Cognitive moral intuitions* have a purely mind-to-world direction-of-fit. You can think of them as being like hunches about what you take to be true or false. These intuitions are not emotions if an appraisal theory of emotions is correct as there is no world-to-mind (non-cognitive) mental state involved. *Appraisal intuitions*, on the other hand, are emotions about what is right, wrong, good or bad in the absence of reasons. If an appraisal theory of emotions is correct, these intuitions can occur when the appraisal process occurs unconsciously delivering an emotion into consciousness. Perhaps the appraisal components can become accessible to consciousness upon reflection, but perhaps not. This issue will be discussed further in the chapters that follow.

Believing that these processes of intuition occur does not make one a *moral intuitionist*. A moral intuitionist claims that moral judgements *are* moral intuitions. These moral intuitions are not thought to be the product of *justifiable* beliefs, or of combinations of moral beliefs or desires. A cognitive moral intuition is slightly different from a cognitive intuition. A *cognitive intuition* is a weakly supported belief. For example, one may have a cognitive intuition that God exists, or that some scientific hypothesis is true, or even that some act is morally bad. Cognitive intuitions are, in principle, empirically verifiable. *Cognitive moral intuitions*, on the other hand, are not empirically verifiable according to intuitionists. For cognitive moral intuitionists, justification is unavailable in principle; one *cannot* empirically verify a cognitive moral intuition. Intuitionists appeal to the intuition's feeling of truth or falsity as evidence for the belief's truth or falsity. This cannot be done to justify cognitive intuitions, e.g. my feeling that some belief is true or false has no bearing on whether that belief is true or false. Appraisal intuitions, as explained, are the intuitions one has when an unconscious appraisal generates an emotion. It appears that some intuitionists take this emotion as a starting point and claim that it is not, in fact,

[8]The class of sentimental judgements is smaller than that of non-cognitive judgements i.e. there are some non-cognitive judgements which do not qualify as sentiments and so not as moral judgements (e.g. likes/dislikes may not be moral judgements whereas approval/disapproval may be).

decomposable. That is, they seem to deny that the emotion (or feeling) that something is right or wrong is the outcome of an appraisal which *contains* a moral judgement (a cognitive or non-cognitive mental state with moral content). I will call this view *anti-appraisal intuitionism*. In essence, it says that an emotional feeling of rightness or wrongness is a moral judgement regardless of how one comes to form that intuition. The significance of the ways in which intuitions are conceived will become apparent in the chapters that follow.

3.8. Concluding remarks

Among psychologists, there is a commonly accepted view (perhaps tacitly accepted) that emotions are, at least in part, non-cognitive, and that moral judgements, being emotional in nature are also non-cognitive. If emotions depend upon, or are, appraisals which consist of both cognitive and non-cognitive mental states (or a state with both properties), some formulation of their first assumption is true. Significantly, though, the assumption that *moral judgements* are non-cognitive (based on their relationship to emotions) is questionable. Appraisals can be conceived of in either Humean, or anti-Humean terms. If they are conceived of in Humean terms, the moral content in an appraisal can be either cognitive or non-cognitive. That is, the moral judgement or moral content of the appraisal can be located within either the cognitive or the non-cognitive mental state. On this view, some emotions (appraisals) depend upon moral judgements which may or may not be non-cognitive.

The following is a brief overview of the ways in which emotions and moral judgements may be related with the assumption that emotions are, or depend upon, appraisals.

(1) Emotions are caused by an appraisal of the significance of cognitive moral judgements and their relation to non-cognitive mental states. Emotions involve a non-cognitive mental state, but the moral judgement is a cognitive mental state with moral content such as a belief. The appraisal isn't necessary to the moral judgement; a cognitive moral belief is sufficient. Such positions include external cognitivism and internal cognitivism. These positions are also sometimes referred to as 'Rationalist' positions.

(2) Emotions are caused by an appraisal of the significance of beliefs for our non-cognitive moral judgements. Appraisals involve a cognitive mental state(s), but the moral judgement is fundamentally non-cognitive. The appraisal 'bottoms out' in

affect. The appraisal is not necessary to the moral judgement; a non-cognitive moral goal or sentiment etc. is sufficient. Such positions include non-cognitivism and Humean sentimentalism.

(3) Emotions *are* moral judgements. An emotion is an appraisal; a single mental state with both cognitive and non-cognitive properties: a besire. Emotions are essential to moral judgements. This position can be referred to as an anti-Humean sentimentalist one.

(4) Emotions can be caused by moral judgements. Moral judgements express cognitive and non-cognitive mental states. Depending on the particular kind of hybrid-expressivist theory, these states may or may not be appropriately related in a way that is necessary to produce an emotion. If they are not appropriately related a third mental state may be appropriately related to one of them resulting in an emotion.

(5) Emotions are caused as described in positions (1) or (2). This is moral thought pluralism. Only one mental state involved in an appraisal needs to have moral content such that (1) and (2) can qualify as moral judgements independently.

Having discussed the nature of emotions and their relation to moral judgements, I will now move on to describe and evaluate the attempts of psychological research to shed light on moral judgements. I work with the assumption that an emotion is an appraisal of some kind.

CHAPTER 4

The prevalence and causality of emotions

In this chapter and the next I discuss three broad empirical methods that have been employed to investigate the relationship between emotions and moral judgements. The first method draws on the correlational research between moral judgements and emotions. The second method discusses the significance of research which demonstrates that we consult our feelings in order to make moral decisions. And the third method examines the content of the mental states involved in appraisals. This chapter concentrates on the first and second methods and Chapter 5 discusses the third method which I take to be the most useful for informing philosophical questions about the nature of moral judgements. The second method is sometimes portrayed as the gold standard of empirical research into the nature of moral judgements. This is because it is the only kind of study belonging to experimental psychology which means that we can be certain about causation. By contrast, the correlational studies of the first method and the examination of the content of mental states in the third method are descriptive studies and so, cannot establish causation. Despite this, I take the third method to be most useful for informing philosophical questions about the nature of moral judgements, which is why they feature in their own chapter. As I will explain, I give lower precedence to studies capable of revealing causation for two reasons. Firstly, the evidence that affect influences moral judgements is not particularly strong. Secondly, moral judgements may still be cognitive even if they are influenced by affect. This chapter finds that the ability of popular psychological studies and studies which examine causation to shed light on the nature of moral judgements is minimal.

4.1. The prevalence of emotions in moral life

A wide variety of methods have been employed to investigate the link between emotions and moral judgements. Greene et al. (2001) began the psychological trend of investigating the neural correlates of moral judgements using brain imaging data. Some other early studies include those by Moll, Oliveirra-Souza and Eslinger (2003) who found

that emotional centres of the brain are active when people evaluate moral sentences and absent when they evaluate factual sentences, and Sanfey et al. (2003) who found that emotional centres of the brain are active when people encounter unfair offers in Ultimatum Games.[1] Since these early studies, brain-imaging data documenting the co-occurrence of moral judgements and the activation of known emotional centres in the brain has been extensively investigated. Jesse Prinz has collated an impressive array of neuroimaging studies which strongly converge on the view that moral decision making involves brain structures associated with emotion (Prinz, 2016, p. 45). In addition, two significant meta-analyses on the topic agree that evidence for the co-occurrence of emotion and moral judgements is strong (Landy and Goodwin, 2015; Chapman and Anderson, 2013).

Further evidence of a correlation between moral judgements and emotions has been found in numerous behavioural studies. Introspectively, moral judgements are emotionally salient. Morally bad actions tend to make us angry or disgusted when performed by others, and ashamed or guilty when performed by ourselves. Phrases such as 'it makes my blood boil', 'my heart skipped a beat', 'you make me sick', 'they acted in bad taste' and 'that's disgusting' are commonplace in the context of moral transgressions. That emotions and moral judgements co-occur is consistent with the heavy use of physiological expressions to describe one's attitude towards moral transgressions.

The following studies have investigated the prevalence of disgust related terminology in the moral domain, finding that people spontaneously report moral transgressions such as racism and child abuse as disgusting when they are asked to list disgusting stimuli (Haidt et al., 1997; Curtis and Biran, 2001). Similarly, people report feeling disgusted by moral transgressions (Haidt and Graham, 2007; Horberg et al., 2009; Hutcherson and Gross, 2011; Simpson et al., 2006). The tendency to describe moral transgressions as disgusting has also been observed in children (Danovitch and Bloom, 2009). Relatedly, word-association studies which require experimental participants to complete word-stems (partially filled in words which can be completed to create different words) reveal that being exposed to moral transgressions leads people to complete word stem tasks with disgust and cleanliness related words, suggesting that moral transgressions prime people to bring to mind disgust related concepts (Zhong and Liljenquist, 2006; Jones and Fitness, 2008). Chapman et al. (2009) examined EMG activation of the levator

[1] For an early significant attempt to interpret these psychological findings see Greene and Haidt (2002).

labii superioris (LL) which, they note, is the muscle capable of causing a raised upper lip and a wrinkled nose which are facial movements characteristic of a disgust facial expression (Chapman and Anderson, 2013, p. 310). They found that the LL was activated when people consumed distasteful liquids, viewed physically disgusting photographs, and when they received unfair offers in economic games (Chapman and Anderson, 2013, p. 310;Cannon, Schnall and White, 2011; Whitton et al., 2013). In these studies, higher levels of inequality were correlated with higher levels of self-reported disgust and more pronounced facial displays of disgust (Chapman et al., 2009).[2]

Together, brain imaging studies and behavioural research overwhelmingly support the claim that emotions and moral judgements regularly co-occur. As I have already discussed in Chapter 3, all of the meta-ethical theories can explain the co-occurrence of moral judgements and emotions. Jesse Prinz has contributed greatly to this research. Following is an excerpt depicting his view of the limits of the correlational data for determining whether emotions are inputs to, outputs of, or identical to moral judgements:

> ...extant studies make little progress adjudicating between the theories outlined here [he refers to theories which claim that emotions are inputs to, outputs of, or identical to moral judgements]. Notice that every theory supposes that emotions regularly arise in the context of making moral judgements. Every theory also supposes that non-emotional aspects of cognition are involved (e.g., we can't morally evaluate a bit of conduct without first representing that conduct). Disagreements concern the role and ordering of these components. The problem is that extant studies shed too little light on those questions. They show that "emotion areas" of the brain are active during moral cognition, and they also regularly implicate brain structures that are not presumed to [be] emotion areas. But they tell us little about how these relate. To put it bluntly, every model presented here [emotions as either inputs to, outputs of, or identical to moral judgements] is consistent with every study cited

[2]Landy and Goodwin (2015) note that it is unlikely, but possible that such facial expressions may serve to indicate to others one's disapproval as suggested by Rozyman and Kurzban (2011a) and Rozyman and Kurzban (2011b).

> in the previous subsection [depicting a correlation of emotions and moral judgements]. (Prinz, 2016, p. 52)

To interpret the significance of the findings that emotions and moral judgements regularly co-occur will require a theory of the nature of emotions. Even if some appraisal theory of emotions is correct, we will need to know whether emotions are besires, or the product of a combination of distinct cognitive and non-cognitive mental states. That is, is the appraisal Humean or Anti-Humean? If it is Humean, evidence of co-occurrence cannot discern where the moral content of the appraisal lies; that is, is the emotion the result of a cognitive moral judgement being appraised for its significance for our non-cognitive goals or desires? Or, is it the result of an appraisal of a non-cognitive moral desire in relation to non-moral cognitive beliefs or percepts? Or, perhaps it can be either of these options. Evidence of co-occurrence does not tell us anything about the *necessity* of either cognitive mental states, non-cognitive mental states, or appraisals for moral judgements. Of the meta-ethical positions described in Section 1.3, all of them except external cognitivism and conditional internal cognitivism say that non-cognitive mental states or states with non-cognitive qualities are necessarily involved in moral judgements. External cognitivists and conditional internal cognitivists claim that moral judgements can occur without non-cognitive mental states, and so can occur without emotion. However, despite this possibility, they also have the capacity to explain the widespread co-occurrence of emotions and moral judgements.

There is clearly a strong link between emotions and moral judgements. This link is consistent with even the most 'rational' of theories of moral judgement, cognitive externalism. In order for cognitivists to make progress from this point, they would need to defend either the possibility or the actuality of the amoralist, an individual who can make moral judgements without experiencing emotions. This is the same place that the discussion about moral motivation in the previous chapter ended up. Making progress on this question depends upon demonstrating that certain kinds of non-cognitive mental states can be *missing entirely* when someone makes a moral judgement.

Although there are serious limits on what the correlational data can tell us about what moral judgements are, the studies are significant for informing us about the regularity with which statements that we ordinarily take to be moral ones occur in conjunction with emotions. In other words, what evidence of co-occurrence can do is give us an indication of how regularly sentences that look like moral ones (such as 'x is wrong') or answers given to stereotypical moral questions (such

as, 'is it wrong to x?') co-occur with emotions. However, in terms of shedding light on deeper meta-ethical questions about the nature of our moral judgements, correlational data is not of any use. In the remainder of this chapter and in Chapter 5 I discuss whether other kinds of experiments fare any better.

4.2. Affect as providing information

There is a mode of psychological investigation which examines the causality of feelings in relation to moral judgements by manipulating people's affective states in order to see the resulting effects on moral judgements. It has been found that feelings can influence moral judgements and argued that this particular finding challenges cognitivism. The thought is, if moral judgements were beliefs, we should not expect that they would be influenced by feelings. Should we find that they are, we would have found evidence that moral judgements are not beliefs, but rather, somehow affectively driven; that is, we would have found evidence that people use their feelings as information in determining whether something is right or wrong. In the remainder of this chapter, I will describe this experimental research and explain why claims that such research disproves cognitivism are overstated.

The psychological research on the influence of feelings on moral judgements displays a bias in examining the effects of disgust on moral judgements to the relative exclusion of other emotions or feeling states.[3] Consequently, this chapter and the next focus heavily on the relationship between disgust and moral judgements. The focus on disgust originates from the findings of Wheatley and Haidt (2005) who demonstrated that people seemingly form moral judgements solely on the basis of their gut-feelings of disgust. Landy and Goodwin (2015) have defended the emphasis on disgust in the literature, claiming that it is the most unproblematic emotion to use in investigating the influence of emotion on moral judgement because it is the only moral emotion that has a pre-normative component and that can thus be induced in the absence of a moral transgression. That is, disgust can be induced using stimuli that are uncontroversially non-moral. This, they note, "allows a cleaner test of the "pure" role of emotion. If disgust was found to influence moral judgements, this would provide clearer evidence for such a pure role of emotion than would demonstrating that a normatively laden emotion,

[3]Although there is a heavy focus on disgust, some experiments test the effects of emotions and sensations other than disgust using stimuli which causes discomfort or pleasure to our other senses, such as loud and irritating sounds, uncomfortably hot rooms, bright lights and so on.

such as anger, influences moral judgement" (Landy and Goodwin, 2015, p. 519). Disgust has been experimentally induced in numerous experiments via a range of different elicitors with the aim of examining its effects on our judgements. I examine the direction and strength of the main claims within this literature. Along the way, I will also address many of the objections and worries that have arisen from these studies, such as the idea that our reports that certain kinds of actions or behaviours are disgusting is merely a metaphor; or the possibility that disgust is present in moral judgements only insofar as core elicitors of disgust are present in the moral scenario (core elicitors are those which are closely tied to the ancient evolutionary function of disgust as I will spell out shortly).

Disgust and moral disgust.

What is disgust? Given the overwhelming focus on disgust in the literature it is worth taking a look at what exactly disgust and moral disgust might be. Ekman and Friesen (1975) characterise disgust as an emotion which promotes avoidance of its eliciting stimuli:

> The prototypical disgust expression is characterized by a wrinkled nose, raised upper lip, and narrowed eyes (Ekman and Friesen, 1975, 1978). These actions are associated with a decrease in the exposed area of the eyes and closure of the nasal cavities, which may serve to protect the vulnerable mucous membranes of the face from exposure to infection (Susskind et al., 2008). The subjective experience of disgust is one of revulsion and offence, often coupled with a behavioural tendency to withdraw from the disgusting stimulus or to remove the stimulus from the self (e.g., by washing, spitting; Rozin, Haidt and McCauley, 1999). Chapman and Anderson, 2013, p. 301.

Things which elicit disgust make us feel nauseous. Uncontroversial examples include faeces, vomit, bodily fluids, decaying matter, and waste in general. Disgust is thought to have its origins in helping us to avoid disease or harmful pathogens. Some argue that the origins of disgust come from distaste specifically, whereby we experience a disgust response to bitter substances which are potentially poisonous and harmful to our health (see Rozin, Haidt and Fincher, 2009). Steiner (1973) has observed that infants just a few hours old express facial expressions characteristic of disgust in response to bitter substances. Other theorists investigating the origins of disgust allow for a more general disease avoidance function whereby disgust helps us to avoid

harmful pathogens more generally via our sense of smell (see Oaten, Stevenson and Case, 2009; Smith and Korsmeyer, 2004).

The status of disgust, at least this very basic form of disgust, as an emotion is questionable. Perhaps it is too psychologically basic to count as an emotion. It resembles our experience of sensations, such as heat induced pain, or the feeling of a person's hand brushing against your arm. This worry is connected to the discussion in Chapter 3 about the different ways one can conceptualise emotions. Specifically, the worry is that if disgust is an emotion, one will have to endorse some kind of feeling theory of emotions. Sensations do not depend on appraisals. And, it is doubtful whether one needs to have any kind of cognitive representations or non-cognitive desires, sentiments, or goals to experience disgust of a very basic kind, such as a disgust response to a type of food. Advocates of besires who wish to defend disgust's status as an emotion need to demonstrate how disgust differs from sensations such that it involves appraisals and sensations do not. For example, Prinz (2004) argues that disgust is a besire which is an appraisal of how states of affairs bear on things that matter to us (which he calls core relational themes). On his view disgust is an emotion because the feeling represents danger, whereas sensations do not. This conception of disgust is controversial and I will not go into it here. It seems, though, that if we take emotions to be appraisals of a Humean kind, then basic disgust may be too psychologically simple to qualify as an emotion.

What is moral disgust? A distinction is made between 'core' or 'physical' disgust (I use these interchangeably) and 'moral' disgust. Core disgust elicitors include those things which are tied more closely to the evolutionary origins of disgust, things which are serious potential contaminates, or disease threats. The origins and elicitors of core disgust are relatively uncontroversial. It is less clear if, how, and why *moral transgressions* might elicit disgust responses within us. Moral disgust is thought to have developed from physical disgust in the sense that we have come to experience literal feelings of disgust towards certain actions, behaviours, or characters. In other words, disgust is thought to have been harnessed to warn us of other more abstract social contaminants. Like core disgust, moral disgust occurs across cultures and there is evidence that children experience disgust at moral transgressions (Haidt et al., 1997; Danovitch and Bloom, 2009).

It is difficult to identify a pattern in the elicitors of moral disgust. Some theorists think that moral disgust can occur in response to any kind of moral transgression, whereas others argue that it has a more specific range of elicitors. I will discuss two important concerns that have been raised about moral disgust. The first concern is that disgust

experienced in moral situations is just a response to core disgust elicitors that are present in moral situations. It would not be surprising if people reacted with disgust at thoughts of blood and gore in moral scenarios. Moral disgust should go beyond this. The second concern is that, if we do experience moral disgust, (a) why aren't all things that are judged to be disgusting also judged to be morally wrong (that is, why are some actions disgusting but not morally disgusting)? and (b), why aren't all things judged morally wrong also disgusting? I'll look at two popular ways of dealing with these concerns. Explanation (1) says that moral disgust is elicited by a non-arbitrary, unified set of concerns. One particularly influential version says that moral disgust occurs to protect our purity or sanctity, protecting us from contamination of the soul, or from 'crimes against nature'. The details of this particular attempt at unification have been developed in a dubious manner. But, as I will discuss, some version of it could fall close to the truth. Explanation (2) attempts to distinguish disgust from moral disgust by showing that although people report feeling disgusted at moral transgressions, moral disgust is actually a blend of emotions or a metaphor. I argue that Explanation (1) is more fruitful. An important reason for preferring the first explanation is that it does not commit one (as the second explanation does) to the claim that other emotions *must* occur if disgust is to qualify as moral disgust.

Concern 1. Does the use of 'disgust' to describe a situation extend beyond things which quite obviously elicit core disgust? Moral transgressions often contain physically disgusting stimuli (Oaten, Stevenson and Case, 2009; Royzman and Sabini, 2001; Case, Oaten and Stevenson, 2012). Some authors have raised the concern that felt disgust, in the case of moral transgressions, may be triggered by the core disgust elements embedded within moral scenarios. It should not be surprising that people react to moral transgressions with disgust if there are physically disgusting stimuli present within the moral situation. For example, many moral transgressions, such as murder or rape, ordinarily contain elements of things we unambiguously find physically disgusting. We want to know, however, whether disgust is also elicited by socio-moral features of our environment in the absence of any physically disgusting stimuli. Case, Oaten and Stevenson (2012) correctly point out that to answer this question we need to find out whether people experience disgust towards moral transgressions which contain no core or physically disgusting stimuli, such as fraud or theft. These transgressions lacking elements of core disgust are referred to as *pure moral transgressions*.

Multiple studies now point to the conclusion that pure moral transgressions do elicit disgust. Chapman et al. (2009) found that being

treated unfairly activates the same facial muscles that become active in response to core disgust substances. Jones and Fitness (2008) found that induced physical disgust led people to make harsher judgements of the pure moral transgressions of theft and fraud. Nabi (2002) found that people spontaneously report on social events with no core disgust elements when they are asked to describe a time that they felt disgusted. People also endorse a disgust facial expression as depicting their experiential reaction to being treated unfairly in an economic game. A meta-analysis by Chapman and Anderson found 13 studies in which people reacted with moral disgust to cases that contained no core or physically disgusting elicitors, such as stealing a library book (2013, pp. 342, 345). Chapman and Anderson conclude that "...when people are asked whether they find pure moral transgressions disgusting, the answer seems to be "yes"" (2013, p. 309). Overall, the evidence that people find pure moral transgressions disgusting is highly suggestive.

Concern 2. Why aren't all disgusting things seen as morally wrong? And why aren't all things judged morally wrong also judged disgusting? We seem to effortlessly sort disgusting stimuli into physical and moral categories. It is an interesting question whether those categories can be clearly distinguished from one another. In particular, what can be said about the class of moral transgressions which we find disgusting? Are all moral transgressions capable of eliciting a disgust response from us, or does disgust have a more limited set of elicitors? If it is not (at least potentially) elicited by all moral wrongs, how are the moral wrongs which do elicit disgust unified?

Explanation (1): Moral disgust is elicited by a non-arbitrary unified set of concerns. Quite probably the most influential view, developed by Haidt and Graham (2007), says that we experience moral disgust in response to concerns about the contamination of the purity of our soul. This concern is taken to be driven by a tendency to feel disgusted by things which remind us of our animal origins:

> Those who seem ruled by carnal passions (lust, gluttony, greed, and anger) are seen as debased, impure and less than human, while those who live so that the soul is in charge of the body (chaste, spiritually minded, pious) are seen as elevated and sanctified. (Haidt and Graham, 2007, p. 106)

This position addresses concern (2) by claiming that core disgust is triggered by contamination threats to our bodies whereas moral disgust is triggered by contamination threats to the soul.

I will not try to evaluate the plausibility of Haidt and Graham's (2007) specific proposal here. However, I will attempt to give some plausibility to the broader idea that we experience moral disgust at things which threaten to contaminate our 'soul' understood as our personally endorsed set of desires (I will refer to these as our values). Recall that some emotion researchers think that disgust is too basic to qualify as an emotion. It is possible that core disgust is not an emotion as it does not involve an appraisal, whereas moral disgust does. Core disgust has quite a clear function of deterring and protecting us from harms, such as poisons, parasites, infections, and diseases. It is cognitively impenetrable, and it has a very robust and identifiable set of chemical elicitors to which our senses of taste and smell are sensitive. We can allow room for some conditioning and for simulation processes to influence this process.[4] Core disgust, on this view, is a sensation, similar to that of the pain experienced when your hand gets too close to fire. Moral disgust, on the other hand, is not triggered by chemical elicitors. It is an emotion as it is triggered by an appraisal. There are several ways the appraisal could unfold. If an appraisal theory of emotions is correct, an appraisal involves a cognitive and a non-cognitive component. It may be the case that moral disgust arises when we appraise that there is some threat to the fulfilment of our values (perhaps more specifically a threat by humans). The idea is that, just as feeling disgust in response to physical contaminants conferred a survival advantage by alerting us to potential harms, feeling disgust at things which threaten to contaminate our values may have conferred a survival advantage. In other words, core disgust may have been 'harnessed' to aid in the protection of things we care about and value – our soul, if you will.[5] Here we have some way to distinguish core disgust from

[4]Conditioning is a powerful mechanism. Interestingly, smell, taste, sight, touch, and sound can all elicit a disgust response. It is worth mentioning that a disgust response to core elicitors is likely to be a conditioned response when it arises via certain senses. It is possible that only one or two of our senses have a reflexive evolutionary trigger for disgust (those of smell and taste) and that our other senses come to elicit disgust through conditioning mechanisms. For example, the sight of physically disgusting stimuli such as vomit can elicit disgust only because the sight has been associated with the smell or the taste at some point. In this way, even basic disgust has abstract elicitors, for example, when one experiences physical disgust at viewing or hearing someone vomit in a television show. Merely imagining a threat of physical contamination can become enough to induce disgust in most people.

[5]It is possible to imagine that we might come to feel disgust at the possibility of the contamination of our values if it led to advantages in survival and reproduction. People tend to believe that it is possible to become physically contaminated by

moral disgust in a way that does not tie moral disgust to an abstract or religious notion of purity. If we are disgusted by lust, greed, gluttony, and anger it is because they, at least regularly, impair our ability to fulfil goals that we take to be of central importance. If this is correct, it should not be controversial that we think it is better that our 'soul' be in charge of our bodies, and that we would be concerned to protect the purity of our values from anything which can contaminate them.

We have an answer to question (a), the question of why not all disgusting things are judged morally wrong: only things which threaten our values are morally wrong, and identifying these threats requires us to make appraisals. We are still left with question (b), the question of if, and why, some moral transgressions cause disgust and others do not. Haidt and Graham (2007) take purity to be concerned with the protection and sanctity of our souls, by which they have in mind a

objects that have been in contact with, or in close proximity to others (Stavrova et al., 2016). In particular, people experience an aversion in the form of disgust to individuals or objects that have undesirable character traits. Many of the things people seem to be morally disgusted by are quite removed from any threats which could obviously result in harms to survival or reproduction. For example, people don't like the idea of trying on Hitler's jumper (Rozin, Haidt and McCauley, 2009). Values and behaviour do have an observable contagion effect. It is natural for human beings to hypothesise about the ways in which values and character traits might transfer from one individual to another. The belief that one can become physically contaminated with the character flaws of others through objects that the undesirable character has been in close proximity to is not such a crazy one for several reasons. To begin with, there are microscopic entities capable of altering human behaviour, specifically their tendency to behave recklessly, e.g. toxoplasmosis. Furthermore, it is likely to be evolutionarily adaptive to have some level of apprehension towards new stimuli or unusual behaviour (that which goes against social norms). At the very least it is adaptive to recognise novel stimuli in your environment and pay some attention to the possibility of harm that an unknown entity may pose. People vary in their openness to new experiences in part due to natural variations in this protective mechanism (conservative/liberal). It is worth mentioning that both, too much and too little openness to experience can be detrimental; too much openness and you risk harm, too little and you risk missing out on potential benefits, for example, the benefits of a new food source in a drought. The ideal position on this scale is not fixed. It will alter depending on environmental factors such as scarcity of resources. For example, in our current environment safe food sources are readily available. Consequently, a high disgust response to novel foods is adaptive; individuals with such a response err on the side of harm avoidance and are on the whole more likely to survive and reproduce than those who unnecessarily place themselves at risk. It is easy to see how the same high disgust response can be detrimental in a different environment. An important point, though, is that people will make errors. Just as some bitter tastes do not correctly track poison or any other harmful substance, not all things we perceive as morally disgusting will contain moral contamination threats.

narrow set of concerns relating to spirituality, chastity, and avoiding carnal desires. For them, purity violations cause disgust and other kinds of moral violations cause other emotions, for example, violations of autonomy are thought to cause anger. There is some reason to be sceptical of their proposal (Cameron, Lindquist and Gray, 2015; Landy and Goodwin, 2015). The idea that moral disgust protects the purity of our soul in the way they describe, is very vague and has strong connections to religion. Some authors claim to have empirical support for the claim that disgust is caused by transgressions characteristic of the 'purity' domain which is taken to be a subset of moral concerns (see Horberg et al., 2009; Seidel and Prinz, 2013). Whilst it is not clear on these accounts exactly which moral violations threaten the purity of our souls, it is clear that these authors do not think that *all* moral violations do so. The research on whether there is a correspondence between concerns about purity (as they seem to portray it) and disgust is, at present, very inconclusive. A meta-analysis consisting of 43 published and 30 unpublished studies by Landy and Goodwin (2015) found no evidence that moral transgressions within the 'purity' domain are more likely to elicit moral disgust than those in the non-purity domain (Landy and Goodwin, 2015, p. 530).[6] Further counter-evidence for the hypothesis that specific types of moral violations such as harm and purity are linked to specific emotions such as anger and disgust was found by Cameron, Lindquist and Gray (2015).

The claim that disgust is elicited by threats to our values is a variation of the story told by Haidt and Graham (2007). It says that moral disgust can be elicited by threats to the contamination of our values *in general*; that is, it can be elicited by *any* moral transgression. Defenders of the idea that moral disgust is caused by concerns about the purity of our values must reply that *all* moral transgressions *can* cause moral disgust. Rather than linking types of moral violations to particular emotions, an alternative plausible story is that moral transgressions are capable of eliciting several emotions, and which one is elicited depends on where the judge's attention to the moral situation is focused in the moral scenario.

Moral situations are regularly quite complex, and one's attention falls only on parts of the situation at a time, each part possibly triggering

[6]Purity domain moral transgressions in their analysis included crimes against nature and bodily and sexual purity type transgressions: "... the mean effect size for non-purity violations was, if anything, slightly larger than the mean effect size for purity violations ..., suggesting that the amplification effect is not restricted to moral transgressions involving bodily purity, sexual purity, or crimes against nature (Landy and Goodwin, 2015, p. 530).

differing emotions. For example, when evaluating moral scenarios, people consider their personal connection to the situation; their ability to act or intervene; their responsibility to act or intervene; sometimes their attention is focused primarily on the victim and sometimes primarily on the perpetrator, sometimes the situation being appraised occurred recently and sometimes it occurred a long time ago, for example, anger seems to dissipate over time, particularly as the victims and perpetrators of moral wrongs die over time. In its place one is likely to experience emotions such as sadness or disgust. Consider the emotions you might experience in response to slavery today as opposed to past slavery. It seems that anger is a more likely response to currently occurring slavery and that sadness is a more likely response to historical slavery, assuming all potential ramifications have been remedied. The ability of moral transgressions to elicit multiple emotions may be the inspiration for the second attempt at answering the question about what distinguishes morally disgusting things from disgusting things in general.

Explanation 2: Moral disgust is used as a metaphor for a blend of emotions or some other more complex emotional state. People report feeling disgusted at pure moral transgressions, but some critics worry that individual's reports may not accurately describe their true emotional state. The disgust-as-metaphor hypothesis says that the disgust reported at moral transgressions is actually a complex blend, or a cluster of emotions rather than a single emotion of disgust (Case, Oaten and Stevenson, 2012). For example, perhaps disgust is the dominant identifiable feeling in an as yet unnamed blend of emotions. Or, it may be possible that people report feeling disgusted as a way to indicate to others that they disapprove of some action despite their actual emotional reaction to the moral transgression being of a more complex kind (Case, Oaten and Stevenson, 2012, p. 215).

The extent to which more than one emotion occurs in response to moral transgressions is extremely hard to discern. Chapman and Anderson (2013) note that current research is suggestive of the idea that moral transgressions are capable of eliciting multiple emotions, but that the incidence of this occurring is obscured in the experimental literature because participants are generally presented with a 'forced choice' response option, whereby they are forced to pick just one emotion to describe their moral opinion. In support of this idea, Simpson et al. (2006) found that individuals self-reported similar ratings of both disgust and anger when presented with photographs depicting moral transgressions (the photographs did not contain core disgust elicitors). A forced choice study by Rozin et al. (1999) which found that pure moral transgressions elicited anger was repeated by Hutcherson and Gross

(2011) without the forced choice requirement. They found that both disgust and anger were elicited by pure moral transgressions (Chapman and Anderson, 2013, p. 307). Simpson et al. (2006) also found that pure moral transgressions elicited similar levels of both anger and disgust (Simpson et al., 2006). There is also some reason to be sceptical about verbal self-reports about emotions, i.e. people report feeling anger, yet exhibit facial expressions of disgust. In moral scenarios individuals often report that other emotions are present in addition to disgust. In particular, there is a high incidence of reported anger in response to moral violations (see Simpson et al., 2006; Marzillier and Davey, 2004; Gutierrez, Giner-Sorolla and Vasiljevic, 2012; Hutcherson and Gross, 2011). In general, it seems that if people are not forced to make a decision between anger and disgust most people report experiencing both (Cameron, Lindquist and Gray, 2015).

According to the disgust-as-metaphor view, moral disgust may be differentiated from core disgust by its being accompanied by, or its being a blend of, other emotions such as anger or sadness, whilst pure disgust scenarios elicit disgust in isolation. It seems that moral scenarios do regularly elicit multiple emotions. If moral transgressions usually generate multiple emotions it will be important to identify the relationship(s) between the various emotion states. There is some evidence that the various emotion components elicited in response to moral transgressions are dissociable (Gutierrez, Giner-Sorolla and Vasiljevic, 2012; Hutcherson and Gross, 2011; Simpson et al., 2006). For the disgust-as-metaphor hypothesis, disgust which occurs in isolation of other emotions is insufficient for a moral judgement. Moral disgust is so because it is a blend of, or accompanied by, other emotions. In contrast, the attempt to distinguish disgust from moral disgust by appealing to some unified set of moral disgust elicitors (explanation 1) does not have this problem. Disgust resulting from an appraisal with moral content is sufficient. One need not experience anger, sadness or any other emotion (although they might do so). To illustrate, consider an individual who is disgusted by paedophiles. For this individual, the thought that someone would abuse children is morally disgusting, but it has never made them angry or sad. It is possible that the kind of appraisal which is capable of producing anger in response to paedophilia has never occurred for this individual. Perhaps, for example, they have never been close to (or responsible for) someone who has been affected. Maybe they have never witnessed the ramifications for victims which are capable of generating an emotion of sadness (a certain kind of appraisal). In this case, the individuals moral response to paedophilia may be limited, but it seems

to me that their disgust response ought to still be considered a moral one despite the absence of other emotions.

4.3. Do feelings influence moral judgements?

The remainder of this chapter examines empirical work which aims to establish causality of the relationship between feelings and moral judgements. Pizarro, Inbar and Helion (2011) outline two possible ways in which feeling (specifically disgust) could causally influence moral judgements. First is what they label the Amplification Hypothesis (AH). According to the AH a feeling (or emotion) can amplify or make moral judgements stronger than they otherwise would have been. In other words, affect can be an input to moral judgements on this view. Secondly, there is a stronger depiction of the causal relationship which they call the Moralisation Hypothesis (MH). It states that feeling (or emotion) can play a causal role in the formation of moral judgements and can be the sole basis upon which moral judgements are formulated. Both the amplification hypothesis and the moralisation hypothesis experimentally manipulate disgust to examine its effect on moral judgements. The two hypotheses are closely linked, however there are not many experiments which explore the moralisation hypothesis. To do so, one needs to show that incidentally induced feelings or emotions can cause people to judge *morally neutral* scenarios as morally right or wrong. Most studies test the effect that incidentally induced disgust has on people's answers to morally laden questions. The truth of the hypothesis that moral judgements can be solely based on feeling is obscured in morally laden experiments by the possibility that cognitive moral appraisals are merely being amplified by affect. In these experiments, though it may be the case that non-cognitive feelings were the basis of the moral judgements, the strongest claim to causality one can make is that affect amplifies moral judgements.

Thalia Wheatley and Jonathan Haidt were among the first to demonstrate the ability of induced disgust to influence people's judgements about the permissibility of certain actions. They hypnotised participants to feel disgust at reading a neutral word; either 'often' or 'take'.[7] They then presented participants with a series of vignettes and asked them to rate both "how morally wrong" and "how disgusting" the behaviour was

[7]Of the 64 participants participants in their study, 45 reached the desired levels of hypnotism exhibiting amnesia to the instructions to feel disgust in response to hearing the selected words until they were later cued to remember. Wheatley and Haidt limited their analysis to these individuals to maximise the degree of certainty that disgust informs judgement.

(see Wheatley and Haidt, 2005). The vignettes were about: "... second cousins who had a sexual relationship, a man who ate his already dead dog, a congressman who took bribes, an ambulance-chasing lawyer, a shoplifter, and a student who stole library books" (Wheatley and Haidt, 2005, pp. 780–781). They found that the participants rated the vignettes both more disgusting and more morally wrong when the hypnotically induced disgust word was present in comparison to when it was absent (Wheatley and Haidt, 2005, p. 783). These findings provide evidence for the amplification hypothesis.

In addition to this, Wheatley and Haidt presented participants with a morally neutral vignette and found that for one third of the participants evaluations of disgust and moral wrongness were forthcoming when the hypnotically induced disgust word was present (Wheatley and Haidt, 2005). This finding is significant as it suggests that the feeling of disgust was the sole source of information that some participants used to evaluate actions and that disgust can cause people to make negative moral judgements in situations which are ordinarily evaluated as morally neutral. This finding provides evidence for the moralisation hypothesis. The participant's describe their own puzzlement at their judgements as follows: "When 'often' appeared I felt confused in my head, yet there was turmoil in my stomach. It was as if something was telling me that there was a problem with the story yet I didn't know why" (2005, p. 783). One nonamnesiac participant commented: "I know about 'the word' but it still disgusted me anyway and affected my ratings. I would wonder why and then make up a reason to be disgusted" (2005, p. 783). Haidt and Wheatley surmise that we use gut feelings as information when we make our moral judgements, concluding that moral judgements are grounded in "affectively laden moral intuitions" (2005, p. 780). They take their findings to "illustrate the philosopher Hume's (1739/1969) famous statement that 'reason is... the slave of the passions, and can pretend to no other office than to serve and obey them' " (Wheatley and Haidt, 2005, p. 783).

Wheatley and Haidt's investigation inspired a wealth of other studies which sought to examine the effects of disgust on judgements, many of which have confirmed their finding that inducing incidental disgust influences peoples moral judgements. For example, Eskine, Kacinik and Prinz (2011) tested the effects of bitter and sweet tasting beverages on moral judgements. They report that moral judgements in the bitter taste condition are significantly harsher than judgements in the sweet and neutral conditions. Moretti and Pellegrino (2010) found that participants in a disgust condition, as opposed to a sad or neutral condition, rejected more unfair offers in ultimatum games. Induced

disgust also influenced how unfair an offer was perceived to be with those in the disgust condition perceiving unfair offers to be more unfair than those in the other conditions. And, Horberg et al. (2009) found that participants who watched a disgusting film clip rated morally neutral vignettes as morally wrong.

Meta-studies. The experiments mentioned above have certainly demonstrated promise for the claim that incidental disgust, among other feeling states, can influence our moral judgements. However, meta-analyses of this literature have turned up deeply conflicting findings which cast doubt on any kind of strong conclusions about the causal relationship between moral judgements and disgust. Chapman and Anderson (2013, pp. 303, 313) collated the data from studies which induced incidental disgust to test its effect on moral judgements. They found that almost all of these studies have demonstrated that incidentally induced disgust influences moral judgements. Contrary to these findings, a meta-analysis by Landy and Goodwin (2015) has generated some more ambiguous conclusions regarding the amplification and moralisation hypotheses which seriously undermine any kind of strong conclusions that the evidence points to thus far. Landy and Goodwin's (2015) meta-analysis included 33 published studies in which disgust was induced incidentally to the moral judgement and which involved the assessment of moral transgressions. They, like Chapman and Anderson (2013), found a small to medium amplification effect whereby disgust was found to enhance the severity of moral judgements. However, they also discovered troubling evidence of a publication bias whereby studies finding no evidence of amplification effects were not well represented in the literature. Moreover, when these studies were accounted for statistically, they found no significant effect of induced disgust on moral judgements (2015, p. 528). Consequently, they urge caution against strong claims of amplification and note that more research is needed to elucidate the debate. They found a similar result in regards to the stronger moralisation hypothesis. That is, there is a small moralisation effect apparent in the published literature which purportedly shows that induced disgust is sufficient to produce moral condemnation. But, once again, when the unpublished studies were accounted for no such effect was found (2015, pp. 530–2). As with the amplification hypothesis, they urge caution in dealing with the moralisation hypothesis, even more so for the fact that the moralisation effect is based on a very small sample of studies. Landy and Goodwin remind their readers that although their meta-study casts doubt on strong claims about the causal effect of disgust on moral judgements, it could still be the case that judgements may be influenced by emotions other than disgust (2015, p. 531).

Explaining the conflicting findings. Several attempts have been made to explain why some experiments find evidence of amplification and moralisation and others do not. One promising idea says that the experiments which find no results have used inadequate disgust induction techniques. This concern asks about the effect of both too much and too little disgust induction. Some disgust induction techniques are more effective than others. For example, asking subjects to imagine a time they were disgusted produced very little disgust in comparison to exposing them to unpleasant smells. The worry that disgust was *under-induced* in experiments is minimal; testing the success of the independent variable (disgust stimuli) is standard practice, unlikely to be overlooked by experimental psychologists. A more interesting suggestion is that disgust was *over-induced* in some experiments and that its salience led to negative findings. The thought is that the participant's *awareness of the source* of the disgust is a crucial factor in determining whether the emotion is used as information by the participant's for making moral judgements. Specifically, it is suggested that when people are unaware as to the source of their disgust they will attribute it to whatever most easily comes to mind. When experiments induce disgust in an obvious way it is thought that the participant's will be more likely to discount the feeling as providing information when making moral judgements. This idea is given credence by a study by Schwartz and Clore (1983) who famously found that the weather significantly influenced people's ratings of their life satisfaction in a telephone interview. However, when the experimenters drew people's attention to the weather before asking them about their life satisfaction, the participant's ratings were no longer affected by it (cited in Case, Oaten and Stevenson, 2012, p. 201). Another suggested reason for the inconsistent findings states that the effect of induced disgust on moral judgements is strong for some individuals and not for others, in particular for those individuals who are highly sensitive to disgust and/or aware of their bodily sensations. However, even if individuals vary in their tendency to use their feelings as information, we should not expect different findings between experiments if random samples of significant sizes were used, which again, is standard practice unlikely to have been overlooked.

4.4. Explaining amplification and moralisation

Research currently indicates that affect is at least sometimes used as information in evaluating moral situations. An important point to stress here is that the studies which demonstrate that an incidental

feeling is capable of influencing judgements do not have to claim that the specific judgements that are made in the experiments are moral judgements. Judgements based on incidental disgust are most likely going to be a mistake on any theory. Ultimately, the studies more abstractly show that our bodily sensations or feelings are (at times) a source of information that we draw on. The studies which find that affect influences judgements give plausibility to meta-ethical theories which emphasise non-cognitive (affective) mental states as being central to moral judgements. Four of the five meta-ethical positions outlined in Section 1.3 claim that some kinds of bodily feelings are inputs to moral judgements (non-cognitivism and the three hybrid positions: hybrid-state theory, hybrid-expressivism, and moral thought pluralism). On these views, moral judgements depend (at least at times) on world-to-mind mental states. None of these theories, however, claim that *sensations* are a sufficient basis for a moral judgement; they posit some kind of world-to-mind mental state which a sensation is not.

Amplification studies cannot be used to comment on which of the meta-ethical theories emphasising non-cognitive mental states is most plausible. In amplification experiments it is still possible that incidentally induced disgust is amplifying cognitive moral judgements. The moralisation studies, on the other hand, claim to demonstrate that affect can be sufficient for producing what appear to be moral judgements. In these experiments we can be certain of the origin of the feelings that are influencing judgements. That is, we can guarantee that a judgement was caused by incidental feelings. All moral theories are likely to claim that these judgements involve some kind of mistake on behalf of the judge. However, there may be some reason to still think of these judgements as moral judgements. I discuss what to say about these cases in Chapters 5 and 6.

Does evidence of amplification and moralisation provide evidence against cognitivism? According to cognitive theories of moral judgement, moral emotions are a consequence of an appraisal which contains a cognitive moral belief. They have a greater challenge to explain why experiments sometimes find that we use feelings as information to form what appear to be moral judgements. On this view emotions are outputs of moral judgements. One may be tempted to think that if moral judgements are cognitive then there should be *zero* influence of incidental feelings on moral judgements. In this section I will explain why this is not so. Cognitive theories can explain moral amplification (AH) produced by incidentally induced sensations. They can also explain why people are inclined to moralise from incidental affect (MH), although they may have to claim that such cases are not moral judgements.

Let's assume that ordinarily moral feelings or emotions are caused by appraisals which contain a cognitive moral judgement. For example, we may have a moral belief and a desire which results in an emotion. One plausible way in which amplification may occur on this view would be if the feeling (or emotion) which results from the appraisal causes us to direct our attention towards the underlying beliefs or desires which caused the initial emotion. It is plausible that such a process can result in the amplification of the initial emotion, perhaps by leading to an altered level of stress on, or commitment to, the belief or desire involved. This feedback process is perfectly consistent with cognitive theories of moral judgement. It can explain amplification. On this view, it is the redirection of attention to the underlying beliefs and desires, one or both of which intensify, which leads to an amplified emotion. This can occur when the appraisal initially occurs unconsciously. For example, suppose that you have unconscious beliefs and desires which produce within you an emotion. Phenomenologically, all that is available to you in this process is the emotion. In this sense, it is the feeling that 'comes first' for you in that it is consciously available to you before the beliefs and desires which produced it. It is plausible that such feelings focus our attention or 'set us on the search' for possible beliefs and desires. This can look like a post-hoc process. Whether it is or not will depend on whether an individual is correctly *identifying* their beliefs and desires which underlie their emotion or *creating* them. If a process such as this is occurring, we should expect that individuals' judgements will be sensitive to the source of incidentally induced affect. This point is consistent with the idea that over-induced disgust induction techniques may be the cause of the conflicting findings that incidental disgust influences judgements.

In the induced emotion studies, moral judgements are influenced by incidentally induced affect. That is, we can guarantee that the feelings which people are 'using as information' in the experiments are not solely the result of an appraisal with moral content in the amplification experiments and not at all the result of an appraisal with moral content in the moralisation experiments. Judgements are either enhanced by or caused by things such as bad smells and gooey substances etc. In moralisation experiments, there is unambiguously a feeling which comes first (and not just phenomenologically) in the process of the subject's 'moral' decisions. Individuals use incidental feeling as the sole basis for their judgement despite not being able to identify any underlying beliefs or desires. Whether or not there are any good reasons to classify these judgements as moral judgements is taken up in Chapter 5 and Chapter 6.

4.5. Concluding remarks

To summarise, the empirical literature finds that emotions and moral judgements often co-occur.The phenomenon of co-occurrence cannot tell us anything about which, if any, meta-ethical theory of moral judgement is likely to be correct. The findings are consistent with cognitive, non-cognitive, and hybrid theories of moral judgement. Experimental studies which manipulate disgust to test its effect on judgements suggest that affect is sometimes used as information when people make moral decisions. This is particularly likely when they are unaware of the source of the affect. Most of these studies are only capable of supporting the amplification hypothesis which is consistent with cognitivism. There is some evidence for the moralisation hypothesis provided by empirical studies which demonstrate that incidental affect is sometimes the sole basis upon which people are willing to base moral decisions. Deciding what to say about these cases is a problem for all theories of moral judgement. Chapters 5 and 6 make some progress in deciding what to say about the moral status of cases where we can guarantee that a judgement was caused by incidental feelings.

CHAPTER 5

Missing cognitive and non-cognitive moral content

The central aim of the empirical literature is to investigate the content of the mental states which underlie stereotypical cases of moral judgements. Theoretically, this investigation can be partitioned into three interesting empirical pursuits. Firstly, one may look for evidence that moral judgements can occur in the absence of any cognitive moral beliefs. This was the primary aim of the studies investigated in Section 4.3. Such evidence would cast doubt on cognitive theories and hybrid-expressivist theories of moral judgement and would support hybrid-state (besire) and non-cognitive theories. Secondly, we can look for empirical evidence that moral judgements can occur in the absence of any non-cognitive mental states. This possibility is explored by those who discuss the possibility and the actuality of the amoralist. Evidence of amoralists challenges non-cognitive, hybrid-state, and hybrid-expressivist theories of moral judgement and supports cognitive and moral thought pluralist accounts. This empirical pursuit was discussed in Section 2.5 and is not revisited here. Thirdly, empirical literature may be used to show that both cognitive and non-cognitive content of a certain *moral* kind is missing in judgements that we stereotypically take to be moral ones. As I will explain, such findings support hybrid-state and simple non-cognitive theories of moral judgement and challenge all others. This chapter examines evidence for this third option.

Evidence that some judgements that we would generally take to be moral ones are missing content of a certain 'moral' kind comes from the discovery of a phenomenon known as moral dumb-founding. Moral dumb-founding is defined as "the stubborn and puzzled maintenance of a moral judgement without supporting reasons" (Haidt, Bjorklund and Murphy, 2000, p. 6). The phenomenon of moral dumb-founding most straight-forwardly supports a simple kind of non-cognitive theory and hybrid-state theories of moral judgement. So far there has not been much reason to take the notion of a hybrid-state (or besire) seriously, so the empirical studies provide an interesting new contribution to the discussion.

Before moving on to look for missing moral content I would like to highlight a recurring concern relating to the tasks of stipulation, description, and explication. In experimental studies, theorists tend to proceed by creating scenarios that we intuitively take to be moral ones and asking people their views about those situations. The experimenters take peoples' answers to the experimental questions to reveal their moral judgements. They then proceed to examine what is going on in the judge's mind in terms of their beliefs and desires etc. It is always possible of course, to stipulate a definition of moral judgement. Once a definition of moral judgement has been stipulated, empirical methods can be used to show us how common or rare those judgements are. However, there should be a good reason for stipulating a definition. Practically we want a theory which captures what we intuitively take to be moral judgements to some degree. That is, we would not want a theory which essentially shows that most of the judgements we take to be moral ones do not stand up to the demands of the stipulated definition. The discovery of judgements that we wish to call moral judgements which do not meet the definition will likely influence our attraction towards various theories. However, we must also be open to the idea that our intuitions about moral judgements are wrong. The issue of justifying a stipulated definition will be the focus of Chapter 6. For now, I will proceed with the psychological method of examining judgements that we ordinarily take to be moral. To begin with I will outline some variations in the claims that are regularly made about the role of reasons and reasoning that are involved in moral judgements.

5.1. Varieties of reason

The various meta-ethical theories of moral judgement make different claims about the kinds of reasons and reasoning that are involved in moral judgements. The claim that moral judgements involve reasoning is very vague. It will be useful to distinguish between four possible interpretations of this claim. The first is that moral *appraisals* are responsive to reasons. The second and third track two different ways in which one may claim that *cognitive moral beliefs* are necessary for moral judgements. The fourth does not make any claim that beliefs of some kind are necessary for moral judgements. Rather, it specifies that moral judgements depend upon reasoning processes occurring within the non-cognitive domain which involve conscious effortful deliberation.

Moral appraisals are responsive to reasons. The first stance on reasoning, which says that moral *appraisals* are responsive to reasons is explicable by all meta-ethical theories of moral judgement. Moral

judgements are regularly a part of an appraisal. The emotions generated by these appraisals are responsive to new factual information, as well as to the recognition that one has other competing desires which are threatened by the one currently producing the appraisal. The non-cognitivist and hybrid-state theorist are only committed to claiming that values or besires are not amenable to reasoning. That is, we do not arrive at, or alter, our values or besires through reasoning. Reasons (or factual information about states of affairs) will be relevant to non-cognitive moral judgements if they exist, and for the hybrid-state theorist, reasons can trigger new besires.

Many of our desires for what ought to be done are instrumental. For example, I have the desire that gun laws be very restrictive *because* I have the desire to live in a society that is safe and conducive to the well-being of its citizens. Factual information is of significance to the instrumental desires that I hold. I see a connection between gun laws and safety and we can reason and argue about this connection. We can also argue about the desire for safety and well-being in relation to other non-cognitive goals. That is, we can argue about whether our desires are instrumental or fundamental. But the ones that I take to be most important upon consideration of conflicts in fulfilment are not amenable to reason. In other words, it is not a problem for hybrid-state theorists or non-cognitivists that reasons, or facts are important to a person's moral decision making, even though it is the case for both, that fundamental world-to-mind mental states or mental state components are unresponsive to reasons.

Two roles for cognitive moral beliefs. Hybrid-expressivists and cognitivists make stronger claims about the kinds of reasons that need to take place for moral judgements to occur. For them, the beliefs involved must contain moral content by explicitly representing moral concepts such as 'good', 'bad', 'right', or 'wrong'. There are two ways that one could claim that cognitive moral beliefs are *necessary* for moral judgements.

Theories of moral judgement which claim that moral judgements require the cognitive endorsement of non-cognitive mental states are hybrid-expressivist. Hybrid-expressivism allows a place for non-cognitive mental states in moral judgements. However, it states that non-cognitive mental states are not sufficient for moral judgements. On this view non-cognitive mental states might be endorsed when one happens to hold both, a moral belief and a moral non-cognitive attitude towards some action, situation or character, or when someone has a cognitive moral belief about the merit of their non-cognitive attitudes.

Cognitivists, on the other hand, claim that moral judgements are fundamentally beliefs. Cognitivists, like hybrid-expressivists, also state that moral judgements necessarily express moral beliefs. They deny that non-cognitive mental states are also necessary to the process of making a moral judgement. Cognitive beliefs are sufficient. This is the strongest possible cognitivist view of moral judgements. It denies that there are moral judgements, or any components of moral judgements which are, in principle, unresponsive to reasons. A belief is a mind-to-world representation. Moral beliefs explicitly reference moral concepts such as 'good', 'bad', 'right' or 'wrong'. The beliefs aim to represent how things are, and they do so without presupposing any non-cognitive mental states, such as desires or goals, as providing the foundation for which beliefs can be true or false.

For both cognitivists and hybrid-expressivists, if we agree that a judgement without cognitive moral content has been made in relation to the permissibility of some situation or action, those judgements would not qualify as 'moral' judgements, as they fail to meet the stipulated requirements requiring a mandatory role for distinctly *moral* reasons. Hybrid-state theorists and non-cognitivists can deny these stronger commitments to the roles of reasoning in moral judgements which require moral beliefs. They need only respond by arguing that non-cognitive mental states or besires can be sufficient for a moral judgement. This point will be further discussed in Chapter 6. Moral dumb-founding experiments question the need for explicit moral beliefs with moral content in moral judgements by claiming to have provided examples of moral judgements which do not contain any explicit moral beliefs.

Non-cognitive reasoning. Non-cognitivists claim that moral judgements express a world-to-mind mental state. Most proponents of this view do not think that just any non-cognitive mental state should be considered a moral judgement. Blackburn (2002), for example, thinks that moral judgements are those non-cognitive mental states that we demand that others share. Intuitively, the idea is that we have many preferences for how we want the world to be, some of which we do not mind if others share, some of which we prefer others to share; and still others which we demand that others share. Only the latter encompass our moral values. This kind of preferencing of non-cognitive mental states is a kind of endorsement. There are other ways a non-cognitivist may differentiate moral non-cognitive mental states from other kinds of non-cognitive mental states; but generally, it will involve our committing ourselves to some goals over others in some way. Importantly, endorsement of a non-cognitive mental state can be a non-cognitive or cognitive act. Here I am referring only to non-cognitive endorsement,

where we endorse, or have preferences for some non-cognitive states, not for any reasons, but simply for the fact that a certain non-cognitive mental state is more appealing to us. This non-cognitive endorsement view agrees that reasoning cannot give us our ends. If one were to require the cognitive endorsement of non-cognitive moral mental states, the position would be rendered hybrid-expressivist.

World-to-mind mental states can come in varying strengths. Moral judgements, if they have this direction-of-fit, seem to require a high level of endorsement. That is, part of what makes them one's moral judgements, as opposed to another world-to-mind mental state is that they are generally of high importance to their possessor and they are enduring rather than fleeting for the individual who has them. Such endorsed mental states are sometimes referred to as sentiments or values. Our representations of how states of affairs are, were, or will be are relevant to world-to-mind mental states. The question about the sufficiency of some kinds of non-cognitive mental states to qualify as moral judgements will be addressed in Chapter 6.

The kinds of reasons and reasoning mentioned above detail the commitments made by cognitivists, sophisticated non-cognitivists, and hybrid-expressivists about the kinds of mental states that moral judgements are. Hybrid mental states are single mental states with both directions-of-fit. They are not endorsed cognitively or non-cognitively in the ways described above. Likewise, some simple Humean appraisals are not endorsed in the ways mentioned above. A simple appraisal may contain a basic non-endorsed non-cognitive mental state and a basic cognitive representation which doesn't contain explicit moral content. This appraisal may produce emotions which may or may not be taken to be relevant to morality. Moral thought pluralists can claim that endorsement of either the endorsed non-cognitive or moral cognitive variety is sufficient for a moral judgement. It is as vulnerable to the empirical literature as cognitive and non-cognitive theories of moral judgement are individually. The topic of endorsement is taken up in Chapter 6.

5.2. Missing cognitive and non-cognitive moral content

Moral dumb-founding. Many experiments now demonstrate that people make what appear to be moral judgements despite not being able to provide any reasons for those judgements. In such cases, individuals are said to be morally dumb-founded. The general framework of such studies involves asking subjects their moral views on vignettes which have been carefully constructed so that none of the standard reasons

we use to justify our moral judgements are applicable. Through these experiments, doubt has been raised about the commonly held view that moral judgements are driven by reasons. The studies suggest that our moral judgements are instead caused by some kind of feeling or emotion, such as a feeling of approval, disapproval, disgust, or intuition. In other words, moral dumb-founding cases provide evidence for theories which say that we use feelings as information to discern right from wrong. This chapter assumes that emotions are appraisals which are either comprised of cognitive and non-cognitive mental states (Humean), or that they are a single mental state with both components (anti-Humean).

One particularly famous moral dumb-founding vignette is as follows:

> Julie and Mark are brother and sister. They are travelling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They kept that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it Okay for them to make love? (Haidt, 2001, p. 814)

Other scenarios devised to demonstrate the absence of reasons in moral judgements include those by Haidt, Koller and Dias (1993), who asked experimental participants their views on, among other things, "... eating one's dead pet dog, cleaning one's toilet with the national flag, or eating a chicken carcass one has just used for masturbation" (Haidt, 2001, p. 817). Haidt et al. found that these kinds of questions left individuals morally dumb-founded, claiming that subjects would "... stutter, laugh, and express surprise at their inability to find supporting reasons, yet they would not change their initial judgements of condemnation" (Haidt, 2001, p. 817). Morally dumb-founded individuals would say things like: ""I don't know", I can't explain it, I just know it's wrong"" (Haidt, 2001, p. 814). Haidt, Bjorklund and Murphy (2000) took care to emphasise that the individuals took themselves to be making *moral* evaluations despite not being able to justify their condemnation (also see Haidt, 2012). Moral dumb-founding experiments are taken to reveal that the general method by which moral judgements are made is via intuition, feeling or emotion.

Cognitive theories of moral judgement claim that moral judgements are beliefs, for example, a cognitive moral judgement may be a belief that goodness is identical to some natural property, or that good actions are those which we could not will to become a universal law. Cognitivism is undermined by cases of moral dumb-founding. To frame it in terms used thus far, moral dumb-founding studies reveal that the cognitive component of an appraisal (emotion) does not always contain complex moral beliefs. In fact, the cognitive (mind-to-world) component contributing to the feeling of rightness or wrongness in moral-founding cases seems to be no more than a basic percept at times. Moral dumb-founding vignettes also threaten the most plausible of non-cognitive theories of moral judgement. It is plausible that moral non-cognitive mental states need to be of a certain kind, for example, moral non-cognitive mental states may be distinguished from non-moral ones by their being ones that we demand others share given our recognition of conflicts. If this is the case, then moral non-cognitive judgements seem to require some degree of conscious processing. Morally dumb-founded individuals cannot point to any of their moral values to which facts of the situations described in the vignettes are relevant. Given that moral non-cognitive mental states require conscious processing, it is difficult to explain why subjects cannot access the values driving their emotion.

Moral dumb-founding studies most straightforwardly support hybrid-state theories of moral judgement, as well as any position which allows that appraisals consisting of very basic non-cognitive mental states (such as desires for survival and reproduction) and very basic cognitive mental states (such as visual representations of danger) can be *moral* judgements. For both besire advocates and simple non-cognitivists, it is likely that the appraisals involved have occurred unconsciously presenting the individuals with only a feeling consciously. Hybrid-state theorists and simple non-cognitivists do not face the same problems as cognitivists and complex non-cognitivists, as the appraisal components at no point require cognitive processing which is accessible to consciousness. In other words, the mental states involved in appraisals can be more or less cognitively complex. A simple appraisal may involve, for example, a drive for survival and a representation of a snake or a cliff. A complex appraisal may involve a desire for autonomy (that one demands others share) and a representation that some policy threatens it. More basic appraisals are more capable of being processed unconsciously, or in a single step process. Consequently, simple appraisals are more prone to leaving someone who is experiencing the outcome of an appraisal dumb-founded.

Post-hoc reasoning and biases. Haidt argues that our sense that we use moral reasoning to arrive at our moral judgements is an illusion and that moral reasoning and justification ordinarily occur after intuitive moral judgements and serve to persuade others to share our views (Haidt, 2001, pp. 45-46, 814). In support of this claim, he has gone to great lengths to demonstrate not only that people make moral judgements in the absence of reasons, but that the reasons people do give for their moral judgements can be demonstrated to be generated post-hoc. Post-hoc reasoning (also known as confabulation) refers to the phenomenon of making up reasons for our own actions or judgements after the fact, which we falsely believe to be the real reasons for our judgements or behaviour. In other words, moral reasons are unknowingly constructed after an intuitive moral judgement has been formed.

The phenomenon of post-hoc reasoning has been robustly demonstrated in numerous experiments. In one experiment conducted by Wilson and Nisbett (1978), participants were invited to choose (and keep) one of four pairs of stockings laid out on a table. Importantly, the stockings were all identical. Wilson and Nisbett found that the majority of people identified stockings from the right hand side of the table as being of higher quality than those on the left hand side (Wilson, 2002, p. 103). The participants were asked to explain why they choose the pair that they did. Each participant made up a reason for their decision which was based on some perceived difference in the quality of the stockings. Wilson and Nisbett concluded that the participants were blind to the real reason behind their judgements, suggesting that the participants' decision was likely to have been due to a 'position effect', whereby right-handed subjects chose the pair closest to their right hand (Wilson and Nisbett, 1978). The experimenters then asked the participants whether they thought that the position of the stockings had an effect on their decision. All but one subject (who was studying psychology) replied negatively (Wilson, 2002, pp. 103–4). In another study, Wheatley and Haidt (2005) hypnotised subjects to feel disgust at hearing a neutral word. Subjects were presented with the following vignette which was morally neutral in nature: "Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He [tries to take/often picks] topics that appeal to both professors and students in order to stimulate discussion" (Wheatley and Haidt, 2005, p. 782). In this study, subjects who were hypnotised to feel disgust judged the actor in the morally neutral vignette to have acted wrongly. Some individuals desperately searched for reasons to justify their judgements: "[o]ne participant wrote: "it just seems like he's up to something." Another confided that the story

evoked bad high school memories, making him view Dan as a "popularity seeking snob" (Wheatley and Haidt, 2005, p. 783). Wheatley and Haidt again found that people persisted with their moral judgements after their post-hoc reasoning attempts failed: "It just seems so weird and disgusting" and "I don't know [why it's wrong], it just is." (2005, p. 783).

Further evidence that we are not the kind of moral reasoners that we think we are comes from findings that we are extremely prone to biases when it comes to 'searching for truth'. We have a 'my side bias' which manifests in the search for evidence that supports our views, rather than evidence that disconfirms them (Perkins, Farady and Bushey, 1991; Kuhn, 1991; Kunda, 1990). We are biased in the level of scrutiny that we subject different ideas to, and we tend to evaluate claims in relation to whether they cohere with our other beliefs as opposed to investigating their truth in a more thorough manner (Perkins, Allen and Hafner, 1983). In addition to these concerns, Haidt describes other ways in which our judgements are biased by strong social motivations that we seem to have (2001, p. 819–821). For example, desires for social acceptance and harmony have repeatedly been shown to influence our judgements, the extent to which was famously demonstrated in the Asch conformity studies (Asch, 1951).

Anti-rationalism. Haidt takes his findings to pose a serious threat to rationalist theories of moral judgement. For Haidt, rationalism is the view that moral judgements are made by processes of reasoning and reflection, involving some conscious steps that the individual is aware of. He adheres to Kathleen Galotti's (1989) view of reasoning which excludes 'one step mental processes' such as gut reactions, flashes of insight or intuitions (Haidt, 2001, p. 818).

Many people have taken issue with Haidt's post-hoc stance on moral reasoning, claiming that we are far more responsive to reasons and more active moral reasoners than he supposes. Cordelia Fine cautions against the dangers of over-interpreting the pervasiveness of post-hoc reasoning:

> ...a direction of causality from judgement to reasoning is only plausible if the individual does indeed lack access to the processes that caused their moral judgement. For this reason, we must be careful not to over-interpret evidence that we sometimes make post-hoc and erroneous justifications of our judgements...If a moral judgement is, contra Haidt, in fact largely determined by conscious controlled processes (either moments before, or on prior occasions), then

> we will presumably be correct in supposing that those reasons were causal in our moral judgement. (Fine, 2006, p. 96)

Distinguishing between post-hoc and genuine reasoning is an extremely difficult task. Importantly, Haidt's work touches upon some interesting and common psychological biases which show us that it would not be wise to use introspection to gauge whether one's own moral reasoning is genuine or post-hoc. To verify Haidt's strong claim that moral reasoning is usually post-hoc, judgements would need to be empirically demonstrated to consistently occur post-hoc. Only then could the phenomenon be used to support the view that moral judgements do not depend upon moral reasoning. If most judgements are found to occur without reasoning processes and one maintains a meta-ethical view that requires reasoning, one will be forced to say of many judgements that we ordinarily take to be moral judgements, that they do not qualify as moral judgements. The empirical data on the pervasiveness of post-hoc reasoning has not yet been developed.

Prinz (2016) claims that Haidt's commitment to the idea that there is a high prevalence of post-hoc reasoning in moral judgements, whether accurate or not, is not required by any theory of moral judgement which claims that non-cognitive mental states or mental state components are necessary for moral judgements. He notes that Haidt's emphasis on the idea that reasoning is predominately post-hoc has led to the general impression that reasoning is generally causally inert. Consequently, evidence of *mere* responsiveness to reasons has been taken as evidence against non-cognitive and hybrid-state views. As discussed in Section 5.1, there is ample scope for reasoning in relation to morality, even if moral judgements themselves are fundamentally non-responsive to reasoning.

In the following section I will discuss the main approaches to addressing the concerns raised by moral dumb-founding studies.

5.3. What to say about cases of moral dumb-founding

Given the moral content that seems to be missing in moral dumb-founding cases, advocates of cognitive, sophisticated non-cognitive, hybrid-expressivist, and moral thought pluralist conceptions of moral judgement need to either explain why the judgements in moral dumb-founding cases are moral ones or deny that they are. In this section I will discuss four possible ways that cognitive theorists and more sophisticated non-cognitive theorists (and by extension hybrid-expressivists and moral thought pluralists) can respond to moral dumb-founding cases. The

first and second ways of responding are generally taken by those who find it counter-intuitive to deny that the kinds of judgements made in moral dumb-founding cases are moral judgements. They argue that, despite appearances, there *are* cognitive moral beliefs or moral values in the proposed examples, or they argue that there *were* moral beliefs or moral values at some earlier point in time. The third way that cognitivists and sophisticated non-cognitivists may respond is by arguing that the judgements made in moral-dumbfounding cases do not qualify as moral judgements. All three lines of objection reflect the cognitivist's and the non-cognitivist's commitment to moral judgements being of a certain kind of mental state. Lastly, they may argue that the kinds of judgements made in moral dumb-founding cases are moral ones of a kind not yet discussed.

There *are* or there *were* moral beliefs or values.

There are *beliefs or moral values.* Many people have the intuition that in the case of Julie and Mark, people who continue to judge their having sex as morally wrong are still making a moral judgement despite their not being able to say why the actions are morally wrong. Those who think that the expression of a feeling is not sufficient for moral judgements may maintain the view that people are making moral judgements about Julie and Mark by arguing that the judges in this case are, despite appearances, expressing moral beliefs or values. This objection rests on the claim that people have a general inability to properly engage with thought experiments.

Rozyman, Kim and Leeman (2015) quite rightly point out that in hypothetical scenarios such as the ones discussed, people have trouble accepting the thought experiments at face value. When we encounter thought experiments, we import beliefs about them and fail to let go of beliefs which we are pre-emptively and explicitly told are not true. Related findings reveal that people have difficulty relinquishing beliefs even after they are told of their falsity (Ross, Lepper and Hubbard, 1975). Ferguson and Sanford (2008) and Ferguson, Sheepers and Sanford (2010) have found that "... people routinely anchor fictional content in real-world knowledge, finding it difficult to comprehend information about a fictional universe that contradicts their real-world assumptions" (cited in Rozyman, Kim and Leeman, 2015, p. 298).

If people are generally bad at taking on thought experiments, then the judgements made in the Julie and Mark scenario do not tell us anything useful about morality at all. There is evidence suggesting that people who are presented with the Julie and Mark vignette are unable to silence their beliefs about the harmful consequences of the act for

the sake of the thought experiment. Rozyman, Kim and Leeman (2015) tested people's beliefs about the Julie and Mark thought experiment. In particular, they asked participants their beliefs about a) whether Julie and Mark would abstain from having sex again; b) Whether they would really keep their sexual encounter a secret; c) Whether their relationship would be negatively affected in the future; d) Whether they would suffer any negative consequences, and e) What real-world effects would come from having sex with one another. They found that individuals tended to strongly believe that there would be severe real-world consequences; they disbelieved that the siblings sexual encounter would have no negative consequences, including negative consequences for their relationship; they disbelieved to a lesser extent the claim that the encounter would be a once off occurrence, but tended to believe that the siblings would keep it between themselves (Rozyman, Kim and Leeman, 2015, p. 300).

Rozyman, Kim and Leeman (2015) also tested whether participants who were willing to acknowledge that they were morally-dumb-founded by endorsing the following statement: "I don't have a strong reason at this point, but I just feel it's wrong for them to do what they did", truly believed that Julie and Mark's actions were harmless (2015, p. 302). They report that all such participants had doubts about the harmless consequences aspect of the vignette (2015, p. 303). Rozyman, Kim and Leeman (2015) suggest that subjects of moral dumb-founding studies did not change much in their beliefs about the act's harmfulness, but, rather, they changed in what they were 'willing to express'. They claim that this casts doubt on the idea that individuals were truly morally dumb-founded (Rozyman, Kim and Leeman, 2015, p. 305–308).

Given this evidence, it may be that participants' judgements in moral dumb-founding experiments are wrongly thought to be non-belief like or non-value laden. That is, people may have reasons for their judgements because they do not accept the thought experiments to begin with. But, we are now in a position where we cannot tell whether the reasoning (either a moral belief or a belief about the relevance of a situation to a person's value) has occurred prior to the feeling or if the belief that there must be harmful consequences is being generated post-hoc to support an intuitive judgement. The fact that participants give reasons which are negated within the thought experiment, such as the possibility that Julie and Mark may have a deformed child, is consistent with both the idea that subjects are bad at taking on thought experiments, and the idea that intuition is followed by post-hoc rationalisation. Prinz argues that the unwillingness of people to change their minds once reminded that certain factors are irrelevant

suggests that the supposed reasons were not vital to the formation of the judgements (2008, p. 31). In other words, the scrambling and jumping from one reason to another suggests that it is highly plausible that justificatory reasons are being invented as Haidt claims, to support an already made intuitive judgement.

There were *beliefs or values.* A second critique of dumb-founding experiments suggests that individuals in dumb-founding cases made complex moral appraisals unconsciously. The content of the appraisals is inaccessible to them now, even upon reflection. The ability to make complex appraisals unconsciously is possible because the complex moral beliefs or sophisticated non-cognitive values that the appraisal is comprised of were made at a previous time in the judge's life. In other words, the feeling of approval or disapproval that individuals have in dumb-founding cases is produced by an appraisal which contained a moral judgement one had made in the past, and the judge was consciously aware only of the resulting emotion. Cognitivists claim that a complex moral belief is present in the unconscious appraisal; non-cognitivists claim that an unconscious value is present; hybrid-expressivists claim that both moral beliefs and some kind of non-cognitive mental state is present, and moral thought pluralists claim that either a moral belief or moral value is present. As I will explain, there seems to be no problem with claiming that appraisals are regularly processed unconsciously. However, all of these theories are committed to claiming that the moral beliefs and/or values were previously, in their creation, conscious.

Cordelia Fine is a proponent of the idea that complex moral appraisals can take place unconsciously. She argues that the moral dumb-founding literature does not pay due diligence to the ways in which our automatic processes can arise. She says that it is at least plausible that some of our automatic judgements only become so after conscious reflection or reasoning has taken place at a previous time in an individuals life (Fine, 2006, p. 93). She draws on the observation that we regularly automatise behaviours that were once conscious. The essence of this idea can be captured for most people by highlighting how many of the things we learn can eventually be done quite automatically, for example, most people can recall how consciously taxing driving a car was when they first began to drive and can compare that to the comparatively automatic ease with which they take to the road as an experienced driver. In the moral case, she draws on the work of Bargh and Chartrand (1999):

> Bargh and Chartrand (1999, 476) described automatic processes as '"mental butlers" who know our tendencies and preferences so well that they anticipate and take care of them for us, without having to be asked'. According to the influential auto-motive model, when a goal is consciously acted upon repeatedly and consistently in a particular situation, the goal becomes automated through its repeated selection (Bargh 1990). The goal is automatically triggered by the situation, in the absence of conscious intent. Bargh et al. (2001, 1015) argue that, 'on the basis of the assumption that goals become automated through their repeated selection in a given situation, such automatic goals should generally be in line with the individual's valued, aspired to life goals and purposes. (Fine, 2006, p. 92).

In this way, judgements made by morally dumb-founded individuals may be the consequence of the unconscious deployment of cognitive or non-cognitive moral judgements that were once conscious. In other words, an individual may arrive at a cognitive moral belief or non-cognitive moral value, and though once these judgements were effortful and conscious, subsequent instances and repeated exposures have led to that particular belief or value being capable of unconsciously activating and pairing with a relevant mental state of the opposite direction-of-fit to form emotions.

In further support of her view, Fine claims that it is a problem for Haidt that people do not always make moral judgements in accordance with their automatic intuitions: "... it is interesting to note that the vast majority of individuals (over 90 percent) report discrepancies between their privately experienced 'should' versus 'would' responses to stereotyped groups" (Fine, 2006, pp. 86, 94). She claims that the fact that people take issue with their automatic judgements, she believes "... in the absence of social pressure or persuasion...", and do not try to justify them is evidence against Haidt's claim that moral judgements are affectively laden moral intuitions. Fine claims that, in the way described by Baurgh and Chartrand, our consciously preferred judgements may potentially over-ride our unwanted intuitive attitudes and may even eventually become automatic (Fine, 2006, p. 94). By way of response, Haidt notes that "the tight connection between flashes of intuition and conscious moral judgements... is not inevitable: Often a person has a flash of negative feeling, for example, towards stigmatized groups... yet because of one's other values, one resists or blocks the normal tendency

to progress from intuition to consciously endorsed judgement." (Haidt and Bjorklund, 2007, p. 818, cited in Kennett and Fine, 2009, p. 90). In this way, intuitions are pitted against other intuitions; they are not pitted against some other kind of superior judgement as is suggested by Fine. The general theme of this debate about 'automatic' and 'conscious' judgements is discussed further in Section 6.3.

The moral dumb-founding experiments claim to show that we use feelings to decide whether something is morally right or wrong. The response that there were moral beliefs or values admits that phenomenologically, the feelings do come first for us (at least in these cases). However, for cognitivists, sophisticated non-cognitivists, and hybrid-expressivists, the emotions must consist of certain kinds of cognitive or non-cognitive mental states which were made at some point in the judge's history. If there was an unconscious appraisal containing moral content which caused the emotion, reasoning or searching for moral reasons is not a post-hoc process. It is an attempt to identify the *real* moral beliefs or values which are at the source of the emotion, or that comprise the appraisal. If, however, there are no unconscious moral beliefs or values comprising the appraisal, reasoning is indeed post-hoc.

The idea that moral reasoning has taken place at some prior stage is an interesting reply. The main problem for the view, as brought to light in dumb-founding cases, is that the appraisal components cannot be identified by experimental subjects upon reflection (Prinz, 2016). As I have expressed, cognitive moral judgements must contain moral content, meaning the representations involved must employ moral concepts such as 'good', 'bad', 'right', and 'wrong' that have a purely mind-to-world direction-of-fit. These are quite complex and abstract representations which involve conscious processing at some point. This high demand for processing makes it reasonable to expect that we should be able to find beliefs with moral content when we look. For example, the kind of belief that 'x natural property is bad' is something that should have at one time been conscious. Similarly, values, as opposed to other kinds of non-cognitive mental states are, on many views, our more highly preferenced desires. The ability to preference some desires over others involves some level of reflection and conscious processing. For example, it may involve the recognition that one has multiple desires. It may involve the recognition that there are conflicts in the fulfilment of ones own desires (perhaps through time), as well as conflicts between people. It involves a preference for which desires be fulfilled given the conflicts. Again, this high demand for processing reduces the plausibility that the components of a moral appraisal can remain unavailable to us consciously, particularly upon reflection. There is an additional

layer of complexity for the Humean appraisal theorist that the anti-Humean besire theorist does not encounter, namely, Humean appraisal theorists think that the relevance of cognitive mental states to non-cognitive ones is cognitive. Having said that, it is possible that this processing is done unconsciously, or that, if it was conscious initially, subsequent processing may occur unconsciously. But, the finding that we cannot work backwards to identify the content of the components which generated the emotion requires explanation.

If one is unconvinced that moral dumb-founding cases really do contain moral beliefs or values either because the thought experiments are not accepted or because reasoning has occurred at a previous time, one will have to argue either that the judgements being made are not moral judgements, or that they are moral judgements of a kind not yet specified.

They are not moral judgements. Cognitivists and sophisticated non-cognitivists also have the option of denying that individuals in moral dumb-founding cases are making moral judgements. The significance of taking this option is dependent upon how often people make judgements that have moral significance in the absence of any cognitive moral beliefs or values. The inclination to deny that unjustifiable judgements are moral ones likely results from the belief that some feelings are the result of an appraisal containing a moral judgement and others are demonstratively not (as was the case for the induced disgust studies in Section 4.2). One may argue that for a judgement to be a moral one it just needs to contain a moral judgement of a certain kind whether or not the individual making it can identify it. This line of thought appeals to those who claim that judgements made in moral dumb-founding cases are moral ones because there are or were beliefs or values that their feeling is based upon. Another option is to claim that the reasons that people have for moral judgements *must* be accessible to them upon reflection. Consequently, morally dumb-founded individuals are not making moral judgements. Lastly, one may claim that a judgement's status as moral in cases where individuals cannot say why they hold the view that they do depends partially on whether the person making the judgement is willing to call it a moral judgement and whether they are willing to make demands of others based on it. This final option is discussed in the following section.

The main problem with denying that judgements made in moral dumb-founding cases, or on the basis of incidentally induced feelings, are moral ones, is that it often seems counter-intuitive when people use their feelings as information in deciding what is right or wrong without

too much concern for explicitly justifying those feelings. These kinds of judgements are, at least at times, strongly imposed on others in the name of morality. That is, it seems plausible and likely that these kinds of feelings motivate the people who have them to put restrictions on other people's behaviour. I will discuss what to say about whether it is best to think of these kinds of judgements as bad moral judgements or not as moral judgements at all in the following section and in Chapter 6.

They are moral judgements of a kind not yet specified. There is one further kind of possible response available to cognitivists, non-cognitivists and moral thought pluralists. They can concede that in moral dumb-founding cases specific moral beliefs and values are missing. That is, they can claim that the judgements are lacking cognitive and non-cognitive moral content. The judgements do, however, share a feature in common with cognitive and sophisticated non-cognitive moral judgements. Specifically, there is an element of endorsement in dumb-founding cases, as well as in incidentally induced sensation studies, that has not yet been discussed. The subjects have a feeling which they themselves, rightly or wrongly, believe to be *morally significant*. This point explains our interest in the question of whether the the individual who is making a judgement takes the act they are judging to be *morally* wrong as opposed to just disgusting. This kind of endorsement is similar to that of a hybrid-expressivist account whereby someone endorses a non-cognitive mental state. It is different because, for the hybrid-expressivist a *specific* non-cognitive mental state is endorsed, such as a desire or goal, by a moral belief that can stand on its own. That is, the belief is not held *because* of the feeling. In dumb-founding cases a feeling with numerous potential sources is endorsed. The endorsement consists of a belief that the feeling they experience is relevant to morality *somehow*. That is, the endorsement is unspecific.

A cognitivist, sophisticated non-cognitivist, or moral thought pluralist may claim that the belief that some feeling has moral relevance is only possible for individuals who have a background of cognitive or non-cognitive moral judgements. That is, in dumb-founding cases, without the belief that the action being judged is somehow relevant to a class of things I take to be morally bad, or that the action somehow conflicts with some moral value of mine, I am not making a moral judgement. Without concepts containing cognitive moral beliefs or non-cognitive moral judgements, the belief that some feeling has *moral* relevance is not possible. Feelings believed to have moral relevance which cannot yet be identified should not be considered 'proto-moral' judgements or 'primitive' judgements that come *before* reasoned moral beliefs or values.

Their existence is dependent upon a history of development of cognitive or non-cognitive moral judgements of the more complex kind. Without them, one could not say that the feelings of approval or disapproval, disgust etc. signify 'moral' wrongs.

Whether the belief that an action is morally relevant is present is an additional question of importance to the question of whether the causal history of the feeling one has is incidental or caused by unconscious moral judgements. What is important is an individual's willingness to make impositions on others. If someone is willing to restrict, control, or criticise others behaviours or attitudes despite not knowing how or if their feeling really is connected to their moral beliefs or values, then they should be held morally responsible for that endorsement (to the extent that their capacities warrant). This option differs from the objection that there were beliefs or values as it does not claim that there are or were any specific beliefs or values occurring unconsciously to form the appraisal. Maybe there were, maybe their weren't. Rather, there is a belief that some action, situation, or character in general is linked to ones body of moral beliefs or values. This is a commitment worthy of moral appraisal in its own right.

5.4. Concluding remarks

Moral-dumbfounding research indicates that people make moral judgements in the absence of distinctly moral reasons. Participants in dumb-founding studies try hard to identify beliefs or desires guiding their moral judgements and come up empty handed. However, the judgements are taken by the individual who has them as more than a personal preference, as indicated by their insistence that the actions in the scenarios are *morally* wrong, and not just disgusting. Individuals appear to rely on feelings or intuitions as information when making these moral decisions. The emphasis on the appeal to feelings as a means for making moral decisions has been claimed to favour any kind of non-cognitive theory of moral judgement. At a closer examination this research cannot support all non-cognitive theories of moral judgement. It only has the capacity for supporting hybrid-state theories and simple non-cognitive theories of moral judgement. It undermines cognitive and sophisticated non-cognitive theories of moral judgements, as subjects fail to identify any cognitive moral beliefs or non-cognitive moral values. The dumb-founding literature raises interesting issues in the debate between cognitivists, non-cognitivists, and hybrid-state theorists. However, it provides very little of use in the way of supporting or challenging cognitivism over sophisticated non-cognitivism (or vice versa).

There are four possible ways of responding to moral dumb-founding studies. One can claim that:

(1) Judgements made in moral dumb-founding cases are moral judgements because there are moral beliefs or values.
(2) Judgements made in moral dumb-founding cases are moral judgements because there were moral beliefs or values.
(3) Judgements made in moral dumb-founding cases are not moral judgements.
(4) Judgements made in moral dumb-founding cases are a kind of moral judgement. Feelings are endorsed as being relevant to a body of moral beliefs or values, but their specific relation is as yet unknown.

It is a problem for options (1) and (2) that individuals are unable to identify reasons for their judgements when given time to reflect. Psychological studies revealing the phenomenon of post-hoc reasoning and human biases also serve to weaken any strong intuitions that one may have about the role of reasoning in arriving at moral judgements. Option (3) denies that morally dumb-founded judgements are moral ones, worthy of moral praise, blame or responsibility. Option (4) focuses on the fact that there is some kind of belief that the feeling or intuition one has is relevant to morality *somehow*, but one cannot yet say how. That is, there is a belief that the feeling must in some way be the product of their cognitive moral beliefs or non-cognitive moral goals. Cognitivists and sophisticated non-cognitivists may claim that this kind of belief is only possible when an individual already has cognitive moral beliefs and/or sophisticated non-cognitive values. This category of unspecific moral judgements explains why it is of interest to us to know whether or not morally dumb-founded individuals take themselves to be making moral judgements.

5.5. Concluding remarks on the empirical literature

In the previous two chapters I have been discussing whether the empirical literature has contributed to helping determine whether one of the traditional theories of moral judgement (cognitivism or non-cognitivism) or one of the three hybrid theories (hybrid-state theory, hybrid-expressivism, or moral thought pluralism) provides a better conception of moral judgements. I explored three significant empirical lines of investigation. The first examined the extent and significance of the correlational data between stereotypical moral judgements and emotions, finding that the data cannot provide any basis for preferring one theory of moral judgement over another. The second looked at whether

people can be experimentally shown to use feelings as information in deciding right from wrong. The data on this was found to be inconsistent; sometimes people do and sometimes they don't. All theories of moral judgement are able to explain why feelings may be used as a source of information in making moral judgements. The third empirical line of investigation examined whether *complex* cognitive moral beliefs and non-cognitive moral values are necessary for moral judgements. Moral dumb-founding experiments provide a means for examining the complexity of the content of appraisals. These experimental studies seem to cast doubt on the view that moral judgements involve reasoning in the form of moral beliefs (which contain explicit moral concepts) or values consisting of appropriately endorsed non-cognitive mental states, challenging both cognitivists and sophisticated non-cognitivists.

In response to moral-dumbfounding cases one can claim that there are beliefs or values, or that there were; one can deny that morally dumbfounded individuals make moral judgements; or one can claim that the judgement is a kind of moral judgement because it is partially endorsed, but lacks the kind of endorsement that cognitive and sophisticated non-cognitive moral judgements have. This final possibility is explored further in the following chapter.

CHAPTER 6

Endorsement and moral judgements

In this work I have sought to explore the nature of moral judgements. I have proceeded with the view that theorists should aim to provide an explicative definition, in which any kind of restrictions or claims about what moral judgements are need to be justified. But, as is the case with most conceptual development, there is inevitably some degree of weight that must be given to starting intuitions about what it is that we are aiming to describe. In this chapter I explore the possibility that the notion of endorsement may capture our intuitions about what is essential to moral judgements. I will begin the chapter with some brief comments on general concept development before discussing the development of moral concepts.[1] I will then explore the notion of endorsement and its relation to moral judgements. Specifically, I explore the idea that there may be both cognitive and non-cognitive beliefs and desires which have the feature of endorsement in common, rendering them both the kinds of mental states which we think should think of as moral judgements apt for moral praise and blame.

6.1. Concept development

General concept development. Think of the basic concept of a chair, the dictionary definition of which is "a seat typically having four legs and a back for one person" ("Chair", 2018). In the world, there exist multiple cases of this group of features. People regularly refer to objects which are either missing one of these features or which have additional features, as chairs. For example, someone might refer to an object as a chair despite it not having four legs. The appropriateness of using the label for cases in which the object being described does not perfectly match a typical example will depend on a variety of factors, some of which include, whether alternative levels of description are available; what concepts the intended recipient of the concept is capable of processing; and what the particularities of specific contexts happen

[1] A detailed discussion of the nature of concepts is beyond the scope of this thesis. For further reference see Margolis and Laurence (2014), Rosch (1973), Lewis (1972), Kripke (1980) and Wittgenstein (1953).

to be. For example, there may be cases where referring to both a stereotypical chair and a stereotypical stool as chairs is appropriate and intended, e.g. you might ask your removalist to put all of your chairs in the dining room, whereby 'chair' you just mean things with the feature 'object designed to be sat on'. There may, though, be cases where your intended usage is more restricted. For example, if you tell someone that there is a chair that they can use in the next room, you may be inclined to also warn them to avoid the stool (if, say, there is also a broken stool in the room), despite your original sentence technically expressing precisely what you meant. People generally adapt their usage with relative ease to help others navigate these ambiguities. They are also quite good at re-interpreting the intended meaning of others when they are presented with incorrect or less than ideal concepts, for example, if someone is asked to retrieve the brie from the refrigerator, many people will return with camembert if there are no other cheeses in the refrigerator. All of these issues arise because there are overlapping features between the concepts. Of course, whether more fine grained concepts are developed in the first place usually depends on whether the distinctions being made are significant or useful for communication. For example, there are more words for 'snow' in countries where it snows a lot, presumably because communicating more fine-grained differences is useful for the people who live there. Distinctions draw attention to the addition or absence of some features[2]. But, even if new concepts are not formed to distinguish objects with overlapping, but different, features, the underlying differences in the objects still remain.

Moral concept development. I will now explore these issues in relation to moral judgements, looking at the problems that arise when various features are proposed to be necessary or sufficient. I assume that 'moral judgement' is a label for some naturally occurring pattern of interest in the world. It is possible that the term 'moral judgement' is an umbrella term with more than one kind of moral phenomenon falling under it, analogous to the term 'seat'. 'Seat' means "a thing made or used for sitting on" of which there are many varieties, such as 'chair', 'couch', and 'stool' ("Seat", 2018). The differing varieties of seats have overlapping features and no particular kind of seat has a naturally privileged status.[3] If 'moral judgement' is a term like 'seat'

[2]You may also think that there are some distinctions that are important or interesting for their own sake; even if they are never discovered by, described by, or of any use to any person.

[3]A specific kind of seat can gain a privileged status by circumstance, for example, stools have a privileged status in the context of high tables.

it may encompass a cluster of more nuanced varieties which are not yet appropriately labelled or commonly distinguished from one another. If there is more than one kind of moral judgement (as there is more than one kind of seat) it may be useful to develop the terminology to be able to discuss them. However, it may be the case that the broader level of description analogous to 'seat' is insufficient as there is some feature which a particular kind of judgement has (analogous to a 'chair' or 'stool') which makes it more representative of the label 'moral judgement'. That is, maybe it is not the feature which all seats share that is important, but the presence of some specific feature of a particular kind of seat which provides the significance of the original pattern of interest. The stakes of understanding the commonly involved features and their significance is much higher in the moral realm than in the ordinary case of distinguishing various kinds of seats. People are regularly rewarded and punished for their moral judgements. For this reason alone, restrictions or expansions of the concept 'moral judgement' enacted by adding or subtracting features must be justified. Importantly, we are looking for distinctions which go beyond facilitating conversation by enriching the descriptive powers available to people. We need to justify those restrictions or expansions. This is what is meant by providing an explicative definition.

The general strategy of the meta-ethical positions I have discussed in this thesis is to restrict the category of moral judgements by giving precedence to a particular kind or combination of mental states that are thought to be involved.

(1) Cognitivists claim that moral judgements express beliefs of some kind.
(2) Non-cognitivists claim that moral judgements express desires of some kind.
(3) Hybrid-state theorists claim that moral judgements express besires.
(4) Hybrid-expressivists claim that moral judgements express some non-cognitive mental state which is endorsed by some kind of cognitive belief with moral content.
(5) Moral thought pluralists reject the restrictions outlined above. They think that moral judgements can be either cognitive or non-cognitive.

Cognitivists are generally drawn to their position because they take moral judgements to be truth-evaluable. Non-cognitivists are predominantly drawn to their position because it makes sense of the motivational nature of moral judgements. Hybrid-state theorists and

hybrid-expressivists are motivated to find a way to explain how moral judgements can be both truth-evaluable and motivating. As I will explain, moral thought pluralists are not motivated to explain these features explicitly. They must draw the category in another way. It is the least restrictive position in terms of the kind of mental states that are claimed by each theory to be involved in moral judgements.

Both hybrid-expressivists and moral thought pluralists reject a common presupposition in meta-ethics that moral judgements must be an expression of just one kind of mental state. For hybrid-expressivists, moral judgements contain both beliefs and desires. They may differ in their views of the kind of content that those beliefs and desires contain For example, the belief may be about what is morally good or bad, or it may be about which kinds of non-cognitive mental states are merited (although, it is possible that beliefs about moral properties and merit require a more fundamental moral judgement which should possibly be taken to be sufficient as moral judgements in their own right). And, the non-cognitive mental state may be a basic desire or it may be a non-cognitively endorsed desire. Hybrid-expressivists can still appeal to truth-aptness and motivation as being essential features of moral judgements. Moral thought pluralists cannot. In fact, they deny both that moral judgements are necessarily truth-apt and that they are necessarily motivating.[4] They require a novel way of defining moral judgements which is capable of unifying different kinds of mental states. This chapter discusses the possibility that a certain kind of endorsement may be able to play that role.

If some cognitive and some non-cognitive judgements can be demonstrated to share a morally significant feature, the burden of proof rests with advocates of more restrictive definitions to justify any restrictions of the concept. That is, there should be a presumption in favour of the broader, moral thought pluralist, conception until such time as some argument for restriction (such as an appeal to motivation or truth-aptness) is found to be convincing.

Below I will restate the common claims of the cognitivist and non-cognitivist. Both cognitivists and non-cognitivists agree that not just any cognitive or non-cognitive mental states should count as moral judgements. I will adopt some reasonable assumptions about the kinds of beliefs or desires that each respective view takes to be 'moral' ones before before going on to discuss the similarities of those beliefs and desires. In brief, moral beliefs require the explicit use of moral concepts.

[4]Cognitive moral judgements are necessarily truth-apt and non-cognitive moral judgements are necessarily motivating.

For example, for someone to have a moral belief they must believe that something is 'good', 'bad', 'right' or 'wrong'. And, in relation to moral desires, I will adopt Simon Blackburn's view that not just any desire should count as a moral one; that our tendency to preference our desires and to demand that others share certain desires is what elevates some desires into the moral realm. I am not defending Blackburn's view in particular. There may be other non-cognitive theories of moral judgement which say that moral judgements are a kind of endorsed non-cognitive mental state. Variations of these views may have different implications for the issues that I describe. It is not my aim to explore all possibilities, but rather to discuss the role of endorsement in relation to a couple of widely accepted meta-ethical positions. The general approach taken in this case, however, can be replicated when analysing alternative positions.

Cognitive moral beliefs. Beliefs are mind-to-world representations. They are the only category of cognitive mental states which are taken to resemble reality by the person who is experiencing them. That is, beliefs represent how things are, were, or will be (they are assented to). If an individual takes their hallucination to represent reality, then that representation is a belief to them. By contrast, if they recognise that their representation is a dream or a hallucination then they do not assent to the content of those states, so they are not beliefs. Beliefs are mental states which involve the believer having at least some degree of commitment to the truth of the proposition. The level of commitment one has to a belief comes in varying degrees; that is, individuals can believe something more or less. Beliefs with a high degree of certainty seem closed, settled, or non-negotiable to the holder. To distinguish a cognitive *moral* belief from a cognitive *non-moral* belief we need to look at the *content* of the belief. To be a cognitive *moral* belief, the belief must make reference to morality or contain moral terms such as 'good', 'bad', 'right', and 'wrong'.

In the case of cognitive moral judgements, the phenomenological feel for the truth of the judgement may come from the belief that one has identified moral properties in the world or identified some fact about the world. For example, an individual may believe that the property of badness is identical to the natural property of pain. If an individual believed that 'good = natural property x', then they can make claims such as 'you ought not to x if you want to increase the good'. This claim may seem true to them in the same way that their beliefs about other truth evaluable claims are (such as the claim that $1 + 1 = 2$). The locus of the moral judgement is contained entirely within the cognitive

domain. That is not to say that the judge will be in any way inclined to act in accordance with their belief. Importantly, the desire is only necessary for motivation and not for deciding what is morally good or bad. But, should they have a related desire, the moral belief may contribute to motivating them to act (assuming the Humean theory of motivation is correct). It is plausible to expect that in the course of everyday moral discourse, the expression of ought claims which contain moral beliefs that we are not motivated by would be rare as we tend to verbalise ought statements (which contain a moral belief) *because* they are relevant to some goal that we have. In other words, we should expect a very large proportion of ought claims which contain moral beliefs to be expressed because their expression aids in the fulfilment of related goals which are, in fact, held by the individuals expressing them.

Non-cognitive moral desires. Non-cognitivists claim that moral judgements express a world-to-mind mental state. Most proponents of this view do not think that just any non-cognitive mental state should be considered a moral judgement. As noted, Blackburn (2002), thinks that moral judgements are those non-cognitive mental states that we demand that others share. Intuitively, the idea is that we have many preferences for how we want the world to be, some of which we do not mind if others share, some of which we prefer others to share; and others which we demand that others share. Only the latter encompass our moral judgements. To make 'ought' claims requires one to have some goal in mind, e.g. 'you ought to x if you want y'. If a judgement is a moral one of the kind described by Blackburn, that goal will have a feeling of closed-ness to it, so much so that it seems odd to state the goal. Moral oughts are more like commands because the goals are like commands. We take non-moral goals to be open to rejection. For example, consider the following claims: 'you ought to go to the festival if you like music', 'you ought to switch your lights off during the day if you want to lower your electricity bill', or 'you ought to buy tickets to the football early if you want to guarantee your entry'. The 'if's' in all of these examples are, for most people, open or conditional in the sense that we genuinely do not mind if other people reject them, for example, someone could say that they don't care about music, wasting money, or the football. We may think that they are strange for rejecting such goals, but we do not demand of others that they should share such goals. We do not view all of the goals of ought statements as open to rejection in this way. There are many goals that we judge others for having (or for not having) and the severity of the judgement varies immensely. For

example, the claim that 'you ought to work on your garden if you don't want the neighbours to think poorly of you' will only motivate someone if they do in fact care about what the neighbours think of them, and we can imagine that someone may not be bothered by such concerns. Somewhere along the line, though, we begin to have an opinion about whether certain goals should be rejected and it is easy to imagine that someone might take the goal of 'having the neighbours like you' as closed in this way. This tendency to experience a sense of closed-ness to the rejectability of various goals is the thing that transforms a personal preference into a moral preference on this view.[5]

The desires that we demand others share are ones which are prioritised over other desires. In other words, our moral desires are born from our recognising that not all desires can be fulfilled. In fulfilling one desire, another is compromised. When we make a decision that the fulfilment of one desire is more important to us than one that it conflicts with, we create and reveal our values. These endorsed desires are described as 'second order desires' by Frankfurt (1971). There are other ways a non-cognitivist may differentiate non-cognitive moral judgements from other kinds of non-cognitive mental states and moral demands from non-moral demands; but generally it will involve our committing ourselves to some goals over others in some way. Importantly, endorsement of a non-cognitive mental state can be a non-cognitive or a cognitive act. Here I am referring only to non-cognitive endorsement, where we endorse or have preferences for some non-cognitive mental states not for any reasons, but simply for the fact that a certain non-cognitive state is more appealing to us (if the desire is endorsed cognitively then the position is a hybrid-expressivist one). This non-cognitive view agrees that reasoning cannot give us our ends.

[5]Conventions are an interesting case which on the surface are a collection of ought statements which some people morally demand others follow, but which they don't take to be moral in themselves, for example, burying versus cremating the dead, using a knife and fork as opposed to chopsticks to eat. It is possible that it is straying from convention that is immoral (despite admitting that one way is not objectively preferable). Conventions can be recognised as both morally demanded and arbitrary. I would propose that they are recognised as conventions because people either demand them to a lesser extent than they demand moral oughts (probably not always the case), and/or more likely, they recognise that their conventional demand is merely conventional yet are willing to demand the actions anyway, perhaps because they believe that the upkeep of conventions aids in preserving some other morally desirable state of affairs (for example, conventions make people more predictable which is good for a society). In this way conventions may be instrumental and we might be able to explain why some people have strong feelings about the rightness or wrongness of the violation of conventions.

Like beliefs, non-cognitive world-to-mind mental states can come in varying strengths. Moral judgements, if they have this direction-of-fit, seem to inherently involve a high level of endorsement on behalf of the person who holds them. That is, part of what makes certain non-cognitive mental states one's moral judgements, as opposed to another world-to-mind mental state, is that they are of high importance to their possessor and they are enduring rather than fleeting for the individual who has them. This also explains our tendency to maintain relatively consistent moral opinions over time. On this view, the capacity to endorse or preference mental states can be considered the mechanism by which we are able to have any values or sentiments at all.

6.2. Endorsement

An important similarity between the two kinds of cognitive and non-cognitive judgements above is that they can both be said to involve a particular kind of *endorsement* on behalf of the individual making them.

Psychologically categorical moral desires. The nature of the non-cognitive moral judgements described above brings to mind Kant's distinction between hypothetical and categorical imperatives. Kant famously sorts 'ought' claims into those where one can reasonably deny a goal if one chooses from those which he claims it is irrational to deny (Kant, 1950). This idea forms the basis of a strong view that reason can uncover objective ends, meaning imperatives we must comply with.[6] The non-cognitivist denies that there are any such ends. However, Jesse Prinz introduces the notion of a *psychologically categorical judgement* to describe a weaker version of Kant's categorical judgements that can be obtained for the non-cognitivist. On this view, it is not categorically wrong, for example, to stab someone because it is irrational, but it is categorically wrong due to our stable psychological commitment to some end. These judgements have an advantage over the Kantian position in their ability to explain the apparent motivational nature of moral judgements. Prinz describes it as follows:

> Like hypothetical imperatives, moral rules depend on ends, and those ends are entirely contingent. One might say that moral rules are neither strongly categorical, nor strictly hypothetical. They are psychologically categorical (we see them as intrinsic ends), and metaphysically hypothetical (if we didn't have those

[6]Some version of this Kantian account is available to cognitive theorists.

> ends, they would have no authority). Kantians who insist on a strong categorical status for morality sever the link between moral demands and human wants. In trying to free moral rules from inclinations, Kantians deprive moral rules of the motivational force that makes them act as imperatives in the psychological sense. Moral rules become categorical, but far less imperative." (Prinz, 2008, p. 136)

Prinz has a specific kind of mental state in mind as being psychologically categorical (a hybrid mental state). I will not go into his position here. However, I will note a slight departure that I take from Prinz's view of the nature of psychologically categorical desires. That is, they can be defined synonymously with second order desires. One way to distinguish moral second order desires from non-moral second order ones is by appeal to Blackburn's view that moral desires are ones we demand other share. The significance of both second order desires and the 'demand that others share' part of Blackburn's description of moral judgements is worth elaborating on. Lots of animals are capable of having desires that they demand others *facilitate* in some way, such as preferences for food, mating, and territory. We must be able to distinguish between 'brute' demands and 'moral' demands. One way to do so is to claim that moral demands are those desires that we give precedence to over other desires, and that we demand that others take on *as their own* (share), whereas brute demands involve the simple imposition of one animal's first order will over another's with no call for others to align their desires with one's own. To demand that someone *share* one's desire (in place of other potential desires) requires one to recognise that multiple, potentially conflicting, desires can exist. In this way, moral desires can be distinguished from brute desire imposition which other animals are capable of.

Psychologically categorical moral beliefs. Beliefs are representations of how we take the world to be. I have said before that to form a *moral* belief one needs to develop moral concepts such as 'good', 'bad', 'right', and 'wrong'. I have also pointed out that any mammal has the ability to represent a *particular* situation or object as undesirable for themselves. For example, a gazelle in some sense recognises that lions are 'bad', and my dog can think that baths and cats are 'bad'. But we do not tend to think that these kinds of assessments are *moral* assessments. Moral concepts seem to require some degree of abstraction. That is, when an individual claims that some particular action or property is 'good', 'bad', 'right', or 'wrong', that individual is

making an abstract claim about all actions and properties *of that kind*. The process of abstraction is also occurring when we try to identify a natural property which is common to all of the things that we think are bad. Some examples of moral beliefs may include, 'natural property x is bad (and would be bad no matter what any person thinks, feels or desires)' (Kirchin, 2012); or, 'y is bad as it is contrary to God's will'; or, 'z is bad as you could not will it to become a universal moral rule'. All of these kinds of moral beliefs require the use of quite complex and abstract concepts.

All beliefs are endorsed to some degree. That is what distinguishes beliefs from other kinds of cognitive mental states. In this sense, endorsement is either on or off; one either takes a representation to be true (or possibly true) or one doesn't. Once one has a belief, the strength of conviction that one can have in the accuracy of that belief comes in varying degrees. There are many ways that we signal the strength of conviction that we have in our beliefs to others. For example, we sometimes call weakly endorsed beliefs by another name such as an 'intuition' or a 'hunch'. It is common to signal when our confidence in a belief is low by saying things like 'it might be...', or 'I think that...'; by contrast, beliefs that we have high confidence in are usually stated as a matter of fact. We also signal our level of doubt through the tone of our voice, our facial expressions, and our general openness to discussions about the truth or falsity of the belief in question.

An individual's *categorical beliefs* are any of their beliefs one might consider to be basic such as those which we have direct sensory access to. These may include beliefs about where objects are located in ones immediate environment. Or, at a slightly more abstract level, beliefs such as 'the earth is round' and 'I have two hands'. I will call *Psychologically categorical beliefs*, second order beliefs, in the sense that they are beliefs that require abstract generalisations, such as, 'flamingos are pink', or 'spiders have eight legs'. In the moral case these beliefs are second order beliefs that explicitly reference moral concepts, such as 'good = natural property x', 'Good = God's will', or 'needless suffering is bad'.[7]

[7]The notion of psychologically categorical cognitive and non-cognitive judgements can explain moral disagreement for both cognitivists and non-cognitivists. That moral beliefs and desires can be psychologically categorical is consistent with the common intuition that moral disagreement between individuals is significant. When one person says that some action is right or good and another person says that an action is wrong or bad they are doing more than reporting on their personal preferences. They take their position to be in some way either the correct one, or the one which is preferable when there are conflicts in desires. For beliefs, the

Hybrid theories of moral judgement. There are three kinds of hybrid theories of moral judgement which all differ in the kinds of endorsement that they take to be involved in moral judgements. Hybrid-state theories claim that moral judgements are besires. It is difficult to conceptualise besires. Jesse Prinz (2004) describes them as appraisals of how one is faring in the world. The appraisal occurs when some stimuli, such as a visual image, smell, or sound, automatically triggers physiological changes in the body, the perception of which gives one information about how one is faring in the world. Consider the experience of seeing a snake, or experiencing vertigo at a great height. The representational component (the visual perception of the snake or the cliff) does not require any complex processing. And, the non-cognitive component is a basic drive for survival. Other mammals presumably also make these kinds of appraisals which do not require any process of being psychologically endorsed.

For Prinz, besires are psychologically categorical. Feelings represent 'core relational themes' which can be thought of as the non-conscious evolutionary goals or drives of survival and reproduction. I have parted ways with Prinz here by denying that core relational themes are *psychologically categorical*. I have said that for a mental state to be psychologically categorical it needs to be psychologically endorsed in that it requires the processing of more basic beliefs and desires. The non-cognitive component in a besire is not a psychologically endorsed non-cognitive mental state as it is for non-cognitive theories of moral judgement of the kind that Blackburn presents. Likewise, the belief-like component of a besire is a basic percept which does not contain *moral concepts* and so does not involve the kind of endorsement required for moral beliefs.

Hybrid-expressivists think that moral judgements express both beliefs and desires. They can differ in their views about whether the non-cognitive component needs to be psychologically categorical. That is, they can claim that any kind of non-cognitive mental state can be cognitively endorsed, or, they can claim that psychologically categorical

---

phenomenon of disagreement is straightforward; beliefs are truth-apt representations of the way the world is and people can disagree about those representations. For desires, the phenomenon of disagreement is intelligible if moral judgements are psychologically categorical. If you have desires which you personally take to be non-negotiable (perhaps because they are what you consider to be your most important desires), and you demand that others share those desires, moral disagreement can occur when others disapprove of or disregard your moral desires.

non-cognitive mental states are cognitively endorsed. For the hybrid-expressivist, the cognitive belief about the non-cognitive mental state is psychologically categorical.

Lastly, moral thought pluralists can claim that moral judgements express either cognitive or non-cognitive moral judgements which involve the kind of endorsement that cognitivists and non-cognitivists posit as individual theories.

Cognitive endorsement of non-specific feelings. Empirical investigation has been successful in demonstrating that people sometimes use bodily feelings as information when deciding whether something is right or wrong. As discussed in Chapter 4, all theories of moral judgement can explain why feelings might be regarded as a source of information. That is, when an individual has moral beliefs and desires, it is possible that those beliefs and desires unconsciously contribute to that individual forming an emotion and that the emotion is the first thing that they become consciously aware of. When this happens, their 'fast and automatic' system is engaged producing what Haidt refers to as 'the sudden appearance in consciousness of a moral judgement, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion' (Haidt, 2001, p. 6). As Jones (2006) notes these automatic judgements may become so only after being made consciously on an earlier occasion. The emotions can be explained by both cognitive and non-cognitive theories of moral judgement by claiming that the appraisal causing the emotion contains moral content of either a cognitive or non-cognitive kind.

Moral dumb-founding research and induced 'emotion' or sensation studies (discussed in Chapter 4) claim to show that there are cases where feelings are used as information for making moral judgements. Importantly, the feelings do not appear to be caused by an appraisal containing either a cognitive or non-cognitive moral judgement of the kinds described above; there are no current, and no identifiable past, appraisals containing moral content. In the case of induced sensation studies it is significant that the feelings driving the judgements are not *non-cognitive* mental states; they are sensations. There is no world-to-mind mental state involved. Consequently, the feelings are not emotions if one adheres to an appraisal theory of emotions. It seems that individuals in these studies commit themselves to the view that their *sensation* is morally relevant. The view that *sensations* can be endorsed as morally relevant is not a hybrid-expressivist one. Hybrid-expressivism requires the cognitive endorsement of a *non-cognitive* mental state which

sensations are not. In the case of moral dumb-founding studies, the source of the feelings is unclear. It is possible that the feelings are caused by unconscious moral appraisals. But it is also possible that the feelings were caused by non-moral appraisals, or by the situation's ability to generate a sensation of physical disgust somehow. As I will explain, for the form of endorsement under consideration, the source of the judgement is not important.

In both moral dumb-founding cases and induced sensation studies the individuals making judgements have no conscious access to any moral beliefs or desires, yet they take themselves to be making *moral* judgements. Do we take the individuals at their word in these cases? One option is to say that a judgement's status as moral depends on whether there is a moral appraisal or not; it does not depend upon whether they can *identify* the content of the appraisal. Another option in cases where specific beliefs and desires cannot be identified is to allow the individual's *personal standards* for what is to count as 'evidence' of what is morally good or bad determine whether their judgement is a moral one or not.

Individuals vary in their inclination to use their feelings as information when making moral judgements. Some people will scarcely question where their feelings may have come from and think that their feeling that an action is wrong is all the evidence needed to make a moral demand. Others may think that there is a moral belief or desire to which their feeling is indebted, but that they just can't think of it at the moment. Those individuals who are inclined towards the latter view are likely to be more reluctant to use their feelings to impose their judgement on others and demand that others share their view until they can identify the underlying beliefs and desires that they take to be responsible for their feeling. It is also possible that they are less inclined to engage in post-hoc reasoning.[8]

Judgements made in moral dumb-founding and induced sensation studies have an element of endorsement. As mentioned previously, it is possible that the endorsement is not the result of any *specific* moral beliefs or desires. Rather, it is a *non-specific* judgement that some feelings are *morally relevant*. This kind of judgement would not be possible for an individual who did not already have moral beliefs or moral desires. Without these pre-existing concepts, individuals can approve or disapprove of *specific* actions, but their approval or disapproval

[8]The relationship between one's view about the need to explain the source of 'moral' feelings and one's tendency to both impose and demand others obey or share ones prescription, and their tendency to engage in post-hoc reasoning can be empirically tested.

expressed in the moment will not contain *moral content*. The moral content depends upon their belief that their non-specific feeling is connected to a body of specific cognitive and/or non-cognitive moral judgements. This kind of endorsement can occur on all accounts of moral judgement except hybrid-state theories. However, it seems unspecific judgements of this kind only have the potential to be considered a kind of *moral judgement* on the moral thought pluralist view as they do not meet the stipulated definitions of moral judgement proposed by cognitivists, non-cognitivists, and hybrid-expressivists.

6.3. In defence of moral thought pluralism

Choosing between hybrid positions. I have suggested that there is an important similarity of endorsement between some cognitive and some non-cognitive judgements which may give both positions sufficient claim to the label moral judgements. This similarity supports moral thought pluralism. Moral thought pluralism is the most explanatorily simple of the hybrid positions in that it does not posit any new kind of mental states (as the besire advocate does), and it does not seek to restrict the class of moral judgements to exclude those which are either purely cognitive or purely non-cognitive (as the hybrid-expressivist does).

A low and a high bar. Hybrid-state theory (and any non-cognitive theory of moral judgement that does not require endorsed non-cognitive mental states) does not require the kinds of endorsement that cognitive and non-cognitive theories of moral judgement posit (or the capacities that make them possible). This sets a low bar for what a moral judgement can consist of. It seems that animals, children, and the mentally disabled are capable of forming besires and that they have basic non-cognitive desires that they *impose* on others. Defenders of hybrid-state theories of moral judgement will need to contribute significantly to questions about the whether other animals, children, and the mentally disabled make moral judgements. On the other hand, hybrid-expressivists set a high bar for a judgement to be considered a moral one. Hybrid-expressivists differ in their opinions about what the content of the cognitive and non-cognitive mental states involved in moral judgements must be. Their commitments in this respect determine just how high the bar for making a moral judgement is. For example, the non-cognitive mental states thought to be involved in moral judgements can be either non-endorsed first order desires, or psychologically categorical second order desires that we demand other share. In addition to there being an endorsed or non-endorsed

non-cognitive mental state, there must also be some kind of cognitive moral belief. All hybrid-expressivist positions require a high standard of endorsement and it is questionable whether a large proportion of judgements that we ordinarily consider to be moral ones can reach those standards.

Comparing psychologically categorical beliefs and desires. The moral thought pluralist can claim that cognitive moral beliefs and non-cognitive moral desires should both be considered moral judgements on the grounds that they are sufficiently endorsed. In the remainder of this chapter I will discuss some reasons to take both endorsed cognitive judgements with moral content and endorsed non-cognitive ones to be moral judgements subject to moral evaluation, criticism, praise and blame.

One may wonder whether one kind of judgement might have more authority, or more of a claim to the label of moral judgement than the other, as cognitivists and non-cognitivists each claim. Kennett and Fine have developed one method of thinking about this question by asking the following: when an individual has judgements which conflict, which of their judgements should we take to have normative authority? They argue that "... the normativity of moral judgement is most plausibly cashed out in terms of reflective endorsement and regulation ... that in cases of conflict the agents considered view deserves the title of the 'real' or authoritative moral judgement" (Kennett and Fine, 2009, pp. 87-88). This they say is the case, even when an individual fails to act on their most considered and endorsed opinion. Moral thought pluralists can agree with this claim to a large extent. They can agree when conflict occurs within a single kind of mental state domain, for example, more considered non-cognitive opinions are more representative of an individual's moral opinion than less considered non-cognitive ones, and more considered cognitive judgements are more representative of an individual's moral opinion than less considered ones. They, however, deny that there is a way of preferencing one kind (either cognitive or non-cognitive) over another when *both* are heavily endorsed. However, it may be possible to compare *across* mental state domains when the disparity of endorsement is obvious, for example, you may be able to compare a moderately assented to moral belief to a strongly endorsed non-cognitive desire. It may also be the case that the more reflective judgement is different at different times. That is, sometimes cognitive moral judgements have been more considered and endorsed and sometimes non-cognitive ones have been more endorsed. It is not clear how one can make comparisons between the two domains, given

that the processes of endorsement in each case are different. Perhaps you could compare how much conscious time an individual has spent on each; or the quality of the reflection that was engaged in, or a combination of factors.

Dual-process models. There is some confusion that seems to arise from a common misconception of the relationship between dual-process models of decision making and meta-ethical theories of moral judgement. Dual process models claim that we have one fast, automatic, intuitive method of making moral judgements as well as a slow, effortful, and conscious one. The misconception is in thinking that the fast, intuitive method is used only to reach our non-cognitive judgements and the slow, conscious one is used only when we make cognitive moral judgements. Rather, it is possible for one to make either fast or slow cognitive judgements and either fast or slow non-cognitive ones. That is, the dual-processing pathways do not correspond to the two different classes of mental states; rather it applies across mental state classes.

To illustrate the duel-process pathways that can occur within both cognitive and non-cognitive mental state domains, consider the large literature now available on *implicit attitudes*. Implicit attitudes are our unconscious attitudes and it has been reliably demonstrated that our explicit attitudes are often in contention with our implicit ones. One might be tempted to think that the implicit attitudes are non-cognitive and the explicit attitudes are cognitive. However, as described above, both implicit and explicit attitudes can be either cognitive or non-cognitive. For example, in the cognitive domain, an individual may have cognitive representations that do not discriminate well between faces of a colour dissimilar to one's own, which may result in acts of discrimination. But they may also have a cognitive belief that all races are equal and should be treated as such. In the non-cognitive domain, an individual may have a non-cognitive fear response (aversion) to people of a less familiar colour which leads them to unconsciously behave in a discriminatory manner. And, they may have another non-cognitive desire which supports the notion of equality of opportunity. As mentioned, it seems that if a person has conflicting mental states within a single domain, that their more reflective mental state is more representative of their moral position, as Kennett and Fine (2009) argue. In the example provided above the more considered views seem to be the individual's *explicit* judgements. Determining which position is an individual's more reflective one is not, though, a matter of finding out what their current explicit attitude is. It is possible that the individual's *implicit* attitude is the outcome of an unconscious appraisal

which contains a highly considered moral judgement and that their current explicit judgement is poorly considered.

The question posed by Kennett and Fine about which judgement should have normative authority is not particularly useful for comparing certain highly plausible versions of cognitivism and non-cognitivism. The problem is, both cognitive and non-cognitive judgements can be deeply considered. One can expend a huge amount of energy and resources analysing their moral beliefs, thinking about what moral properties may be, trying to identify them, looking for counter examples, and so on. Likewise, one can spend a lifetime considering and amending their values, thinking about what they want out of life, introspecting on the details of the relationships between their desires, and so on. When one considers just how much a person can reflectively consider these two vastly different kinds of mental states it is not obvious that we can say that one kind is able to be 'more considered' than the other, or that one kind 'deserves the title of the 'real' or authoritative moral judgement'. When one has beliefs and desires which seem to be in contention, it is not even clear how one could begin to compare the two. For the moral thought pluralist, these judgements are not ultimately comparable.[9] Just as a 'stool' and a 'chair' may lay equal claim to belonging to the category 'seat', certain beliefs and desires may lay equal claim to the category 'moral judgement'. Perhaps, too, non-specifically endorsed judgements also have a claim to the label, albeit an inferior one due to their dependence on pre-existing cognitive and non-cognitive moral judgements.

There is a significant consequence of restricting the concept of moral judgement by appeal to mental states that all of the meta-ethical positions suffer to a greater extent than moral thought pluralism. Judgements which fail to satisfy the standard that they set are not 'bad' moral judgements; they are not moral judgements at all. For example, traditional cognitivists and traditional non-cognitivists restrict the class to expressions of mind-to-world and world-to-mind states respectively. The other, non-prescribed kind, (if they exist) are not 'bad' moral judgements; they are not moral judgements at all. Good and bad moral judgements can be made only *within* the specified class. The empirical literature has the potential to tell us how significant dismissing an entire class of possible moral judgements might be. For example, if it turns out that people very rarely (or never) make cognitive moral judgements, then it is not very significant if cognitive judgements with moral content

[9]Although in some specific cases it may be obvious that more reflection and endorsement has occurred for one kind of judgement.

are not considered to be moral judgements at all (as opposed to being considered 'bad' moral judgements). The kind of empirical data required to test the prevalence of cognitive and non-cognitive judgements with moral content is not currently available. Essentially, with the data, one will be able to see the extent to which their theoretical commitments cohere with common intuitions about stereotypical cases. That is not to say that intuitions are right, just that bringing to light discrepancies between theoretical positions and intuitive beliefs about what moral judgements are is a good starting point for thinking about which theory, if any, should be revised or rejected.

In the following section I will relate the notion of psychologically categorical endorsement to the notion of directed action. I will explore the idea that psychologically categorical moral beliefs and desires are particularly important for their contribution to directed action, and that both moral beliefs and desires should be considered endorsed in a way that leaves them vulnerable to moral praise and blame.

6.4. Directed action

Psychologically categorical directed actions. If the Humean theory of motivation is correct, beliefs and desires explain directed action. Significantly, beliefs and desires that are strongly endorsed by an individual are important because the individual who holds them is heavily inclined to act in accordance with them when they are paired with a relevant counterpart. Essentially, one is constantly primed to act in certain ways by one's strongly endorsed beliefs and desires. Just as my belief that I cannot fly (taken to be beyond reasonable doubt) prevents me from considering some possibilities of getting from A-B, my belief that needless suffering is bad (paired with some relevant desire) prevents me from considering some methods of attaining some things that I want. My beliefs and desires determine, and give others information about, the possible ways in which interactions with me may unfold. My highly endorsed beliefs and desires should be of particular interest to others as they contribute to determining whether I am generally a harmless, dangerous, or beneficial person to be in close proximity to.

In virtue of our shared human nature much of our directed action is similar and predictable. For example, we have similar abilities to represent the world around us and a common desire for survival which leads us to approach and avoid various situations in a way that is reasonably predictable. The greatest variance in directed actions arguably arises from the divergences in our psychologically categorical beliefs and desires, which is where our moral beliefs and desires reside if moral

beliefs and non-cognitive desires are of the kind described. Psychologically categorical beliefs are, for example, beliefs about some pattern that exists among basic percepts (e.g. flamingos are pink, and 'spiders have eight legs'). Psychologically categorical *moral* beliefs make explicit reference to moral concepts (e.g. 'good' = natural property x, and 'wrong actions are those that you could not will to become a universal law'). Psychologically categorical moral desires are second order desires which we demand others share. Both kinds of psychologically categorical mental states seem to have more potential for being 'other than they were' than do basic, first order beliefs and desires. For psychologically categorical beliefs, including moral ones, there is more potential for representational error than there is for basic sensory representations. And, for desires, the scale and depth of reflection and consideration of one's desires and possible conflicts between them can alter which desires we demand that others share. Both our commitments in these domains, and the process by which we come to have those commitments is the focus of moral assessment.

Intentions: assessments of directed action. When we judge an individual's intentions, we are analysing the significance of that individual's involvement with some situation. Ordinarily, assessing an intention first involves determining whether or not some action was the product of a directed action; perhaps an individual was sleep-walking, in which case the action was not intentional. Our moral assessments of intentions are sensitive to the content of the beliefs and desires that lead to behaviour. They are also sensitive to the capacities people have to form beliefs and desires. By gathering information about these two factors we determine whether there are sufficient levels of endorsement from the agent whose actions or character is under scrutiny to warrant moral praise and blame, and we form an opinion about the moral severity of their actions.

Type and strength of endorsement. As mentioned, on the Humean view, directed action requires a belief and an appropriately related desire. If the action in question was a directed action, we want to know whether it was a product of *moral* beliefs or desires. That is, we want to know whether the beliefs and desires had moral content. Perhaps the belief was one that is not really endorsed by the actor, for example, maybe the belief is one that they do not hold very strongly or it does not contain any moral concepts; and, perhaps the desire was one that they do not endorse, but succumbed to in a moment of weakness. We are also interested in how thoroughly the individual has considered their moral beliefs and desires and possible alternatives to them. For example, one can commit themselves to a particular belief or desire after having considered (and rejected) a large or small number of

alternative possibilities.[10] Generally, we consider someone to be more morally praiseworthy or blameworthy when they act on *moral* beliefs or desires (or when they consider moral beliefs and desires and disregard them), and when the degree of commitment to their moral attitude is high (that is, when they act on strongly held psychologically categorical moral beliefs or desires).

In cases where individuals act directedly and something 'bad' happens, we tend to be interested in whether or not they foresaw the 'bad' thing happening and acted to bring it about anyway. The ignorant are generally viewed less morally culpable than the knowledgeable. People are not generally held morally responsible for actions which are the result of unforeseeable accidents unless it was somehow their responsibility to be in a position to foresee that incident. In these cases they can be judged for neglect. To be sure, being in the presence of the ignorant may have deleterious consequences (so much so that you do best to avoid them), however evil, the morally deplorable, or the abhorrent, are those who most endorse the 'bad' thing that has, or that may occur. Relatedly, we sometimes say that someone acted rightly despite their bringing about something we consider to be bad. We tend to feel this way when they could not have reasonably foreseen the consequences of their actions, or when comparable 'bad' things will occur no matter how they act.[11]

In addition to these cases, we also judge people's intentions even when their fulfilment does not, or will never occur. That is, we morally assess attitudes to, or willingness to contribute to, hypothetically 'bad' states of affairs. This can occur when we judge someone's planned but thwarted terrorist attack. Or, when one's motivation to act is overridden by some other goal; consider a would-be mass murderer whose goals are not reached due to laziness or a lack of financial resources. Or, perhaps we are critiquing beliefs and desires which have not yet formed part of an appraisal. That is, we judge in anticipation that the beliefs or desires could lead to some actions if paired with a relevant counterpart.

The final variation of these kinds of considerations I wish to point out are cases when one acts according to one's moral beliefs or desires at the time but is given the opportunity to demonstrate or articulate

[10]There are some environmental factors that come into play in relation to the type of moral beliefs and desires people have. That is, the kinds of beliefs and desires we have depends a great deal on our exposure to the moral beliefs and desires of others.

[11]In these cases, the actor generally tends to feel despair and regret at having being involved in what often is the case, harming another individual. This is an attitude apt for moral appraisal in its own right

that one no longer holds those desires or beliefs; that is, they are given the opportunity to demonstrate that they are remorseful. We want to know whether an individual who did act upon moral beliefs and desires *maintains* those beliefs and desires after the act. The answers to all of these questions that we are interested in give us information about the likelihood that some directed action will occur again.

One noteworthy complication when thinking about endorsement capacities arises because many moral situations call for the judge to make multiple related moral judgements. For example, it appears that even in the very ordinary case of someone stepping on your toe, there are actually two 'moral' judgements being made which ought to be distinguished. First there is the judgement that something 'bad' (or 'good') has happened (e.g. someone's toe has been hurt and that is 'bad'), and secondly, there is the judgement that some individual is morally culpable for contributing to that 'bad' thing. In other words, there is a judgement about states of affairs and about causes of states of affairs. Sometimes moral discussion centres on one and not the other, for example, we discus the moral status of states of affairs when we discuss whether climate change, genetic engineering, animal suffering, or abortion are really 'bad' without discussing any individual's contribution to climate change, genetic engineering, animal suffering, or abortion. Other times, that something 'bad' has happened is not in dispute. What is in focus instead is whether or not some individual who has causally contributed to the 'bad' things occurring is morally blameworthy. Appraisals in all of these situations require either cognitive or non-cognitive (or both) moral judgements and a failure to separate one judgement (say of the 'bad' that has occurred) from another (the judgement of an individual's intent) may lead to one judgement influencing the other in ways that we may not approve of.[12]

As I will discuss in the following section, when we judge people for their directed actions, we tend to also be interested in the extent to which the person being judged is capable of forming either cognitive or non-cognitive moral views about the 'goodness' or 'badness' of states of affairs, or of their ability to endorse and adhere to their desires.

Capacities for endorsement. Being capable of making moral judgements is an ability often thought to be uniquely human. Although we generally do take each other to be moral agents, we do not think that all human beings are morally competent. There are a host of

[12]There are people who will claim that when it comes to judging others morally, there is no need to consider the actor's intentions. Conversely, there are others who will claim that intention is the only morally relevant variable. Most people intuitively take both sources of information to be morally significant.

seemingly unrelated mental capacities that affect our tendency to be considered a moral agent. These capacities all affect or determine our ability to endorse mental states either cognitively or non-cognitively. Consider the ability to recognise conflicts in desires and the ability to abstractly formulate a belief such as 'pain with no positive consequence is bad whenever it occurs'. If psychologically categorical endorsement, whether cognitive or non-cognitive, is necessary for moral judgements, then any capacity which impacts on how capable we are of endorsing in this way will influence the depth of moral judgement that we are capable of making. We tend to, in part, be morally culpable insofar as we have these capacities. Our grasp of the capacities which contribute to whether someone is capable and the extent to which they are capable is not clearly defined and explicit.

The standard seems, at least in part, to be set by what is average for human beings. That is, there are some capacities which, when a person possesses them at a level which is below average (or not at all) either eliminates or reduces the extent to which we consider them moral agents. For example, our capacities to understand other minds or to foresee consequences could be, on average, better or worse than they actually are. Were they better, we would expect more from people morally and were they worse, we would expect less. The average power of our capacities is just a matter of happen-stance and could be otherwise in ways that would alter our moral expectations of people. For example, liken the average ability of people to perceive conflicts in their own desires over time or to look for counter examples to their moral beliefs, to the average ability of people to do basic maths or to see at a distance. Stereotypical impairments of capacities that affect *all* kinds of directed action may include things like blindness, mental disabilities or psychological disorders such as depression, psychopathy, or anxiety. Certain mind altering substances such as drugs and alcohol, as well as other more everyday factors such as our general arousal levels, hunger levels, moods, and so on are all believed to have an impact on our ability to make decisions, including moral decisions. Defects in decision making capacities can be present from birth, acquired, temporary, or permanent. Children's ability to make decisions (including moral decisions) is commonly thought to develop over time and they are generally judged to an extent suited to their abilities and with recognition of their potential for improvement.[13]

[13]Historically, many of these factors have not been taken into consideration in moral assessments, for example, children have been treated as adults. On this, I will merely say that we regard our views on this to have progressed rather than simply changed, for example, we tend to think that there is some reason for our treating

Some mental capacities affecting *cognitive moral judgements* will include an individual's capacity to represent or form second order mind-to-world beliefs. This includes the number of possible beliefs about an object that they were capable of considering before they were drawn to a particular view, their capacity to think abstractly, to manipulate concepts, to recognise when representations may be inaccurate, and to actively seek data that falsifies their representations. All of these abilities can break down in many ways. Schizophrenia is an extreme example of an impairment in forming ordinary cognitive beliefs. An impairment in a *moral* representation may occur when someone believes that some natural property is morally bad, but fails to recognise instances where they take it not to be bad in specific cases. Many of us have a tendency to adjust our moral condemnation of people based on our understanding of their capacities for cognitive representation.

For *non-cognitive* moral judgements, we are interested in individuals' capacities to prioritise desires and to demand that others share desires. Among other abilities, this depends upon an individuals ability to recognise their own desires, to recognise the consequences of the fulfilment of their desires, and to recognise when the fulfilment of one of their desires conflicts with the fulfilment of another. It also depends upon their capacity to have a preference for one desire over another when such conflicts arise, as well as their capacity to recall and adhere to their preferred desire. Again, adjustments are made to attributions of wrong-doing based on the degree to which we take an individual to have these capacities.

Directed actions and moral thought pluralism. There is a further point to be made in favour of moral thought pluralism which highlights the link between directed action and the *explanatory component of interest*. According to non-cognitivism, moral content lies in the non-cognitive component of an appraisal. In contrast, for the cognitivist, moral content lies in the cognitive content of an appraisal. One reason to prefer moral thought pluralism is that when we attempt to explain an individual's morally motivated actions, the explanatory component of interest need not be confined to one side. For example, when explaining the actions of a suicide bomber, the moral belief that to act rightly one must follow the word of God, as opposed to the desire, for example to act rightly, is what gives us the most insight into the actor's behaviour. In contrast, the *desire* for a 'pure' race. as opposed to the *belief* that

children differently from adults when we appraise their behaviour. Similarly, there seems to be more of a focus on rehabilitation over incarceration when it comes to dealing with criminals. Likewise, medical abnormalities are taken into consideration more readily.

some race is inferior, is more explanatorily informative in explaining why someone might go on to murder members of a particular race. After all, the belief that a race is inferior is no reason to murder.

6.5. Conclusion

This chapter has explored the idea that both moral beliefs and desires involve a level of commitment from the agent who has them that warrants moral praise and blame. I have argued that when an individual has highly endorsed moral beliefs and desires, it is not obvious that one kind is better deserving of the label 'moral judgement'. Both psychologically endorsed moral beliefs and desires are central contributors to our directed actions and in morally assessing directed actions it seems that sometimes the explanatory component of interest is located in an individual's beliefs and other times in their desires.

CHAPTER 7

Conclusion

Traditionally there has been an assumption that moral judgements express either beliefs or desires. Mental states are widely accepted to have just one direction-of-fit, such that only our non-cognitive desires are inherently motivating and only our cognitive beliefs are truth-apt. As a result, the competing views of cognitivism and non-cognitivism have unique advantages and disadvantages. Specifically, cognitivists are easily able to explain the intuition that moral judgements are truth-apt, while non-cognitivists can easily explain the intuition that moral judgements are motivating. As discussed, cognitivists and non-cognitivists have developed some quite sophisticated methods of incorporating the inherent advantages of their rival's position into their own theories. But because of the inherent nature of mental states, attempts to do so regularly introduce new problems. For cognitivists there are concerns about how moral judgements and moral motivation can be connected, whereas for non-cognitivists there are concerns about moral relativism and standards of moral truth. Defining moral judgements in terms of either beliefs or desires has remained dissatisfying to many.

In this thesis I have outlined three kinds of hybrid theories that have been developed as a result of efforts to capture the positive qualities of both cognitivism and non-cognitivism: hybrid-state theory, hybrid-expressivism, and moral thought pluralism. These theories differ in their abilities to explain moral truth and moral motivation. For example, hybrid-state theorists can easily explain motivation, but they face similar challenges as non-cognitivists in explaining how besires can be truth-apt. Hybrid-expressivists can also explain motivation, while their ability to explain moral truth depends upon the nature of the moral beliefs that they take to be involved. For moral thought pluralists, moral beliefs are truth-apt (but not motivating) and moral desires are motivating (but not truth-apt).

In recent years, experimental psychology has weighed in on the traditionally philosophical debate about the nature of moral judgements. The main goal of the empirical research in this area has been to discover what kind of mental content is involved in judgements that we

ordinarily consider to be moral ones. Three broad pursuits are involved: correlational observations, experimental studies, and descriptive studies. I have found that the latter has proven most useful for contributing to discussions about moral judgements. A general finding of correlational studies is that moral judgements co-occur with emotions and if appraisal theories of emotions are correct, then these emotions involve a non-cognitive mental state of some kind. Although these findings have been claimed to discredit cognitivism, all meta-ethical theories of moral judgement are able to explain the correlational data. Experimental studies find that we in fact have a tendency to use feelings as a source of information when making stereotypical moral judgements. Again, these findings have been wrongly taken to discredit cognitive theories of moral judgement; cognitivists are able to explain why people might be inclined to think that their feelings are a reliable source of information about moral matters, although for cognitvists these judgements are not moral judgements. Descriptive studies are able to delve deeper into the nature of the mental content that is involved in moral judgements. These studies question the need for explicit moral beliefs and complex moral desires of the kinds described by cognitivists and non-cognitivists. They are not useful for comparing cognitivism and non-cognitivism, because moral dumb-founding studies seem to show that both moral desires and moral beliefs are missing when people make moral judgements. However, they are useful for comparing hybrid theories of moral judgement.

Cognitivists do not take just any beliefs to be moral judgements, just as non-cognitivists do not take all desires to be moral judgements. For cognitivists, moral beliefs involve the explicit use of moral concepts such as 'good', 'bad', 'right', and 'wrong', and for non-cognitivists, I have taken moral judgements to be the expression of a world-to-mind mental state of a kind such as that advocated by Blackburn (i.e. a desire that we demand others share). There are other ways that moral non-cognitive mental states can be distinguished from non-moral ones. The commitments one makes in this respect are important for those who wish to defend moral thought pluralism, as a defence of this position requires identifying similarities between different kinds of mental states.

For the hybrid-state theorist, moral judgements do not consist of mental states of the kinds described by moral cognitivists or non-cognitivists. The existence of hybrid mental states is controversial. And, even if they do exist, it is a further question whether they ought to be considered *moral* judgements. At face value, moral dumb-founding studies seem to support hybrid-state theory, as it appears that people make moral judgements in the absence of any complex moral beliefs or desires.

However, there are other ways of explaining what is going on in moral dumb-founding cases, which, in contrast to cognitivism, non-cognitivism, and hybrid-expressivism, do not require that the judgements consist of complex moral beliefs or desires. Hybrid mental states such as besires, if they exist, do not require any *psychologically categorical* endorsement, and for this reason may be too simple to contain moral content.

For hybrid-expressivists, moral judgements occur when we cognitively endorse a non-cognitive mental state of some kind. A central problem for hybrid-expressivists is to specify the relationship between cognitive and non-cognitive mental states. For the hybrid-expressivist, at least one, and potentially both kinds of psychologically categorical endorsement are required. Specifically, cognitive endorsement is required in the form of a belief with moral content. Non-cognitive endorsement may or may not be required, depending on whether the hybrid-expressivist claims that the non-cognitive mental states are of first or second order. Either way, hybrid-expressivism sets a high standard for a judgement to be considered a moral one and it is an interesting empirical question whether judgements that we generally want to label as moral can reach it.

Moral thought pluralists claim that moral judgements can express more than one kind of moral thought. I have explored the idea that they can express either moral beliefs or desires. There must be some feature which unites the differing kinds of moral judgements. The specific ways of distinguishing moral beliefs from non-moral beliefs and moral desires from non-moral desires can differ from what I have discussed in this thesis. But, once one identifies restrictions within the cognitive and non-cognitive classes, discussions about what those judgements have in common may ensue. I have suggested that there is a specific kind of endorsement that is capable of uniting some cognitive and non-cognitive moral judgements as those that are worthy of moral praise and blame.

Our beliefs and desires contribute to our directed actions. I have mentioned that there is likely greater variance in beliefs and desires between individuals as they become more abstract. For example, there is likely to be greater agreement among individuals about basic representational beliefs (such as the layout of medium sized objects in a shared environment) than about abstract beliefs (such as what natural properties things which we consider 'bad' share). Similarly, there is likely to be less variance in individuals' desires for what are considered basic needs (such as water, food, and shelter) than there is variance in second order desires (such as how to spend one's leisure time). In other words, people have some first order beliefs and desires which are generally similar and predictable. These similar beliefs and desires generate

predictable directed actions. Variance in directed actions arises more from the divergences in our second order (psychologically categorical) beliefs and desires. These psychologically categorical beliefs and desires have a greater potential for being 'other than they were' than do basic beliefs and desires. There is more potential for making errors when it comes to forming abstract beliefs than there is when forming basic sensory representations. And, the scale and depth of reflection and consideration of our first order desires, as well as the possible conflicts between them, can alter which second order desires we develop and demand that others share. The commitments made in both of these cases are morally appraisable.

Support for moral thought pluralism also comes from the observation that the explanatory component of interest when it comes to directed actions need not be confined to either our beliefs or our desires. That is, sometimes our beliefs are more informative for explaining our moral actions and sometimes our desires are. Further, it is not clear that one kind (either cognitive or non-cognitive endorsed mental states) can be claimed to be more considered or endorsed than the other. For moral thought pluralists, endorsed moral beliefs and desires both have the potential to reach sufficient levels of commitment from the agent making them to warrant moral praise and blame. Moral thought pluralism also gives us a way of explaining what is going on in moral dumb-founding cases without appealing to hybrid mental states nor claiming that there are specific moral beliefs or desires at work. They can claim that such a judgement without an explicitly identifiable belief or desire can still be considered a moral judgement, though this kind of moral judgement may be considered inferior as its existence depends on a pre-existing body of moral judgements.

All attempts to restrict the definition of moral judgements to one category or another, whether it be cognitivism, non-cognitivism, hybrid-state theory, hybrid-expressiivism, or some other position, must explain why one kind better deservers the label. If one is not persuaded by current arguments for these restrictive theories, one should hold a presumption in favour of moral thought pluralism until such time as a more restrictive theory is found satisfactory.

Bibliography

Aharoni, E., W. Sinnott-Armstrong and K. A. Kiehl (2012). 'Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction'. *Journal of Abnormal Psychology* 121.2, pp. 484–497.

Altham, J. E. J. (1986). 'The legacy of emotivism'. In: *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic*. Basil Blackwell, pp. 275–88.

Anscombe, G. E. M. (1957). *Intention*. Harvard University Press: London.

Asch, S. E. (1951). 'Effects of group pressure upon the modification and distortion of judgement'. In: *Groups, Leadership and Men*. Carnegle Press.

Ayer, A. J. ((1936) 1971). *Language, Truth and Logic*. Penguin: Harmondsworth.

Bargh, J. A. and T. L. Chartrand (1999). 'The unbearable automaticity of being'. *American Psychologist* 54.7, pp. 462–479.

Björnsson, G., F. Björklund, C. Strandberg, J. Eriksson and R Francën Olinder (2015a). 'Motivational internalism and folk intuitions'. *Philosophical Psychology* 28, pp. 715–734.

Björnsson, G., F. Björklund, C. Strandberg, J. Eriksson and R. Francën Olinder (2015b). *Motivational Internalism*. Oxford University Press.

Björnsson, G. and R. Francën Olinder (2012). 'Internalists beware - we might all be amorlaists!' *Australasian Journal of Philosophy* 91, pp. 1–14.

Blackburn, S. (1993). *Essays in Quasi-realism*. Oxford University Press: Cambridge.

— (1998). *Ruling Passions*. Clarendon Press: Oxford.

— (2002). 'Précis of ruling passions'. *Philosophy and Phenomenological Research* 65.1, pp. 122–135.

Blair, R. J. R. (1995). 'A cognitive developmental approach to morality: investigating the psychopath'. *Elseveir Science, Cognition* 57, pp. 1–29.

Cameron, D. C., K. A. Lindquist and K. Gray (2015). 'A constructionist review of morality and emotions: no evidence for specific links between moral content and discrete emotions'. *Personality and Social Psychology Review* 19.4, pp. 371–394.

Cannon, P. R., S. Schnall and M. White (2011). 'Transgressions and expressions: affective facial muscle activity predicts moral judgments'. *Social Psychological and Personality Science* 2.3, pp. 325–331.

Case, T. I., M. J. Oaten and R. J. Stevenson (2012). 'Disgust and moral judgment'. In: *Emotions, Imagination, and Moral Reasoning*. Ed. by R. Langdon and C. Mackenzie. Psychology Press: New York, pp. 195–218.

"Chair" (2018). In: *Merriam-Webster*. Merriam-Webster incorperated. URL: `https://www.merriam-webster.com/`.

Chapman, H. A. and A. K. Anderson (2013). 'Things rank and gross in nature: a review and synthesis of moral disgust'. *Psychological Bulletin* 139.2, pp. 300–327.

Chapman, H. A., D. A. Kim, J. M. Susskind and A. K. Anderson (2009). 'In bad taste: evidence for the oral origins of moral disgust'. *Science* 323, pp. 1222–1226.

Cholbi, M. (2006). 'Moral belief attribution: a reply to Roskies'. *Philosophical Psychology* 19.5, pp. 629–638.

Cima, M., F. Tonnaer and M. D. Hauser (2010). 'Psychopaths know right from wrong but don't care'. *Social Cognitive and Affective Neuroscience* 5, pp. 59–67.

Cohon, R. (2010). 'Hume's moral philosophy'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2010 Edition. URL: `http://plato.stanford.edu/archives/fall2010/entries/hume-moral/`.

Curtis, V. and A. Biran (2001). 'Dirt, disgust, and disease: is hygiene in our genes?' *Perspectives in Biology and Medicine* 44.1, pp. 17–31.

Damasio, A. (1994). *Descartes' Error*. G. P. Putnam's Sons: New York.

Danovitch, J. and P. Bloom (2009). 'Children's extension of disgust to physical and moral events'. *American Psychological Association: Emotion* 9.1, pp. 107–112.

Dolan, M. C. and R. S. Fullam (2010). 'Moral/conventional transgression distinction and psychopathy in conduct disordered adolescent offenders'. *Personality and Individual Differences* 49.8, pp. 995–1000.

Dreier, J. (2004). 'Meta-ethics and the problem of creeping minimalism'. *Philosophical Perspectives* 18.1, pp. 23–24.

Eggers, D. (2015). 'Unconditional motivational internalism and Hume's lesson'. In: *Motivational Internalism*. Oxford University Press.

Ekman, P. and W. V. Friesen (1975). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Prentice Hall: Michigan.

— (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Prentice Hall: Michigan.

Ekman, P and W. V. Friesen (1989). 'The argument and evidence about universals in facial expressions of emotion'. In: *Handbook of Social Psychophysiology*. John Wiley and Sons, Ltd.: New York.

Eskine, K. J., N. A. Kacinik and J. J. Prinz (2011). 'A bad taste in the mouth: gustatory disgust influences moral judgment'. *Association for Psychological Science* 22.3, pp. 295–299.

Ferguson, H. J. and A. J. Sanford (2008). 'Anomalies in real and counterfactual worlds: an eye-movement investigation'. *Journal of Memory and Language* 58.3, pp. 609–626.

Ferguson, H. J., C. Sheepers and A. J. Sanford (2010). 'Expectations in counterfactual and theory of mind reasoning'. *Language and Cognitive Processes* 25.3, pp. 297–346.

Fine, C. (2006). 'Is the emotional dog wagging its rational tail, or chasing it?' *Philosophical Explorations* 9.1, pp. 83–98.

Finlay, S. (2006). 'The conversational practicality of value judgement'. *Journal off Ethics* 8.3, pp. 205–223.

Fisher, A. (2011). *Metaethics: An Introduction*. Acumen: Durham.

Frankfurt, H. G. (1971). 'Freedom of the will and the concept of a person'. *The Journal of Philosophy* 68.1, pp. 5–20.

Galotti, K. M. (1989). 'Approaches to studying formal and everyday reasoning'. *Psychological Bulletin* 105, pp. 331–351.

Gendler, T. S. (2008). 'Alief and belief'. *Journal of Philosophy* 105.10, pp. 634–663.

Gibbard, A. (1990). *Wise Choices, Apt Feelings: a Theory of Normative Judgment*. Harvard University Press: Cambridge.

Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley and J. D. Cohen (2001). 'An fMRI investigation of emotional engagement and moral judgment'. *Science* 239.5537, pp. 2105–2108.

Greene, J. and J. Haidt (2002). 'How (and where) does moral judgment work?' *Trends in Cognitive Science* 6.12.

Gupta, Anil (2015). 'Definitions'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2015. Metaphysics Research Lab, Stanford University. URL: `http://plato.stanford.edu/archives/sum2015/entries/definitions/`.

Gutierrez, R., R. Giner-Sorolla and M. Vasiljevic (2012). 'Just an anger synonym? Moral context influences predictors of disgust word use'. *Cognition and Emotion* 26.1, pp. 53–64.

Haidt, J. (2001). 'The emotional dog and its rational tail: a social intuitionists approach to moral judgement'. *Philosophical Reviews* 108.4, pp. 814–834.

— (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. Pantheon Books: New York.

Haidt, J. and F. Bjorklund (2007). 'Social intuitionists answer six questions about moral psychology'. In: *Moral Psychology: Volume 2*. Ed. by W. Sinnott-Armstrong. MIT Press: Cambridge.

Haidt, J., F. Bjorklund and S. Murphy (2000). 'Moral dumbfounding: when intuition finds no reason'. *Lund Psychological Reports* 1.2.

Haidt, J. and J. Graham (2007). 'When morality opposes justice: conservatives have moral intuitions that liberals may not recognize'. *Social Justice Research* 20, pp. 98–116.

Haidt, J., S. Koller and M. Dias (1993). 'Affect, culture, and morality, or is is wrong to eat your dog?' *Journal of Personaliy and Social Psychology* 65.4, pp. 613–628.

Haidt, J., P. Rozin, C. McCauley and S. Imada (1997). 'Body, psyche, and culture: the relationship between disgust and morality'. *Psychology and Developing Societies* 9.1, pp. 107–131.

Hanna, Robert (2017). 'Kant's theory of judgment'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2017. Metaphysics Research Lab, Stanford University.

Horberg, E. J., C. Ovies, D. Keltner and A. B. Cohen (2009). 'Disgust and the moralization of purity'. *Journal of Personality and Social Psychology* 97, pp. 963–976.

Hume, D. (1738–40). *A Treatise of Human Nature*.

Hutcherson, C. A. and J. J. Gross (2011). 'The moral emotions: a social-functionalist account of anger, disgust, and contempt'. *Journal of Personality and Social Psychology* 100, pp. 719–725.

Jacob, P. (2014). 'Intentionality'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Winter 2014 Edition. URL: `http://plato.stanford.edu/archives/win2014/entries/intentionality/`.

James, W. (1884). 'What is an emotion?' *Mind* 9.34, pp. 188–205.

Jollimore, T. (2017). 'Impartiality'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Winter 2017 Edition. URL: `https://plato.stanford.edu/archives/win2017/entries/impartiality/`.

Jones, A. and J. Fitness (2008). 'Moral hypervigilance: the influence of disgust sensitivity in the moral domain'. *Emotion* 8, pp. 613–627.

Joyce, R. (2015). 'Moral anti-realism'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2015 Edition. URL: `http:`

//plato.stanford.edu/archives/fall2015/entries/moral-anti-realism/.

Kant, I (1950). *Foundations of the Metaphisics of Morals*. 1st ed. University of Chicago Press: Chicago.

Kauppinen, A. (2008). 'Moral internalism and the brain'. *Social Theory and Practice* 34, pp. 1–24.

— (2013). 'Ethics and empirical psychology'. In: *Empirically Informed Ethics*. Ed. by M. Christen, C. van Scaik, J. Fischer, M. Huppenbauer and C. Tanner. Springer: New York, pp. 279–305.

— (2015). 'Intuition and belief in moral motivation'. In: *Motivational Internalism*. Ed. by G. Björnsson et al. Oxford University Press.

— (2016). 'Moral sentimentalism'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2016 Edition. URL: http://plato.stanford.edu/archives/fall2016/entries/moral-sentimentalism/.

Kennett, J. and C. Fine (2008). 'Internalism and the evidence from psychopathy and acquired sociopaths'. In: *Moral Psychology: Volume 3*. MIT Press: Cambridge.

— (2009). 'Will the real moral judgment please stand up? The implications of social intuitionist models of cognition for meta-ethics and moral psychology'. *Ethical Theory and Moral Practice* 12.

Kirchin, S. (2012). *Meta-ethics*. Palgrave Macmillan: UK.

Kriegel, U. (2012). 'Moral motivation, moral phenomenology, and the alief/belief distinction'. *Australian Journal of Philosophy* 90.3, pp. 469–486.

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press: Massachusetts.

Kuhn, D. (1991). *The Skills of Argument*. Cambridge University Press: New York.

Kunda, Z. (1990). 'The case for motivated reasoning'. *Psychological Bulletin* 108.3, pp. 480–498.

Landy, J. F. and G. P. Goodwin (2015). 'Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence'. *Association for Psychological Science* 10.4, pp. 518–536.

Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford University Press: Oxford.

Levy, N. (2006). 'The wisdom of the pack'. *Philosophical Explorations* 9.1.

Lewis, D. (1972). 'Psychophysical and theoretical identifications'. *Australasian Journal Philosophy* 50, pp. 249–258.

Locke, J. ((1690) 1975). *An Essay Concerning Human Understanding*. Ed. by P. Nidditch. Clarendon Press: Oxford.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. Penguin books: New York.

Margolis, Eric and Stephen Laurence (2014). 'Concepts'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2014. Metaphysics Research Lab, Stanford University.

Marzillier, S. and G. Davey (2004). 'The emotional profiling of disgust-eliciting stimuli: evidence for primary and complex disgusts'. *Cognition and Emotion* 18.3, pp. 313–336.

McDowell, J. (1998). *Mind, Value, and Reality*. Harvard University Press: Cambridge.

McLaughlin, B. and K. Bennett (2014). 'Supervenience'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Spring 2014 Edition. URL: `http://plato.stanford.edu/archives/spr2014/entries/supervenience/`.

Miller, A. (2003). *An Introduction to Contemporary Meta-Ethics*. Wiley.

Moll, J., R. de Oliveirra-Souza and P. J. Eslinger (2003). 'Morals and the human brain: a working model'. *Neuroreport* 14, pp. 299–305.

Moore, G. E. (1903). *Principia Ethica*. Cambridge University Press.

Moretti, L. and G. di Pellegrino (2010). 'Disgust selectively modulates reciprocal fairness in economic interactions'. *Americal Psychological Association: Emotion* 10.2, pp. 169–180.

Nabi, R. L. (2002). 'The theoretical versus the lay meaning of disgust: implications for emotion research'. *Cognition and Emotion* 16, pp. 695–703.

Nichols, S. (2002). 'On the geneaology of norms: a case for the role of emotion in cultural evolution'. *Philosophy of Science* 69, pp. 234–355.

— (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgement*. Oxford University Press: Oxford.

Oaten, M. J., R. J. Stevenson and T. I. Case (2009). 'Disgust as a disease-avoidance mechanism'. *Psychological Bulletin* 135.2, pp. 303–321.

Perkins, D. N., R. Allen and J. Hafner (1983). 'Difficulties in everyday reasoning'. In: *Thinking: The Expanding Frontier*. Ed. by W. Maxwell. The Franklyn Institute Press: Pennsylvania.

Perkins, D. N., M. Farady and B. Bushey (1991). 'Everyday reasoning and the roots of intelligence'. In: *Informal Reasoning and Education*. Ed. by J. F. Voss, D. N. Perkins and J. W. Segal. L. Erlbaum Associates: Michigan.

Pizarro, D., Y. Inbar and C. Helion (2011). 'On disgust and moral judgment'. *Emotion Review* 3.3, pp. 267–268.

Platts, M. (1979). *Ways of Meaning*. Routledge and Kegan Paul.

Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotions*. Oxford University Press: New York.

— (2008). *The Emotional Construction of Morals*. Oxford University Press: Clarendon.

— (2015). 'An empirical case for motivational internalism'. In: *Motivational Internalism*. Ed. by G. Björnsson et al. Oxford University Press.

— (2016). 'Sentimentalism and the moral brain'. In: *Moral Brains: The Neuroscience of Morality*. Ed. by S. Matthew Liao. Oxford University Press.

Radcliffe, E. S. (2006). 'Moral internalism and moral cognitivism in Hume's metaethics'. *Synthese* 152.3, pp. 353–370.

Reichlin, M. (2014). 'Neuroethics and the rationalism/sentimentalism divide'. In: *Morality in Times of Naturalizing the Mind*. Ed. by C. Lumer. Walter de Gruyter Inc.: Boston.

Roojen, M. van (2015). *Meta-ethics: A Contemporory Introduction*. Routlidge: London.

— (2016). 'Moral cognitivism vs. non-cognitivism'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Winter 2016 Edition. URL: `http://plato.stanford.edu/archives/fall2015/entries/moral-cognitivism/`.

Rosati, C. S. (2016). 'Moral motivation'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2016 Edition. URL: `http://plato.stanford.edu/archives/fall2016/entries/moral-motivation/`.

Rosch, E. H. (1973). 'Natural categories'. *Cognitive Psychology* 4, pp. 328–350.

Roskies, A. (2003). 'Are ethical judgments intrinsically motivational? Lessons from acquired sociopathy'. *Philosophical Psychology* 16.1, pp. 51–66.

Ross, L., M. R. Lepper and M. Hubbard (1975). 'Perserverence in self-perception and social perception: biased attributional processes in the debriefing paradigm'. *Journal of Personality and Social Psychology* 35.5, pp. 880–892.

Royzman, E. B. and J. Sabini (2001). 'Something it takes to be an emotion: the interesting case of disgust'. *Journal for the Theory of Social Behaviour* 31.1, pp. 29–59.

Rozin, P., J. Haidt and K. Fincher (2009). 'From oral to moral: is moral disgust an elaboration of a food rejection system'. *Science* 323, pp. 1179–1180.

Rozin, P., J. Haidt and C. McCauley (1999). 'Disgust: the body and soul emotion'. In: *Handbook of Cognition and Emotion*. Ed. by T. Dalgeish and M. Power.

— (2009). 'Disgust: the body and soul emotion in the 21st century'. In: *Disgustand its Disorders: Theory, Assessment, and Treatment Implications*. Ed. by B. O. Olatunji and D. Mckay. American Psychological Association: Washington.

Rozin, P., L. Lowery, J. Haidt and S. Imada (1999). 'The CAD triad hypothesis: a mapping between three moral emotions (Contempt, Anger, Disgust) and three moral codes (Community, Autonomy, Divinity)'. *Journal of Personality and Social Psychology* 76.4, pp. 574–586.

Rozyman, E. B., K. Kim and R. F. Leeman (2015). 'The curious tale of Julie and Mark: unraveling the moral dumbfounding effect'. *Judgment and Decision Making* 10.4, pp. 296–313.

Rozyman, E. B. and R. Kurzban (2011a). 'Facial movements are not goosebumps: A response to Chapman and Anderson'. *Emotion Review* 3, pp. 274–275.

— (2011b). 'Minding the metaphor: the elusive character of moral disgust'. *Emotion Review* 3, pp. 269–271.

Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom and J. D. Cohen (2003). 'The neural basis of economic decision making in the ultimatum game'. *Science* 300.5626, p. 1755.

Saunders, L. F. (2015). 'What is moral reasoning?' *Philosophical Psychology* 28.1, pp. 1–20.

Sayre-McCord, G. (1988). *Essays on Moral Realism*. Cornell University Press: Michigan.

— (2014). 'Metaethics'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Summer 2014 Edition. URL: `http://plato.stanford.edu/archives/sum2014/entries/metaethics/`.

Schroeder, T. (2015). 'Desire'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Summer 2015 Edition. URL: `http://plato.stanford.edu/archives/sum2015/entries/desire/`.

Schwartz, N. and G. L. Clore (1983). 'Mood, misattribution, and judgments of well-being: informative and directive functions of affective states'. *Journal of Personality and Social Psychology* 45.3, pp. 513–523.

"Seat" (2018). In: *Merriam-Webster*. Merriam-Webster incorperated. URL: `https://www.merriam-webster.com/`.

Seidel, A. and J. Prinz (2013). 'Sound morality: irritating and icky noises amplify judgments in divergent moral domains'. *Cognition* 127, pp. 1–5.

Simpson, J., S. Carter, S. Anthony and P. Overtson (2006). 'Is disgust a homogenous emotion?' *Motivation and Emotion* 30.1, pp. 31–41.

Smith, B. and C. Korsmeyer (2004). 'Visceral values: Aurel Kolnai on disgust'. In: *On Disgust*. Carus publishing company, pp. 1–22.

Smith, M. (1994). *The Moral Problem*. Blackwell Publishing: USA.

— (2015). 'Evaluative judgements, judgements about reasons and motivations'. In: *Motivational Internalism*. Ed. by G. Björnsson et al. Oxford University Press.

Sousa, R. de (2017). 'Emotion'. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Winter 2017. Metaphysics Research Lab, Stanford University. URL: `https://plato.stanford.edu/archives/win2017/entries/emotion/`.

Stavrova, O., G. E. Newman, A. Kulemann and D. Fetchenhauer (2016). 'Contamination without contact: an examination of intention-based contagion'. *Judgment and Decision Making* 11.6, pp. 554–571.

Steiner, J. (1973). 'The gustofacial response: observation on normal and anencephalic newborn infants'. In: *Oral Sensation and Perception: Development in the Fetus and Infant: Fourth Symposium*. Ed. by J. F. Bosma. Us Govt. printing office.

Stevenson, C. L. (1937). 'The emotive meaning of ethical terms'. *Mind* 46.181, pp. 14–31.

— (1944). *Ethics and Language*. Yale University Press: New Haven.

— (1963). *Facts and Values*. Yale University Press: New Haven.

Strandberg, C. and F. Björklund (2012). 'Is moral internalism supported by folk intuitions?' *Philosophical Psychology* 26, pp. 319–335.

Street, S. (2006). 'A Darwinian dilemma for realist theories of value'. *Philosophical Studies* 127.1, pp. 109–166.

Susskind, J. M., D. H. Lee, A. Cusi, R. Fieman, W. Grabski and A. K. Anderson (2008). 'Expressing fear enhances sensory aquisition'. *Nature Neuroscience* 11, pp. 843–850.

Wheatley, T. and J. Haidt (2005). 'Hypnotic disgust makes moral judgments more severe'. *Psychological Science* 16.10, pp. 780–784.

Whitton, A. E., J. D. Henry, P. J. Rendell and J. R. Grisham (2013). 'Disgust, but not anger provocation, enhances levator labii superioris activity during exposure to moral transgressions'. *Biological Psychology* 96, pp. 48–56.

Wilson, T. D. (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Harvard University Press: Massachusetts.

Wilson, T. and R. Nisbett (1978). 'The accuracy of verbal reports about the effects of stimuli on evaluations and behavior'. *Social Psychology* 41.2, pp. 118–131.

Wittgenstein, L. (1953). *Philosophical Investigations*. Trans. by G. E. Anscombe. 1st ed. Macmillan Publishing Company.

Wood, E. (2014). 'Björnsson and Olinder on motivational internalism'. *Australasian Journal of Philosophy* 92, pp. 379–383.

Zabzebski, L. (2003). 'Emotion and moral judgment'. *Philosophy and Phemonenological Research* 66.1, pp. 104–124.

Zangwill, N. (2015). 'Motivational externalism: formulation, methodology, rationality, and indifference'. In: *Motivational Internalism*. Ed. by G. Björnsson et al. Oxford University Press.

Zhong, C. and K. Liljenquist (2006). 'Washing away your sins: threatened morality and physical cleansing'. *Science* 313.5792, pp. 1451–1452.