MORAL JUDGMENT AND MOTIVATION

by

XIAO ZHANG

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Philosophy
School of Philosophy,
Theology and Religion
College of Arts and Law
University of Birmingham
March 2020

Abstract

In this thesis, I explore motivational internalism and externalism, which concern the relationship between moral judgments and motivation. I first introduce the basic terms and different forms of internalism and externalism, including the externalist objections to internalism based on the famous counterexamples. I then argue against externalism by defending and developing Michael Smith's fetishism argument. I not only respond to the externalist objections to the fetishism argument but also further argue against different externalist explanations of moral motivation that intend to avoid the fetishism charge. Finally, I re-examine different forms of internalism in order to argue for a new form of internalism that can better preserve our internalist intuitions whilst accommodating the externalist counterexamples. My ultimate conclusion will be that the most plausible form of internalism to accept is constitutional, unconditional, relatively strong, direct internalism that is formulated in terms of dispositional desires.

Acknowledgement

First and foremost, my sincere gratitude goes to Jussi Suikkanen for being an excellent supervisor. I could have hardly had better philosophical support than him for the whole doctoral experience. Jussi not only helped me out with developing my rudimentary ideas into well-structured arguments but rather he also helped me with presenting the arguments precisely. This invaluable experience reshapes my way of thinking of philosophical issues and dealing with challenges in philosophy. A massive thank you also goes to Maja Spener whose feedback prompted me to reconsider some crucial debates involved in this thesis from different perspectives. I would also like to thank my parents for their support through this whole PhD journey. The Department of Philosophy at University of Birmingham has been a friendly and vibrant place to conduct research, where I have received a lot of insightful thoughts from seminars and conferences. It is my pleasure to have once been a member of this fantastic philosophical community. Finally, this thesis was made possible through the generous funding from the China Scholarship Council, for which I am deeply thankful.

Table of Contents

Chapter 1: Introduction

One thing we can observe in many situations is that, if someone changes her mind about a moral issue (i.e., makes a new moral judgment), normally her motivation will change accordingly. For example, if you come to realize that a party advocates exactly your views and represents your interests, then you should have some motivation to vote for this party in the election. In philosophical moral psychology, the view that accepts the previous connection between moral judgments and motivation at face value is usually called 'motivational judgment internalism'. Although 'internalism' itself is a rather vague label that can be used to refer to a number of different meta-ethical views, 'motivational judgment internalism' (hereafter 'internalism') is a relatively clear view. In one way or another, it claims that moral judgments necessarily motivate or, to put this in another way, that there is a necessary connection between moral judgments and motivation.

In the last few decades, 'internalism' has invoked numerous debates concerning whether some form of internalism should be accepted. In his well-known book *The Moral Problem* (1994), Michael Smith famously defended a certain new, weak version of internalism. Shortly after this influential book was published, a large number of different objections introduced by Smith were made and various internalists tried to defend their view against these objections. Yet, if those objections were sufficiently plausible, then we would have strong enough reasons to give up internalism. However, the situation in this respect seems to be somewhat open as not all of the objections to Smith have been evaluated sufficiently and carefully by those who are sympathetic to internalism. Therefore, I think that it is necessary to resume the discussion and see whether plausible responses could be given especially to the unanswered objections. At the

same time, I also hope that in this way we can reveal that the opposite view—externalism—is implausible and thus should not be accepted.

Internalists have, of course, not only have tried to defend Smith's weak version of internalism, but they have also explored whether there are some other forms of internalism that could reflect our moral intuitions about moral judgments in an equally plausible way. As a consequence of this, new forms of internalism continued to be introduced and defended. All these new forms of internalism try to achieve two things at the same time: Firstly, they still try to defend internal moral connection between moral judgments and motivation, so as to support many of our internalist intuitions, like the ones mentioned above. Secondly, they also try to enable internalism to accommodate many of the externalist counterexamples, that is cases where agents do not seem to be motivated by their moral judgments. This attempt has led to more and more sophisticated forms of internalism that indeed weaken the internal connection between moral judgments and motivation and also makes forms of internalism conditional on various factors. One worry emerges thereby, it is whether these internalist views can still retain the original attraction of internalism. In this thesis, I will hope to defend a new form of internalism even if there are certain ways in which my view will be much stronger than some of the other recently introduced internalist views.

This means that my thesis will have three key claims. Firstly, it will create a map of logical space by explaining what different forms of internalism there are available for us on the table (and also how they differ from externalism) in Chapter 2. Secondly, I will aim to provide a conclusive argument against externalism. I do this by developing and defending Michael Smith's fetishism argument in Chapters 3-5. Finally, in Chapters 6-7, I will investigate which

specific form of internalism is the most plausible one that we should accept. I will conclude that by far the most plausible form of internalism to take is constitutional, unconditional, relatively strong, direct internalism with dispositional desires.

In Chapter 2, I will first introduce the most fundamental concepts such as 'moral judgment', 'motivation' and 'motivational judgment internalism' that I will rely on in the rest of this thesis. Furthermore, I will spend the majority of this chapter to introduce different forms of internalism that have been discussed recently in metaethical literature (Sections 2.4-2.7). Additionally, I will also outline some of the main arguments for these views as well as some of the crucial counterexamples to them. After these discussions, I will then finally present a map of the logical space of what forms of internalism there can be, which can both help us to locate the existing forms of internalism and the differences between them and also guide us to new forms of internalism that have not yet been explored. In the end of Chapter 2, I will also explain in more detail externalism, the view that is the main alternative to different forms of internalism.

In Chapter 3, I will focus on the so-called fetishism argument. The fetishism argument starts from an observation of an ordinary phenomenon: if our moral judgments change, this typically causes changes also in what we are motivated to do. Michael Smith (1994) has famously argued that only internalism can provide a compelling explanation of the previous connection between our moral judgments and motivation. Smith claims that, because they deny internalism, the externalists will have to explain the same connection by relying on something other than the moral judgments themselves, for example, on additional desires to do whatever is right. It indicates that, in this externalist framework, ordinary agents would actually have to care more about the abstract property of moral rightness itself much more than the basic considerations

that make different actions right and wrong such as that someone needs help. According to Smith, this would make ordinary agents moral fetishists if externalism were true.

The externalists have adopted two strategies in response to the previous objections. The first strategy has been to argue against the idea that caring about the moral rightness itself turns a moral agent into a moral fetishist. It has been suggested that the desire to do whatever is right is harmless at least when it exists with an agent's other, more concrete desires to do the things that are right. Furthermore, under certain circumstances where an agent lacks specific desires to do right things, a general desire to do whatever is right could be claimed to be necessary for motivating the agent to act in the morally right way. I will consider this first externalist strategy in Chapter 4. I will argue that the externalists' first strategy will still cause other serious issues as a consequence, which makes it implausible to accept the externalist proposals.

The externalists' second strategy of responding to the fetishism argument has been to offer new, alternative and externalist-friendly explanations of the close connection between moral judgments and motivation. The externalists have been tried to argue that these alternative explanations can completely avoid Smith's fetishism objection to externalism because this time the externalists would not explain the connection between moral judgments and motivation by relying on any additional desires with regard to the moral rightness. In Chapter 5, I will examine four externalist alternative explanations in detail and argue that, as far as these theories can in many cases avoid the fetishism objection, they prove to be implausible for other reasons.

Chapters 3-5 thus show that we should at least reject externalism. This conclusion then naturally leads to the following question: which form of internalism should we accept instead? Relying

on the different forms of internalism that were introduced in Chapter 2, I will try to answer this question in Chapter 6 and 7.

The first contrasting forms of internalism that will be discussed in Chapter 6 will be *de re* internalism and *de dicto* internalism. Despite that these two forms of internalism are often understood as conflicting theories, I will argue that they are not mutually exclusive indeed. This is why we cannot argue that we should accept one of them because we have reasons to reject the other. As a result, I will discuss the following four views. Those combinations of views include 1) *de dicto* internalism and *de re* externalism; 2) *de dicto* externalism and *de re* externalism; 3) *de dicto* internalism and *de re* internalism; 4) *de dicto* externalism and *de re* internalism. On the basis of the arguments already discussed in Chapters 3-5, I will argue that we should reject the combinations 1) and 2), which also means that we should definitely reject *de re* externalism and accept *de re* internalism instead. I will also remain neutral about whether or not we should also accept *de dicto* internalism.

Chapter 6 will also include my discussion on the second type of contrasting forms of internalism—unconditional and conditional internalism. When I introduce the conditional forms of internalism in Chapter 2, I already at that point explain why so many philosophers have found these theories very plausible. Yet, in Chapter 6, my ambition is to try to argue for a certain new form of unconditional internalism, something whose plausibility was not fully acknowledged in the past. I will first outline the new form of internalism, which I call 'unconditional internalism with dispositional desires' as it is formulated in terms of dispositional desires. My argument for this view will then be based on a thought experiment, which I hope will illustrate how this new form of internalism is no longer vulnerable to the

traditional counterexamples to unconditional internalism such as depressed people and amoralists.

Chapter 7 will begin by investigating a third type of contrasting forms of internalism, strong and weak versions (though, here the discussion will be formulated in terms of dispositional desires too). Based on the key differences between motivation and dispositional desires, I will first introduce three different ways in which the strength of dispositional desires can vary. One of the benefits of this groundwork will be that there will then be three different ways to argue for stronger forms of internalism, two of which turn out to fail finally. In the third way, it will be argued that the dispositional desires that are required by genuine moral judgments must be able to produce not only motivation but also the so-called reactive attitudes. This is the case even if these dispositional desires need not be able to produce especially strong motivation or motivation in many different kinds of cases.

In Chapter 7, my exploration of which form of internalism is the most plausible one and the one we should accept will end up by discussing the fourth type of contrasting forms of internalism—direct and deferred versions of internalism. I will begin from the arguments that are often used to support deferred forms of internalism, I then hope to show that, the cases used in these arguments, will also require the deferred internalists to make inconsistent *ad hoc* assumptions about when the genuineness of moral judgments depends on the other moral judgments and when it does not. If this is right, I will conclude that the evidence is on the side of direct forms of internalism. I will thus conclude, on the basis of Chapter 6-7, that we should accept constitutional, unconditional, relatively strong and direct internalism formulated in terms of dispositional desires.

Chapter 2: Different Forms of Internalism

2.1 Introduction

This chapter explains some of the most important basic concepts used in this thesis and it will formulate a taxonomy of different forms of internalism. In Section 2.2, I will first explain how I understand and use the terms 'moral judgment' and 'motivation' throughout this thesis. In Section 2.3, I will introduce a very general idea of motivational judgment internalism which is a philosophical theory about the connection between moral judgment and motivation.

Sections 2.4–2.7 then describe different forms of internalism. These four sections include all the main forms of internalism and thus they constitute a helpful guide for understanding the core topic of the debates in which this thesis is taking apart. I will introduce four key distinctions that can be used to explain how different forms of internalism differ from one another. These distinctions are not mutually exclusive, taken into combination, they help us to discern sixteen resulting forms of internalism that are more complex and fairly precise. These distinctions can then be used to formulate, in a precise way, sixteen different forms of internalism. After reading these four sections, it should become evident how the different more sophisticated forms of internalism have been formulated by the defenders of internalism as responses to the problems of the previous, simpler formulations of the core idea of internalism.

Section 2.4 is about the strength of the connection between moral judgments and motivation: this connection could be either strong or weak. Strong internalism suggests that moral judgments lead to strong (overriding) motivation, and because of this, this view has an important advantage: it can provide a compelling explanation of hypocrisy. Yet, strong internalism also faces a serious challenge because it seems unable to make sense of weakness

of will. As a consequence, it is natural to move from strong internalism to weak internalism as the resulting weaker views seem better able to accommodate weakness of will within the internalist framework.

Section 2.5 is about whether the connection between moral judgments and motivation exists unconditionally or conditionally. This section starts from a discussion of the simplest forms of unconditional internalism. Although unconditional internalism can reflect the internalist intuitions well, many people have still thought that there are intuitive counterexamples to it. In Section 2.5.2, I will introduce these examples, which rely on different types of amoralists, depressed and listless people and also bad and evil people. Faced with these challenges, many internalists have adopted conditional forms of internalism. The internalists argue that an internal connection between moral judgments and motivation only exist in certain circumstances. Because of this, they think that their views can deal with the previous counterexamples.

In Section 2.6, I will introduce the distinction between direct internalism and deferred internalism. Direct forms of internalism claims that each moral judgment must be accompanied by the relevant motivation directly, whereas the contrasting views claim that the relevant moral motivation in internalism could be deferred. The deferred forms of internalism thus claim that not all genuine moral judgments need to motivate, as long as they are connected to other moral judgments that do so.

Section 2.7 then introduces the distinction between constitutional internalism and non-constitutional internalism. Constitutional forms of internalism are substantial views about the nature of moral judgments and how such judgments are connected to motivation. In contrast,

the non-constitutional versions of internalism claim that our internalist intuitions cannot support this form of internalism. Rather, the defenders of the non-constitutional views claim that internalism can at most be true at the level of the meaning of the terms 'moral judgment'. Thus, according to non-constitutional forms of internalism, the term 'moral judgment' applies to a mental state only if it is accompanied by motivation. On this view, if a moral judgment is not accompanied by motivation, then we would not use the term 'moral judgments' to describe those states.

Section 2.8 is a brief summary of my discussion of the different forms of internalism. It creates a map of the logical space of different internalist views. This map can be used to locate the existing versions of internalism that have been discussed in the literature so far. It also shows us that there are at least some versions of internalism that have not been explored yet. Finally, Section 2.9 introduces the opposite view, externalism, which argues that there is no necessary connection between moral judgments and motivation.

2.2 'Moral Judgment' and 'Motivation'

A moral judgment is usually assumed to lead to at least some motivation in the agent who makes that judgment.[1] Suppose that we are engaged in a discussion. During the conversion, I tell you that I think that donating a certain amount of money that is well within my means to a charity is the morally right thing to do. Let's imagine that while I am making this very point to you, someone from the local charity happens to knock on the door and ask me for a small amount of money for a charity that supports homeless people. In this situation, it is natural to think that

[1] See Smith (1994, 60 and 71-72) for his basic observation of how moral judgments are usually expected to be connected to motivation in cases like the example above.

you would doubt the sincerity of my moral judgment if I refused to give any money to the fundraiser. After all, others would be very puzzled if I tried to explain to you that I do believe that I should give money to the charity—it is just that I have no desire to do so. This example illustrates how we intuitively think that moral judgments are practical and action-guiding at least in some way. By contrast, we would not assign similar motivational force to a scientific judgment about facts such as that water consists of hydrogen and oxygen atoms.

The previous general description of the example relies on two assumptions. Firstly, it is often assumed in the debates concerning moral motivation that moral judgments and motivation are two different mental states.[2] Secondly, many people in the debate also assume that there must be a causal relation between the moral judgment and the resulting motivation. For instance, according to many of those who think that there is an internal connection between moral judgments and motivation, my moral judgment that it is right to give money for the charity has the power to produce motivation in me. The first assumption is based on what is today called the general theory of Humean moral psychology (Smith 1994, 7-8). The second assumption has been developed further as the thesis that has become known as motivational judgment internalism or simply internalism. In the rest of this section, I will clarify the two fundamental concepts of internalism—moral judgment and motivation—which will then be used throughout the whole thesis.

'Judgment' or 'moral judgment' has multiple meanings. Sometimes it is used to refer to the speech act that I perform when I tell you that donating money to a charity is morally right. On

[2] Expressivists, (for example, Simon Blackburn (1998) and Allan Gibbard (1990, 2003)) reject this assumption as they believe that moral judgments themselves already consist of some form of motivation.

the basis of this speech act, you cannot tell yet whether I am just lying or whether what I am saying reflects what I truly believe. I will not adopt this first meaning of 'normative judgment' in this thesis. Rather, I will use the terms 'judgment' and 'moral judgment' to refer to a certain mental state in which an agent can be. I will assume that you must satisfy certain conditions—you must be in a certain mental state—in order to count as someone who can sincerely accept a moral sentence such as the sentence 'donating money to a charity is right'. I will use the term 'moral judgment' as a neutral way of referring to that mental state, whatever it is like. This is to say that when I satisfy the relevant conditions by being in a certain mental state, you will not doubt my sincerity when I say that donating money to a charity is morally right. If this is the case, then we can call the mental state I am in this case a moral judgment. This second way of understanding the meaning of the term 'moral judgment' that is adopted by most philosophers involved in the debate (e.g. Björnsson et al. 2015, 2; Svavarsdóttir 1999, 167; 2006, 161). Because of this, I will use the term 'moral judgment' hereafter to refer to the mental state, in virtue of which an agent counts as sincerely accepting an indicative moral sentence (Suikkanen 2014, ch. 8).

The way in which I will use the term 'motivation' in my thesis relies on a Humean theory of moral psychology. According to the Humean theoretical picture, all mental states can be divided into two kinds: belief-like states and desire-like states. (Hume 2007[1739-1740], 266-267, book II, ch.3, sect.3) The key difference between these states is often indicated by saying that the belief-like states aim at truth whereas desire-like states aim at something else, like realization or changing the world. The previous metaphor is then unpacked further by explaining how these two kinds of states have opposite directions of fit. In order to explain what this means, I first need to use the following analogy.

Let us imagine that George has been asked by his wife to buy food from a nearby store. To ensure that George will not forget all the different ingredients, his wife has made a shopping list for him. In the supermarket, George walks along the shelves and picks up items one by one in accordance to the list: milk, eggs, pizza, steaks, carrots and so forth. While George is choosing what he wants to buy, a stranger notices George. For some reason, this person is interested in the stuff that George puts in the shopping cart. Every time George puts an item into the cart, the stranger writes down the name of the item in his list. Eventually, the stranger's list is identical with George's shopping list (Anscombe 2000, 56-57).

In this example, George's list and also his own state of mind have the world-to-mind direction of fit. This is because the function of George's list and his state of mind is to make the world (that is, what there is in George's shopping basket) match the list and what George has in his mind. Because George's world-to-mind direction of fit psychological state that makes him pick out certain items and thus moves him to act, this state is often understood as a paradigmatic desire state. In contrast, the stranger's list and his corresponding state of mind can be said to have the mind-to-world direction of fit. The purpose of this list and the state of mind is to match what there is in the world (that is, what there is in George's shopping basket). The stranger's list is thus only satisfied when its content fits the items in George's shopping basket. Because of this, it is thought that the stranger's mind-to-world direction of fit mental state is a paradigmatic belief state.

In this picture of Humean theory of psychology, we can call world-to-mind direction of fit mental states, desires and other desire-like states, states of being motivated.[3] Motivation can thus be regarded to consist of desires the functional role of which is to cause the agent to act in a way that changes the world to match how the agent wants it to be. Motivation is regarded to consist of desires to change the world in the way the agent wants to it to be. By contrast, we can call mind-to-world direction of fit mental states, namely belief-like states, states of judging.[4] Judgments indicate that an agent assents or confirms that what the world is like and forms views according to this fact. If the world changes, then the agent has to modify his own views to match what the world seems to be like.

2.3 Motivational Judgment Internalism

If you share the common intuition behind the donation case introduced in beginning of the previous section, you will probably find at least some forms of internalism to be intuitively plausible. It follows from all forms of internalism that we can usually expect that, when an agent makes a moral judgment, she will have at least some motivation to act in the way that fits her judgment. After all, all forms of internalism assert that there is at least some kind of an internal, modal connection between moral judgments and motivation. However, since there has been so much discussion on the topic, it is noteworthy that the label of 'internalism' has been used to convey very different views even within metaethical moral psychology. In order to

[3] There are subtle differences between the notions of motivation and desire. Yet, for the majority of this thesis, I will use the terms 'motivation' and 'desire' interchangeable. In Section 6.3.2, I will explain the main differences between motivation and desire, and I will explore differences between them further in Section 7.3.

[4] Above, I intentionally defined the term 'moral judgment' in a way that leaves it open whether moral judgments are belief-like or desire-like states.

introduce all the different forms of internalism clearly and then argue for one of them, I will need to introduce a classification of different versions of internalism next.

Roughly speaking, at the most general level, all internalist views are about the connection between two seemingly different mental states: A and B. Initially, we get the first set of different internalist views depending on what specific mental states A and B we take to be connected by an internal connection. Then, after A and B have been fixed, we get different forms of internalism based on what the strength of the connection between A and B is thought to be. We also get different versions of internalism based on under which conditions we take the connection between A and B to exist.[5] With these three principles kept in mind, we are able to explore a variety of different forms of internalism.

As already suggested, the kind of internalism on which I will focus in the rest of this thesis and which also attracts lots of attention in the meta-ethical literature is motivational judgment internalism (MJI).[6] [7] According to all forms of MJI, the variable A in the previous schema is

[5] I borrow this idea from Mark van Roojen (2015). For other general introductions of internalism, please see David O. Brink (1989), Stephen Darwall (1983), Sigrún Svavarsdóttir (1999), Russ Shafer-Landau (2003), Jussi Suikkanen (2014), and Mark van Roojen (2015). For a recent introduction of motivational judgment internalism, see Gunnar Björnsson et al. (2015).

[6] Brink uses *appraiser internalism* to for the same view (Brink 1989, 40). Shafer-Landau uses the same term *motivational judgment internalism* as me (Shafer-Landau 2003, 143). Van Roojen's term for this view is *morals/motives internalism* (van Roojen 2015, 59). Usually, the version of internalism which concerns the connection between moral judgments and motivation is simply called *motivational internalism* (Björnsson et al. 2015, 1; Svavarsdóttir 1999, 163).

[7] For the sake of clarity, we need to distinguish *existence internalism* from *motivational judgment internalism*. According to Williams (1981, 101-113), existence internalism posits a link between reasons and an agent's motivation. On this view, a rational agent can have a reason to φ in circumstances C only if she is at least somewhat motivated to φ in circumstances C (or would be after certain kind idealized deliberation). Brink (1989, 41) has also defended what could be called *hybrid internalism*. Hybrid internalism claims that, if an agent judges that to φ in circumstances C is the right thing to do, she will have a reason to be motivated φ in circumstances C. However, there is currently no widely accepted formulation of this kind of internalism. Neither *existence internalism* nor *hybrid internalism* is directly connected to the main topic of my thesis, and so I will not concentrate on them here.

moral judgments and the variable B stands for motivation. MJI thus focuses on the relationship between a subject's moral judgments and her motivation (when these are understood in the way explained in the previous section). Different versions of MJI thus all posit that there is at least some form of an internal connection between moral judgments and motivation. Different forms of MJI, however, disagree with each other about what the exact nature of the connection between moral judgments and motivation is and how and under what circumstances the connection exists. Despite this, one very general form of MJI could be formulated in the following way:

> *Motivational judgment internalism*: Necessarily, if an agent judges that it is right to φ in circumstances C, then she has motivation to φ in circumstances C.

One clarification needs to be made at this point. MJI is about the connection between moral judgments and motivation. This means that, even if MJI claims something about the nature of moral judgments, it is not a claim about the moral facts at all. This is because, MJI—as a very general formulation of internalism—applies to both true and false moral judgments (Wedgwood 2004, 414). We can illustrate this point with a simple example. Suppose that someone judges that it is morally right to waste food. In this case, we would expect the person who has made this judgment to be motivated to waste extra food given that this is what she believes to be right. Likewise, if someone experiences joy when smoking, he might draw the wrong conclusion that smoking is the right thing to do because it benefits his health. This is a judgment which is obviously mistaken. Yet, on the basis of the mistaken judgment, it is natural to think that the agent will continue to smoke. These two cases illustrate that the truth and falsity of the relevant moral judgments has no impact on MJI.

In the last two sections, I introduced the three most fundamental concepts which I will use in this thesis: ‘moral judgment’, ‘moral motivation’ and ‘motivational judgment internalism’. In the next couple of sections, I will investigate further the different versions of internalism that result from varying strength of the connection between moral judgments and motivation and under which conditions the connection exists.

2.4 Strong and Weak Internalism

Let me begin from a simple version of strong internalism. Before this however, in order to be clear at this point, I think it is helpful to distinguish first between two kinds of strength here, since they are usually discussed together in a confused way. One kind of strength concerns the connection between moral judgments and motivation and especially how strong motivation must follow from a moral judgment. More precisely, it could be argued that either strong moral motivation or weak moral motivation must follow from a moral judgment as I will explain in more detail below. Yet, sometimes when philosophers discuss the strength of a given form of internalism, they have something more general in mind. This is because the strength of a given form of internalism does not always appear to be based on the strength of relevant motivation. It also depends on other factors such as, for example, in how many different circumstances the relevant connection between moral judgments and motivation is thought to obtain. Thus, an internalist view according to which moral judgments lead to strong motivation could still be judged to be a weak form of internalism if the strong motivation is argued to exist only in a limited number of situations.

During the debates in the last two decades or so, most internalists have defended a weak form of internalism in the narrower sense where moral judgments are only thought to lead to some motivation that can be overridden by other desires. Yet, many of these internalists have also accepted that perhaps moral judgments can provide the previous kind of weak motivation only under certain conditions. As a consequence, it has been thought that these philosophers have made their weak internalism even weaker. Despite this, so as to be clear, in this thesis, I will use the language of strong and weak internalism only to refer to the nature of the connection between moral judgments and motivation and the strength of the corresponding motivation.[8]

This Section 2.4 consists of three parts. I will first introduce strong internalism and then explain the advantages of the view: how it can help us to explain certain moral phenomena. Secondly, I will describe one of the challenges to strong internalism based on the idea that in some cases we seem to suffer from weakness of will. The last sub-section will state the basic idea of weak internalism and explain how this view can accommodate weakness of will.

2.4.1 Strong Internalism and Hypocrisy

The simplest possible version of internalism is also a very strong one. It could be formulated in the following way:

> *Strong internalism*: Necessarily, if an agent judges that it is right to φ in circumstances C, then she has overriding motivation to φ in C.

[8] For a related discussion of the strength of different forms of internalism, see Section 7.5.2 below.

Strong internalism claims that there is a very strong, modal and internal relationship between moral judgments and motivation. Firstly, it should be emphasized that if there really were such connection, it would be able to explain why there exists a reliable connection between moral judgments and motivation. Secondly, strong internalism would entail that if someone failed to have the strongest motivation to act as she claims that she judges, the agent could not count as genuinely making a moral judgment. On this view, what an agent does always reveals what her judgments are.

Remember the donation case introduced at the beginning of Section 2.2. Suppose that I have made a judgment that making a donation that is well within my means is the right thing for me to do. According to strong internalism, it follows from this very judgment that I will hereafter have a strong desire to make a donation whenever it is possible—so strong in fact that it will outweigh any other desires I might have. Otherwise, my utterance only remains at the level of insincere speech and cannot be read to express genuine moral judgments.

Historically, one of the first explicit defenders of internalism was Charles Stevenson. In the following paragraph, Stevenson uses a simple example in order to make his readers aware of their intuitions that seem to support strong internalism:

> When you tell a man that he oughtn't to steal, your object isn't merely to let him know that people disapprove of it. You are attempting, rather, to get him to disapprove of it … If in the end you do not succeed in getting him to disapprove of stealing, you will feel that you've failed to convince him that stealing is wrong (Stevenson, 1937, 19).

In the previous quote, Stevenson first suggests that when you have successfully managed to convince another person that certain moral claims are true, that person comes to make a genuine moral judgment. Stevenson's intuition then is that convincing another person about a moral issue requires that the person persuaded comes to disapprove of the act they now judge to be wrong. This case thus seems to indicate that making a genuine moral judgment intuitively requires overriding motivation to act accordingly.

On the basis of the previous example and the intuitions it elicits, Stevenson then put forward his own formulation of strong internalism. According to it, 'a person who recognizes X to be 'good' must *ipso facto* acquire a stronger tendency to act in its favour then [sic] he otherwise would have had' (Stevenson 1937, 16). Many philosophers have thought that the previous statement entails that moral judgments must lead to the strongest motivation in an agent—motivation that can outweigh the agent's other motivations. Let us return to the example in last paragraph. In that example, if it is accepted that the man judges that he oughtn't to steal, then he must have the strongest motivation to avoid stealing (or, in Stevenson's words, he must acquire a stronger tendency to avoid stealing that he otherwise would have had). If he claims that stealing is wrong but still has more motivation to steal than not, on Stevenson's view, the man could not count as having made the relevant moral judgment.

R. M. Hare is also sometimes thought to have accepted a similar view as Charles Stevenson. This is because there are certain paragraphs in which Hare comes close to defending a version of strong internalism. Hare, for example, argues that:

> It is a tautology to say that we cannot sincerely assent to a second-person command addressed to ourselves, and at the same time not perform it, if now is the occasion for performing it and it is in our (physical and psychological) power to do so (Hare 1952, 20).

According to this passage, the connection between an agent's moral judgments and motivation must be strong, so strong in fact that making a moral judgment (assenting to a second-person command addressed to ourselves in Hare's words) must lead to motivation that can make the agent perform the action in question unless external forces prevent her from doing so. Hare thus not only believes that moral judgments are action-guiding, but rather he also seems to suggest that those judgments can issue so strong motivation that it will lead to the relevant actions automatically.

At this point, it should also be granted that strong internalism really has several theoretical advantages. It implies that there is a very strong connection between moral judgments and overriding strong motivation (and consequently even actions). Because of this, if strong internalism were true, this would entail that we could tell whether an agent has made a sincere moral judgment by observing her behavior. In order to illustrate this point, I now turn to the phenomenon of hypocrisy.

Imagine that there is a politician who is running in an election for a high-ranked position. In order to get the votes of the local citizens, this politician makes every effort to please them. During different election rallies, the politician makes many promises to the voters. He says that it is the right time to create more job opportunities for the public; he judges that more of the

budget should be allocated to health, education, and public transport; and he also believes that the government bureaucracy should be reduced. Although these plans might be good ones to have in the actual situation, the politician has promised too much. It would be beyond his ability and power to deliver all the things he has promised once he is finally elected. More importantly, the politician just wants to obtain the position and he has no intention to actually put his promises into effect. Yet, whenever being interviewed or speaking publicly, the politician repeats his promises and convinces the public of them for the only sake of securing a large number of votes. However, when the politician's term of office finally ends, he has failed to keep any of the promises he made.

In the previous case, many of us would find it intuitive to call the politician in question a hypocrite. Hypocrites usually claim that certain actions are right, and they also will tell us that they will perform those actions. However, there is an evident problem that the actions of the hypocrites do not match with what they say out loud—their thoughts do not reflect the judgments they claim to have made.

It seems that one advantage of strong internalism is that it can explain this phenomenon and especially our intuitions about when individuals are being hypocritical. If strong internalism were true, there would be an internal connection between moral judgments and overriding motivation. In this situation, we would be predicting that moral judgments will lead to overriding motivation and furthermore to actions. This would explain why, if there is no overriding motivation and the further actions do not appear, we will seriously doubt whether the agent in questions has made a sincere moral judgment in the first place. Usually, we tend to

call such agents hypocrites, which seems to support the idea that we expect agents to be motivated by their judgments exactly like strong internalism describes.

2.4.2 Weakness of Will and Weak Internalism

Many objections to strong internalism were made already immediately after the publication of Stevenson's ground-breaking work on the topic. Henry Aiken, one of the earliest critics of internalism, already questioned the idea that moral judgments must always lead to strong motivation. As he puts it:

> I may recognize, for instance, that the music of Tschaikowski is 'good' since many honest and discriminating people have affirmed its power to move and to please and yet not in the least be impelled to listen to it … Moreover, during periods of weariness or satiety, especially, 'goods' which we believe and gladly acknowledge to have the profoundest import to ourselves often leave us quite cold, and our judgment that they are 'good' has no magnetism or persuasive power whatever (Aiken 1944, 461; cf. Kauppinen 2007, 111).

In this quoted phrase, Aiken in fact raises two objections to strong internalism—the view according to which moral judgments are strongly magnetic. Aiken first tries to show that we can judge things to be right or wrong without having any motivation to act accordingly as long as other experts endorse those judgments. It is, of course true, that those experts not only have made the corresponding judgments but they are often motivated by their judgments too. Nevertheless, unlike the experts who can be motivated by their judgments, ordinary people in many cases seem to share the experts' judgments without being motivated by them. The second

objection implies that, under certain unusual circumstances, judgments that usually motivate us can leave us cold. Thanks to exceptional situations such as when we are tired, the relevant judgments can have no influence on us at all.

With the second objection, Aiken put forward the akratic challenge to strong internalism. I will focus on this objection more below. But before that, it is worthwhile to consider Hare's famous response to the first objection:

> If I were not accustomed to commend any but the most modern styles of architecture, I might still say 'The new chamber of the Houses of Commons is very good Gothic revival'… I might mean… [the chamber is] to be commended within the class of Gothic revival buildings… The sense… we are concerned is that… 'the sort of Gothic revival building about which a certain sort of people—you know who—would say 'that is a good building' (Hare 1952, 124).

Here Hare argues that, in the first kind of cases introduced by Aiken, we only make what he calls 'inverted commas moral judgments'. Hare's claim is that these so-called inverted commas moral judgments are not genuine moral judgments which have moral content. Instead, Hare assumes that these judgments are actually ordinary empirical judgments that just happen to be about other people's moral judgments. Inverted commas moral judgments are not about what is good and bad or right and wrong but rather they are about what other people regard as such. Aiken's own example of the judgment that Tschaikowski's music is good seems like a good example of an inverted comma judgment. It shows that thinking that music which is appreciated by the experts is good is not identical with sincerely judging that the music is good yourself.

This suggests that, unless an agent's judgment is accompanied with the relevant motivation, we do not necessarily think that the agent genuinely accepts the judgments made by others.

In response to Aiken's second objection, Hare adopted a different strategy. Hare writes:

> If a person does not do something, but the omission is accompanied by feelings of guilt, &c., we normally say that he has not done what he thinks he ought. It is therefore necessary … to admit that there are degrees of sincere assent, not all of which involve actually obeying the command (Hare 1952, 169-170).

Here Hare seems to grant that even genuine moral judgments can sometimes fail to lead to sufficiently strong motivation that would lead to action. This means that at this point Hare seems to be giving up his strong internalism. Yet, the view he seems to accept in the previous passage is still a form of internalism. It claims that when an agent has made a genuine moral judgment and yet failed to have overriding motivation to act accordingly, the agent must feel guilty or have some other residual feelings due to the failure of being motivated. Hare seems to suggest that emotion will expose that the agent has made a sincere moral judgment.[9] In this situation, Hare seems to actually give up strong internalism and adopt a version of weak internalism instead.

Many philosophers have followed Hare in accepting a weak form of internalism. They have moved on to weak internalism because such views do not require that moral judgments always come with overriding motivation. Weak internalism can be formulated in the following way:

[9] A further discussion of reactive attitudes can be found in Section 7.5.3.

Weak internalism: Necessarily, if an agent judges that it is right to φ in circumstances C, she has at least some motivation to φ in circumstances C.

This internalist view still suggests that there is an internal, modal connection between moral judgments and motivation. But weak internalism also admits that the motivation that is entailed by a sincere moral judgment needs not always be the strongest one an agent has: it can in many cases be overridden by the agent's different non-moral desires. For example, even if someone believes that keeping a promise she has made is right, she can still fail to have sufficiently strong motivation to act in accordance to her judgment even when she would be physically and psychologically capable of doing so. Most contemporary internalists have defended different forms of weak internalism. These internalists include at least Simon Blackburn (1998), James Dreier (1990), Allan Gibbard (1990, 2003), Michael Smith (1994; 1996a; 1996b) and many others. It seems that, even if strong internalism did play an important role in the historical debates, weak forms of internalism are more popular today.

2.5 Unconditional and Conditional Internalism

Let us then turn to unconditional and conditional forms of internalism. After weak internalism was introduced by Hare, it quickly became evident that there seem to exist counterexamples even to weak forms of internalism: cases where we intuitively would think that an agent has made a moral judgment even if she has no motivation at all to act accordingly. In the introduction to Section 2.4, I mentioned that, if we think that the internal connection between moral judgments and motivation exists only under certain conditions, we will accept a version of internalism that is already weak in the narrow sense and even weaker in the broad sense.

Thus, in response to the externalist critics and their counterexamples to weak internalism, most contemporary internalists have adopted the 'let us add conditions' strategy. This response eventually led to the development of different forms of conditional internalism.

Section 2.5 proceeds as follows. In Section 2.5.1, I will start by introducing and evaluating weak unconditional forms of internalism—views that are less popular today. In Section 2.5.2, I will then discuss the well-known examples that have been used to challenge both strong and weak forms of unconditional internalism—challenges that appear to be even more effective than the weakness of will challenge to the strong unconditional form of internalism that was discussed above. After this, I will introduce different versions of conditional internalism in Section 2.5.3.

2.5.1 Unconditional Internalism

Stevenson's original strong form of internalism was already a version of unconditional internalism because he thought that there must always be a connection between moral judgments and overriding motivation. Yet, more often unconditional internalism is formulated in the following way as a version of weak internalism:

> *Unconditional (weak) internalism*: Necessarily, if an agent judges that it is right to φ in circumstances C, she will always have at least some motivation to φ in circumstances C (e.g. there are no additional conditions that an agent must satisfy in order to become motivated by her moral judgments).[10]

[10] Daniel Eggers (2015, 85) formulates unconditional internalism in a different way: 'necessarily, if a person judges that it is morally wrong to ∅, then she is, at least to some extent, motivated to refrain from ∅-ing'. He thus advocates a view that is very close to my formulation of the view defended in this

The basic idea of unconditional internalism is that moral judgments essentially lead to at least some motivation in all possible situations—unconditionally, without a requirement that any further condition would need to be satisfied. Consider the following example. Imagine that I promise to meet you at a café at 3:00 p.m. tomorrow. Let us also assume that I am strongly of the view that keeping a promise is the right thing to do. According to unconditional internalism, if I have made such a moral judgment that it is right to keep my promise, I will necessarily have at least some motivation to keep the promise I have made. On this view, moral judgments lead to motivation in all circumstances, rather than only when certain further conditions have been met.

2.5.2 Amoralism, Depression and the Bad People

At the end of Section of 2.4.2, it became evident that many internalists initially hoped to deal with the cases of weakness of will by moving from strong internalism to weak (unconditional) internalism. The internalists tended to grant that moral judgments merely require agents to have some motivation to act accordingly but this motivation needs not always be overriding. Nevertheless, after the internalists adopted weak internalism, new counterexamples to internalism were put forward. The point of this second set of problem cases was to try to show that intuitively we at least sometimes accept that an agent has made a sincere moral judgment even if they have no motivation at all to act accordingly. If such cases existed, then even weak unconditional forms of internalism would have to be false. These purported cases include amoral individuals, depressed and listless people, psychopaths, and even evil people who desire

thesis. Only a few internalist have sympathy for unconditional (weak) internalism—they include pretty much only Danielle Bromwich (2016), Daniel Eggers (2015) and James Lenman (1999), who have all defended the view.

to do bad things.[11] Although the counterexamples vary, they share a common feature: the agents in them seem to remain wholly indifferent to morality. I will introduce these counterexamples next.

Let us begin from the famous, widely-debated case of amoralism. The critics of unconditional internalism have often raised the counterexamples which are based on the possibility that amoralists could exist. Generally, an amoralist is someone who remains indifferent to what she concedes to be a moral consideration. An early, typical description of an amoralist can be found from the works of David Brink (1989, 48). According to Brink, an amoralist is someone who is skeptical about the justification and rationality of moral considerations. Consider, for instance, someone who thinks that eating meat is wrong but is not fully certain about it. This person will question whether rationality requires him not to do what is wrong and, thus, she would count as an amoralist on Brink's view. From the amoralist's own deliberative perspective, it is still an open question for her whether she should act according to her own judgments. To answer this question, an amoralist would require further support for her moral views. The problem, however, is that Brink has not offered a specific example of a person who would be an amoralist. Instead, he seems to think that it is sufficient to focus on whether amoralist skeptics could exist without assuming their actual existence

Sigrún Svavarsdóttir (1999) has also famously described a character how could be argued to be an amoralist in the relevant sense. In contrast to Brink's amoralist skeptics, Svavarsdóttir's model of an amoralist is an agent—call him Patrick—who is a cynic. The cynic Patrick too

[11] Nick Zangwill (2008, 101) suggests that all these phenomena can be labelled as forms of 'moral indifference' as the relevant agents do not seem to care enough about demands of morality.

stays unmoved by the moral judgments that he concedes he makes. The only difference between two kinds of amoralists is that, a cynic does not show any doubt—he does not question the justification for the moral considerations. Rather, he just remains unmotivated by his moral judgments. Svavarsdóttir also assumes that she is describing the story about Patrick in 'purely observational terms' that could be accepted by both internalists and their critics.[12]

In Svavarsdóttir's example, Virginia and Patrick have the following conversation:

> Patrick rather wearily tells her that he has no inclination to concern himself with the plight of strangers. Virginia then appeals to explicit moral considerations: in this case, helping the strangers is his moral obligation and a matter of fighting enormous injustice. Patrick readily declares that he agrees with her moral assessment, but nevertheless cannot be bothered to help. Virginia presses him further, arguing that the effort required is minimal and, given his position, will cost him close to nothing. Patrick responds that the cost is not really the issue, he just does not care to concern himself with such matters. Later he shows absolutely no sign of regret for either his remarks or his failure to help (Svavarsdóttir 1999, 176-177).

In this case, Patrick admits that he agrees with Virginia that there is a moral obligation that requires us to fight against injustices. As a consequence, Patrick cannot be understood as

[12] Adina Roskies (2003, 2006, 2008) discusses examples of patients who have ventromedial damage in hope of supporting the counterexample based on amoralists. These patients appear to be normal in terms of intelligence and reasoning abilities that are measured by a wide range of standard psychological tests. Nevertheless, it is also notable that the ventromedial damaged patients find it difficult to do what they allegedly think are the appropriate actions in the situations they are in. Because of this mismatch between the patients' judgment and behaviour, Roskies believes that the ventromedial damage patients are 'acquired sociopath', i.e. empirically observed amoralists. For responses to Roskies' view, see Michael Cholbi (2006a, 2006b) and Jeanette Kennett & Cordella Fine (2008).

someone who is skeptical about the existence of moral requirements. In other words, he does seem to be someone who makes a genuine moral judgment according to which he is under a moral obligation to help. However, Patrick refuses to offer any help even if doing so would cost him very little. Svavarsdóttir then argues that, despite Patrick's reaction, we should still think that Patrick makes a genuine, sincere moral judgment.

It is true that internalists could still in this case employ Hare's inverted commas strategy to argue that Patrick has not made a sincere moral judgment given that he has no motivation at all to act accordingly. This would be to think that Patrick has only made a non-moral judgment according to which most people are happy to offer help in the situation he is in. Yet, many people have worried at this point that, if we adopted Hare's inverted comma strategy here, we would be begging the question: we would be assuming the truth of internalism to explain our intuitions that do not match the view away. This is why most critics of internalism do not think that Hare's strategy would be a very persuasive defense of the view here.

Let us then consider the counterexamples based on depression and listlessness. In these examples, the depressed and listless agents are claimed to make sincere moral judgments even if they, like the amoralists in the previous cases, have no motivation to act accordingly. Unlike the amoralists, the depressed and listless people can be motivated by their moral judgments in other contexts where they are not depressed or listless. Many people who suffer from depression just cannot be motivated by the very same moral judgments that have motivated them before. Michael Stocker (1979) and Alfred R. Mele (1996, 2003) have both discussed examples of depressed agents. I will consider their objections to unconditional forms of internalism next.

Stocker asks us to consider a politician, who used to be very concerned about the happiness and sorrows of other people around the world and who devoted himself to improving those people's lives. Yet, after decades of trying to help others, this politician only cares about the fortunes of his own family and friends. He still believes that what he used to do is right, but the politician simply does not want to do so anymore (Stocker 1979, 742).

Stocker's claim is that in this case what the politician believed to be good (making the lives of other people better) has ceased to attract the politician for understandable reasons (years of unrewarded services and fruitless efforts). Perhaps the politician has become tired because many of his attempts to try to help others have turned out to be ineffective. In this recognizable case, it does not seem plausible to think that the politician ceases to have motivation to help others because he has changed his moral judgments. Rather, here the lack of motivation could be thought to be based on a long list of considerations which includes 'spiritual or physical tiredness, accidie, weakness of body, illness, general apathy, despair, inability to concentrate, a feeling of uselessness or futility' (Stocker 1979, 744). Due to the previous reasons, the politician could gradually become less and less motivated by his moral judgments to the point where he has no motivation at all to act accordingly. This, of course, would mean that unconditional internalism could not be true.

The role of depression is made even more explicit in another case introduced by Mele:

> Consider an unfortunate person—someone who is neither amoral nor wicked—who is suffering from clinical depression because of the recent tragic deaths of her husband and children in a plane crash. Seemingly, we can imagine that she retains some of her

> beliefs that she is morally required to do certain things—some of her 'MR beliefs,' for short— while being utterly devoid of motivation to act accordingly, or what I term 'MR motivation.' She has aided her ailing uncle for years, believing herself to be morally required to do so. Perhaps she continues to believe this but now is utterly unmotivated to assist him (Mele 2003, 111).

In Mele's case, the unfortunate mother's depression and her lack of motivation are very clearly connected, whereas in Stocker's case, this relationship is merely an implicit one. However, both cases seem to pose a problem for internalism: both the politician and the unfortunate mother are motivated by their moral judgments in many situations even if they are not always motivated by the same judgments as the relevant cases show.

There is at least one reason for why these examples based on depression and listlessness are more difficult for the internalists than the previous cases of amoralism. It cannot be responded to these cases that the agents in them show no sign of making sincere moral judgments at all (as someone like Hare might say about the amoralists). As we can discern that the agents in these cases used to be motivated by their sincere moral judgments, it is more difficult to postulate that these agents would have given up or changed their moral judgments. The depression and listlessness examples thus seem to pose a more serious challenge for the internalists: these cases really appear to reveal that sincere moral judgments do not necessarily motivate agents like the unconditional forms of internalism claim.

The final type of counterexamples concerns evil people who seem to desire explicitly the bad. Blackburn (1998, 61 and 63) introduces the figure of Satan who desires evil in Milton's epic

poem Paradise Lost. Satan was an angel. But due to his rebellion against God, Satan was deprived of his power and exiled from the paradise. After that, even if Satan remembers what is good and he can also recall the time when he stayed with God, Satan still chooses deliberately to pursue evil. For him, a judgment that an action is evil does not appear to yield motivation not to do the action but rather the opposite motivation to do the action instead. Milton's poem thus describes how Satan responds to his past of being an angel with a firm will to do evil. Actually, Satan not only seems to rebel against God, but rather he also seems to rebel against the angel whom he used to be

Milton's Satan seems to respond his own moral judgments in two ways. Satan remains unmoved by the moral judgments about what is morally right—the judgments that used to motivate him in the past. Furthermore, Satan is now motivated by his moral judgment about what is bad and evil even if his motivation is exactly the opposite of what we would normally expect. This example thus leaves internalists with two problems to deal with. The first of these challenges is to explain why Satan is not motivated by the moral judgments that used to move him. The second task is to come up with an explanation of Satan's inclination towards evil that does not threaten internalism.[13]

Before proceeding to the next topic, the following chart will summarize the three kinds of counterexamples introduced in this section. Particularly, it shows how critics of the internalist account for the relevant cases. I hope that this chart will help us to compare their similarities and difference between the alleged counterexamples to internalism.

[13] The first problem will be tackled in Section 2.5. The second issue will be considered in Section 4.3.2.

Questions Answers	Does the agent make moral judgments?	Does she have corresponding motivation?
Amoralists	Yes	No
Depressed and listless people	Yes	Used to have but not now
Evil people	Yes	Used to have Now, the evil motivates instead

We can draw some provisional conclusions from the previous table. Both weak internalists and their critics can agree that the agents in counterexamples make genuine moral judgments. However, the weak internalists and their critics disagree about what conclusions should be drawn from these cases. The externalist Critics tend to think that there is no internal connection between moral judgments and motivation in the cases discussed above. In contrast, many internalists think that there is still an internal connection between moral judgments and motivation in the same cases and they try to defend their view through different ways (see discussions in Section 2.5.3, for example).

2.5.3 Conditional Internalism

The last section introduced some of the typical counterexamples to the unconditional forms of internalism. The previous examples are also enough to give us a sense of the challenges that these counterexamples posit for the defenders of the unconditional forms of internalism. Taken together, they suggest that moral judgments do not necessarily motivate all agents in all conditions even if those moral judgments motivate the agents in many other situations.[14] In

[14] A number of philosophers (for example, Copp (1995 & 1997), Lillehammer (1997), Shafer-Landau (1998 & 2003) and Svavarsdóttir (1999)) criticize this view from their externalist perspectives. Generally, these critics argue that moral judgments can motivate an agent only if she has certain other

response to this objection, many internalists adopt a straightforward strategy. They have argued that certain conditions need to be set on when moral judgments are able to produce motivation.

The resulting conditional forms of internalism can be formulated with the following schema:

> *Conditional (weak) internalism*: Necessarily, if an agent judges that it is right to φ in circumstances C, either she has at least some motivation to φ in the circumstances C or she fails to satisfy certain conditions D.

According to the conditional internalists, moral judgments thus necessarily motivate unless certain abnormal conditions prevent them from doing so. Different ways of specifying the relevant conditions then lead to different versions of conditional internalism.

Before we move on to discuss these versions, it is worthwhile to first consider a potential objection to this strategy, which was first stated by Alex Miller (2013). On his view, internalists tend to think that, when the additional conditions have not been met, something prevents the normal way in which moral judgments give rise to motivation. This entails that, when the relevant conditions have been met and we are not in the exceptional circumstances, the cases of amoralism, depression and listlessness as well as evil people cannot exist. This entailment thereby leads to the concern that the conditional internalists formulate their view merely by precluding all the situations where the counterexamples could be put forward. All that is left of internalism is thus the claim that internalism is true except when internalism is not true.

---

desires, such as the desire to do the right thing. I will introduce this opposite view below in Section 2.9 and argue against it in Chapters 3-5.

Actually, it seems that the conditions in which there is no connection between moral judgments and motivation have quickly been given an insubstantial characterization. Because of this consequence, the resulting forms of conditional internalism become trivial. This is why Miller suggests that, unless a more substantial description of the relevant conditions is given, conditional internalism will fail to be a substantial, interesting philosophical view (Miller 2013, 219- 221).

This is why, in the rest of this section, I will investigate versions of internalism that try to avoid Miller's objection. These views introduce different kinds of conditions: psychological normalcy, practical rationality and moral perception. I will discuss how these versions of internalism try to both avoid the previous counterexamples to internalism and provide a substantial description of the conditions in which moral judgment must lead to motivation (so as to avoid Miller's concern).

Psychologically normal: First, let us consider a form of conditional internalism according to which conditions D in the previous schema are the conditions of psychological normalcy. Many philosophers who defend this version of conditional internalism are also expressivists. They usually tend to think that moral utterances express desires-like attitudes, or, in other words, they argue that moral judgments are some type of desire-like attitudes.[15] Simon Blackburn has thus famously argued that moral judgments consist of a number of kinds of related desires (Blackburn 1998, 67). For instance, if someone judges that eating meat is morally permissible,

[15] There is another view based on the function of moral judgments which also endorses the motivating function of moral judgments. The other theory draws an analogy between moral judgments and certain natural objects such as body-parts. Bedke (2009, 196) argues that moral judgments have evolved for certain purposes, one of which is to motivate moral behaviors, in the same way as natural objects such as body-parts (consider for example hearts or eyes) obtain their purposes through evolution.

this judgment will consist of a variety of different corresponding desires. This network of desires not only includes the desire to eat meat, but also a desire to desire to eat meat, a desire for other people to eat meat, a desire for other people to desire to eat meat, a desire for other people not to desire not to eat meat and so forth. Desires within this network are all related and support each other.

Similarly, Allan Gibbard has argued that moral judgments consist of certain kind of sophisticated planning attitudes towards the relevant actions (Gibbard 2003, 153-154). For the same case concerning eating meat, according to Gibbard, if an agent has judged that eating meat is morally permissible, she then has an attitude of planning to eat meat at least in many of the circumstances she could be in.

Many conditional internalists who formulate their view in terms of psychological normalcy thus think that moral judgments consist of certain sophisticated desire-like states. These psychological states could then be argued to give rise to motivation to do specific actions only in conditions in which a given agent is psychologically normal. This would explain why there would also be circumstances in which agents are not motivated by their moral judgments. Even if the following list of abnormal conditions is only provisional, it should, however, include at least psychological conditions such as 'despondency, severe depression, physical or mental exhaustion' (Eriksson 2006, 174). The list of abnormal psychological conditions should also include 'states of listlessness … sleep deprivation, sickness or personal loss' (Björnsson 2002, 335). The intuitive thought then is that, under these abnormal situations, the connection between moral judgments and motivation may not happen as usually expected because the agent's psychology is not functioning in the normal way.

Generally, the previous psychological conditions are abnormal ones, which is why the type of conditional internalism under discussion can argue that it is not a problem if moral judgments fail to motivate accordingly in the abnormal conditions. Under such conditions, our moral judgments are argued to malfunction and thus it is not an issue that there is no corresponding motivation. Suppose that a person is suffering from a serious emotional disturbance or depression. These physical disturbances will presumably influence the consequences of that person's moral judgments—even if these judgments normally result in motivation, the previous kind of disturbance can plausibly be thought to block the motivating force of her judgments.

Yet, even though on this view moral judgments can fail to function as they are supposed to do under the previous types of unusual circumstances, this is still compatible with the claim that moral judgments retain their necessary motivating function under the normal circumstances. If this is the case, then counterexamples such as depression will not be a challenge for the resulting forms of internalism. The defenders of this view can always argue that the agents in the proposed counterexamples fail to satisfy the condition of being psychologically normal.

Practical rationality: We can then turn to the second form of conditional internalism which is the practical rationality version. According to it, the conditions D in the basic schema of conditional internalism given above consist of practical rationality. Thus, on this view, if an agent judges that it is right to φ in circumstances C, she either will have at least some motivation to φ in the circumstances C or she is practically irrational.

There are, of course, many different views about what is required for being practically rational. However, according to Michael Smith, who is the best-known defender of conditional internalism based on practical rationality, in order to be fully rational, an agent has to meet four requirements: she should have no false beliefs, she should have all the relevant true beliefs, she should have a systematically justifiable set of desires and she should not suffer from any physical or psychological disturbances (Smith 1994, 156-161; 1996a, 160).[16] Smith claims that when rationality is understood in the previous way, the practical rationality version of conditional internalism entails the following proposition: a rational agent who makes a moral judgment will necessarily have some motivation to act accordingly. The principle also allows that an irrational person can make a moral judgment without having any motivation. However, even in these cases, rationality requires the relevant agents have at least some motivation.

The first and second requirements can be explained together. This element of being rational can be illustrated with an example from Bernard Williams (1981, 102). Let us imagine that a person desires to drink gin from a glass in front of him even if actually the liquid in the glass is petrol. Given that the person believes that there is gin in the glass and he only desires to drink from the glass because of this belief, she will desire to drink it. But this does not seem to be what a rational agent would desire in this case, because the person's desire is based on a false belief: he mistakes petrol for gin. If the agent had the true belief about what there is in the glass instead of the false one, she would not drink the liquid in the glass because that action would not give him what he wants. This is why the actual agent's desire is based on a false belief and, therefore, we intuitively take it to be an irrational desire.

[16] Originally, Smith (1994) did not make the absence of physical or psychological disturbance a condition of full rationality, whereas later on he did explicitly endorses this requirement (Smith 1996a, 160).

By the third requirement, 'a systematically justifiable set of desires', Smith means that rational agents have a set of coherent and unified desires. This is to say that a rational agent's desires do not first of all contradict with each other. They do not pull the agent towards different directions at the same time. Additionally, the desires in the set support each other: they are in harmony with each other.[17]

For illustration, if I feel cold in a low-temperature room, I may have a desire to turn up the heater and to put on more clothes. I could also have a desire not to open the windows as doing so would bring more cold air into the room. I can even have a higher-order desire to desire myself to desire to take measures to keep the room warm. In addition, I can continue to desire others in the room to desire as I do. My desires in this case are what Smith calls a systematically justifiable set of desires. It is evident that my desires aim at the same direction and they support one another rather than contradict with each other. Because of this, having such a set of coherent and unified desires should be thought as rational. However, suppose that I would desire to turn up the room temperature and open the windows at the same time. In this situation, all my desires would be pulling towards entirely different directions and I would not get what I want as a result—I would never end up feeling warm up. As a consequence, maintaining a set of incoherent and disjointed desires seems clearly irrational.

And finally, the last element of Smith's characterization of practical rationality is similar to the psychological normalcy condition discussed above. Smith considers cases of depressed people (Smith 1994, 155). In these cases, emotional disturbances are assumed to have a crucial

[17] I will discuss in detail a systematically justifiable set of desires in Section 5.5.2.

influence on the agents' mental states. Yet, it does not appear like exceptional, psychological disturbances are successfully precluded by the previous requirements of rationality (Smith, 1994, 158). Therefore, to define the requirements of rationality accurately, we need to add explicitly that in order to be fully rational, one must not suffer from any psychological disturbances.

Smith uses the previous understanding of rationality for two purposes. He uses it both for explaining how rational agents are motivated by their moral judgments and also for dealing with previous counterexamples to internalism. Now, let us see why, according to Smith's version of conditional internalism, a rational agent would have at least some motivation to act in accordance with her moral judgments precisely because of her practical rationality.

Smith begins first by analyzing the concepts employed in moral judgments (in a similar way as we could try to analyze other concepts). Smith claims that moral concepts can be reductively understood to be about reasons for actions (Smith 1994, 62). Take the concept of 'a bachelor' for illustration. When I think that Mark is a bachelor, what I am thinking of is that Mark is a male and unmarried. This is because the concept of bachelorhood can be reductively analyzed in terms of being male and unmarriedness. On Smith's view, moral judgments are always in part judgments about what you have reasons to do. When an action is judged to be right or wrong, a part of this thought is always that there are at least some reasons either to perform or refrain from doing the action. Consider a specific example concerning moral judgments. When an agent believes that helping innocent people is the right thing to do, according to Smith, she is thinking that she has good reasons to help innocent people.

According to Smith, the content of an agent's moral judgments, that is, the content of the thought that there are reasons for actions can be investigated further. His proposal is that, when an agent believes that there are reasons for her to carry out a certain action, she essentially believes that her fully rational version would want her to do that action in the actual situation she is in (Smith 1994, 151-152). So, for example, an agent's judgment that it is right to help the innocent people is a judgment about what she has reasons to do. And, the content of this judgment, according to Smith is that her fully rational version would want the agent to help the innocent in the situation she is in.

At this point, the agent has two options: either she will desire to help those innocent people to get rid of the plight or she will lack that desire. Smith then claims that, as a rational agent, in this situation the agent would desire to help the innocent people merely due to her rationality. Because practical rationality can be thought to consist at least in part of a disposition to have coherent mental states, practically rational agents are disposed towards coherence. It is then plausible to suggest that a desire to help innocent people coheres better with the belief according to which the agent's fully rational version would want her to help the innocent. In contrast, if the agent lacked that desire, it could be argued that what she wants to do does not cohere with what she herself believes that her better version would want her to do. This means that, when an agent is practically rational, she will desire to act in accordance with her moral judgments, or so Smith argues.

Secondly, Smith and the other defenders of this version of conditional internalism can respond to the different counterexamples to internalism exactly in the same way as the defenders of the psychological normalcy version of the view. For depressed and listless agents, it could be

argued that they are disturbed by certain psychological disorders, which makes them practically irrational in light of Smith's account of the fourth element of practical rationality. Because these agents are not fully rational and so are also not under the relevant circumstances in which moral judgments are claimed to motivate necessarily, they are not counterexamples to the introduced form of conditional internalism. Conditional internalists like Smith can allow agents to fail to be motivated by their moral judgments as long as this failure of motivation can be explained by relying on different kinds of practical irrationality.

Morally perceptive: The third form of conditional internalism to be introduced here is conditional on moral perception. The conditional internalists who accept this version of the view tend to believe that moral judgments are able to motivate agents who are morally perceptive in the relevant situations. John McDowell is thought to have defended this particular version of conditional internalism (McDowell 1978, 1979). [18] [19] In a debate with Phillip Foot, McDowell famously argued that an agent's conception of a situation can on its own suffice to motivate the agent.

The notion of perceptual capacities lies at the heart of McDowell's view of how moral judgments motivate. According to him, an agent who can be motivated by her moral judgments masters a kind of an ability. That perceptual capacity can be obtained through moral education and then, once required, be exercised even in entirely new environments (McDowell 1978, 23).

[18] William Tolhurst (1995) and David Wiggins (1991) could be read to express similar views.
[19] McDowell's view should be classified as a strong version of conditional internalism. According to him, moral considerations constitute independent reasons that can override reasons based on non-moral considerations. This is why, on his view, agents who perceive moral considerations correctly would have corresponding motivation to act according to their moral judgments. (McDowell, 1978, 26; 1979, 335).

Based on concrete cases, moral education also cultivates children to acquire certain set of rich cares and concerns. Having the capacity in question thus means that the agent has acquired certain knowledge-how that enables the agent to identify different kinds of moral situations and how to fulfil the moral requirements in those circumstances.

The application of perceptual capacity thus enables an agent who has it to recognize moral situations when she faces them. Furthermore, the agent's cares and concerns that are an important part of her moral sensitivity will then enable her both to see the situations in the right way and also to be motivated by those perceptions (McDowell 1979, 343). McDowell thus suggests that a morally capable agent can form the relevant moral motivation in response to the encountered moral facts on the basis of her finely tuned moral sensitivity that consists of different cares and concerns shaped by the agent's upbringing (McDowell 1979, 333).

An example can be considered to illustrate how an agent with the previous kind of a perceptual capacity responds in an ordinary moral scenario. Suppose that the agent's friend is in trouble—she needs help in order to get out of the difficult situation she is in. Should the agent be equipped with the capability of moral perception and the sensitivity it is based on, the agent will be able to recognize that her friend needs help. Since the agent, because of her cares and concerns, is sensitive to the fact that it is both kind and necessary to help her friend, the fact that the agent's friend is in trouble will be sufficient to motivate the agent to help her friend. On this view thus, when an agent who makes moral judgments is morally sensitive, moral facts themselves can lead to relevant moral motivation in the agent.

The conditional form of internalism is thus in this case conditional on the moral perceptual capacity. As with the defenders of the previous forms of conditional internalism, the defenders of this view too have tried to address the counterexamples of amoralists and depressed people, which were freshly discussed in Section 2.5.2.

The defenders can thus argue that the unmotivated agents in the relevant cases lack the necessary abilities that are required for being morally perceptive. Since the amoralists lack the required perceptual capacities, they will also lack the cares and concerns which the morally sensitive agents have. As a result, the amoralists are unable to perceive the moral scenarios in the same way as the morally sensitive agents. This is why, when the amoralists encounter different kinds of moral situations, they will be unable to respond to those moral issues in the right way by being motivated to do the right actions.

At the beginning of this section, I introduced Miller's objection towards conditional internalism. Miller's concern is that internalists rely on circular ways of excluding counterexamples and as a consequence, conditional internalism becomes trivially true. In this Section 2.5.3 on conditional internalism, I have investigated three forms of conditional internalism that are conditional in different ways. These conditional forms of internalism will not be trivially true simply because they finally formulate conditions which they then use to deal with counterexamples. Furthermore, the defenders of the previous three forms of conditional internalism argue that they have provided an independent, substantial explanation of the conditions in which moral judgments motivate. Since moral agents in the counterexamples fail

to satisfy the proposed internalist conditions, it is understandable that they remain unmotivated by their moral judgments.[20]

2.6 Direct and Deferred Internalism

All the versions of internalism discussed so far in the previous Sections 2.4 and 2.5, even if they are different from one another in many ways, they still share a common feature. All these views assume that at least when an agent makes a moral judgment in the idealized circumstances, necessarily, she will have some motivation to act accordingly in that very same situation. According to these views, the relation between moral judgments and motivation is direct, which means that every moral judgment has to be accompanied with motivation at least when the relevant conditions have been met.

However, some critics of conditional internalism claim that we can still find agents who make sincere moral judgments, satisfy all the relevant conditions set by the conditional internalists, and yet fail to have any motivation. In response to these concerns, some internalists have explored new ways of trying to accommodate the previous kind of agents in the internalist framework. These internalists concede that even agents who satisfy the conditions mentioned in Section 2.5 can sometimes make moral judgments without having any corresponding motivation. According to them, this can happen as those moral judgments are suitably related to other moral judgments that lead to motivation in the required internal way.

[20] The critics have not always been convinced by the idea that conditional forms of internalism can avoid the resulting triviality. For discussion, see Sayre-McCord (1997).

Here, two types of moral judgments can be distinguished from one another on the basis of Hilary Putnam's famous example of elm trees and beech trees (Putnam 1973, 704). Putnam asked us to imagine a situation in which we, you and me, are unable to tell the difference between elm trees and beech trees. Putnam thought that we can still both use the concepts of elm trees and beech trees to refer to certain species of trees. This is because there are at least some experts who can tell the difference between the relevant species. Yet, if there were no one who could recognizes the elm trees and beech trees, it would not be clear whether these concepts could be used meaningfully by any of us. Thus, according to Putnam in the previous case, even if we do not know the full meaning of the concepts of elm and beech trees, we can still make judgments by using these concepts because our judgments are suitably related to the judgments of the specialists who are able to distinguish between the trees.

It could then be suggested that the moral judgments that do not motivate are in one important respect similar to our judgments than employ the concepts of elm trees and beech trees in the previous example. Sometimes moral judgments do not have an internal link with motivation. Still, they count as genuine moral judgments as long as they are connected in the right way to the proper moral judgments that do motivate. We can call the resulting new forms of internalism versions of deferred internalism so as to distinguish these views from the previous direct versions of internalism. The new deferred version of internalism can be formulated with the following schema:

Deferred (weak) internalism: Necessarily, if an agent judges that it is right to φ in circumstances C, then she has at least some motivation to φ or her moral judgment is

connected in a certain way W to some other moral judgments that are accompanied by motivation.

The internalists who accept a deferred version of internalism can adopt the other internalists' explanations of why the paradigmatic moral judgments necessarily motivate. This means that many of the insights of the internalist views discussed in Sections 2.4 and 2.5 of how moral judgments motivate can still be thought to apply here. The deferred part of internalism is only required to explain how an agent whose own moral judgments fail to motivate her in a given situation must be connected to other moral judgments that are accompanied by motivation. In the rest of this Section 2.6, I will explain how the deferred part of internalism works. I will consider two forms of deferred internalism: individualist deferred internalism and communal deferred internalism.

Let me first begin from an individual version of deferred internalism. Simon Blackburn (1998) illustrates this type of deferred internalism with the already discussed case of Satan. As I mentioned in Section 2.5.2, Satan not only remains unmoved by his moral judgments concerning right actions, but he even pursues evil intentionally. As Blackburn puts it, cases like Satan are parasitic upon a background connection between moral judgments and motivation (Blackburn 1998, 61). Before falling from the paradise, Satan was an angel who was acquainted with what the right thing to do is in each case. He just became angry because he was exiled from the paradise by God, which made him want to do exactly the opposite of what is good. Yet, even after Satan has become evil, his moral judgments are still connected to other moral judgments that motivate necessarily—namely, his own previous judgments. It is merely that

Satan's own moral judgments about what is right only trigger the opposite motivation that before—motivation for doing what is wrong.

Likewise, the counterexamples of depressed and listless agents (see Section 2.5.2) can also be explained away by relying on the previous type of deferred internalism. The defenders of these views can claim that the politician is still making the same genuine moral judgments with exactly the same content as before. It is just that the politician's moral judgments just cease to motivate him. Yet, the defenders of deferred internalism can argue that the politician's judgments, the ones that no longer motivate him, are still genuine moral judgments because they are connected in the right way to his own previously motivating judgments. This is why if the individualist deferred internalism were true, then many of the counterexamples to internalism would appear to go away.

We can then consider the communal versions of deferred internalism. Some critics of internalism might think that, even if individualist deferred internalism can deal with cases such as the evil people and depressed agents, it is cannot explain away amoralism because an amoralist is never motivated by her moral judgments and so there are no previous moral judgments with motivation to be related to in the right way. Critics can also further propose that, if we can imagine an individual amoralist who is unmotivated by her moral judgments, then it must also be possible to imagine a whole community of this kind of amoralists.

The defenders of the communal forms of conditional internalism at this point argue that a community which would only consist only of amoralists does not seem to be quite imaginable.[21] To see why this would be the case, let us first imagine a planet called Amorality, a community which consists of only amoralists. This planet looks similar to our planet in many aspects. The people there also use the expressions 'right' and 'wrong' for certain actions—the same actions we call right and wrong, even if they seem to use these words in one crucial respect in a way different from ours. When you visit the planet Amorality, you will observe that no one there is ever motivated by his or her moral judgments. This means that moral judgments do not prompt the citizens of the planet Amorality to do the right things or refrain them from doing the wrong things. This description indicates that, on this planet, morality no longer serves as a practice that guides actions and, as a consequence, good behaviors are not encouraged on this planet and wrong actions are not disapproved of. Let us then imagine that, in one case, one of your hosts tells you that it would be wrong for him to help an elderly person across the road, but then he happily does so. At this point, the deferred internalists can refer to two intuitions about the previous case.

First, they would argue that in the previous case, it would not make sense to disagree with the host or try to convince him that actually helping an elderly person is right. After all, even if the host accepted that helping the elderly person were right, her corresponding motivation would not change as a consequence as the host must already be motivated to help the elderly person to cross the road. So, moral disagreement seems to be both impossible and unnecessary in the planet Amorality case. Secondly, many deferred internalists also claim that, when you return

[21] See James Dreier (1990), James Lenman (1999), Jon Tresan (2009b) and Matthew Bedke (2009; 2019) for discussions of similar cases.

back home from the planet Amorality, you could not translate the host's utterances into our own moral language. For example, you would not tell your friends that, on planet Amorality, they think that helping an elderly person across the road is wrong. This is because it seems clear that the citizens of the planet amorality do not use moral language with a similar practical significance as our own moral language. Relying on these two intuitions about moral disagreement and translation, the deferred internalists argue that amoralists are not making genuine moral judgments.

If the citizens of the planet Amorality are not making sincere moral judgments, we should thus not think that a whole community of amoralists could exist. Even if there could be amoralist individuals in a community, we seem still to believe implicitly that at least some people must be motivated by their moral judgments in a community for anyone to be able to make moral judgments in that community. The implausibility of a community which would merely consist of amoralists makes that communal versions of deferred internalism appealing. Unlike in the case of the individual versions of deferred internalism, according to the communal views an agent's deferred moral judgments need not link to his or her previous moral judgments. Instead, on this view, even an agent whose moral judgments have never motivated her can make genuine moral judgments as long as those judgments are connected in the right way to the moral judgments of others that motivate them. This is why many internalist have accepted a communal version of deferred internalism as such views seem to be able to also deal with the difficult cases of amoralism.

2.7 Constitutional and Non-constitutional Internalism

The previous Sections 2.4, 2.5 and 2.6 help us to understand different internalist views of how moral judgments motivate and also how different versions of internalism try to deal with the various counterexamples. All the previous different internalist views seem to agree that moral judgments are very specific kind of psychological states that are connected to motivation in some special way. Jon Tresan (2006, 2009a, 2009b), however, suggests that the internalists should consider the notion of moral judgments from a semantic perspective, which focuses on the meaning of words 'moral judgment'. Internalism understood in the traditional way as a view of the motivating role of certain kind of psychological states is a *de re* claim of the nature of those states and, in other words, thus a constitutional theory. Internalism understood in the new way recommended by Tresan is a semantic view of the meaning of phrase 'moral judgment'. This would make it a *de dicto* claim and hence, in other words, a non-constitutional theory. As the *de re* or constitutional versions of internalism have been discussed in detail above, this section will mainly focus on the *de dicto* or non-constitutional internalism.

In order to understand *de dicto* internalism, we should begin from Tresan's example of two different sentences that appear to have very different meanings:

(1) Necessarily, planets are accompanied by stars.

(2) Planets are necessarily accompanied by stars.

Although these two sentences use the same words, they differ in meaning. The relevant type of necessity in (1) only claims that a certain proposition is true and thus here the modal operator 'necessarily' has a wide scope. In contrast, the relevant type of necessity in (2) claims that

certain objects have a certain property necessarily and thus here the modal term has a narrow scope (Tresan 2009a, 54-55). (1) means that the sentence of 'planets are accompanied by stars' is necessarily true. This sentence, of course, must always be true because, according to the definition of a planet, things that can be called 'planets' must be orbiting stars. If there is an astronomical body which is not orbiting a star, we just cannot call it a planet. However, according to Tresan, (2) means something very different. It seems that, if we want to discover whether (2) is true, we cannot merely rely on semantic analysis but rather we would have to investigate certain celestial bodies, the planets, themselves to see whether they must be orbiting stars. But this *de re* claim about the astronomical bodies called planets turns out to be false. These very things could well exist in a possible world even if they did not orbit stars there (Tresan 2006, 145).

Tresan has then argued that the modal claims put forward by the internalists can too be understood as either *de re* or *de dicto* claims. The internalists could have in mind either one of the following two claims:

(3) Necessarily, moral judgments are accompanied by motivation.

(4) Moral judgments are necessarily accompanied by motivation.

(3) is a formulation of a *de dicto* internalist view. The idea is that, if this claim were true, it could be shown to be true by analyzing the meaning of the terms 'moral judgment' in the sentence 'moral judgments are accompanied by motivation' in which those terms are embedded. This view claims that the semantic meaning of term 'moral judgment' already includes a connection to moral motivation. Yet, note that, in the case of the planets, the truth of (1) tells

us nothing about the nature of planets (the relevant celestial bodies would continue to be exactly the same even if they did not orbit stars). Similarly, here too (3) is not supposed to tell us anything about the nature of moral judgments. It is just that certain mental states that are called moral judgments only when they are accompanied by motivation. Yet, these states would remain exactly the same even in situations in which the agents who are in these states are not motivated by them.

By contrast, according to Tresan, (4) captures the more traditional *de re* internalist views. Proposition (4) claims that moral judgments are certain specific kinds of substantial mental states. Recall that, if we wanted to know whether (2) is true, we will have to investigate planets in the actual world and what modal qualities they have. Likewise, if we wanted to justify (4), we would have to investigate what moral judgments consist of in the actual world and what modal qualities they have as a consequence. In the actual world, if during the investigation, we find that the nature of moral judgments is such that moral judgments are able to produce motivation, then (4) could be claimed to be true. This is why, (4) cannot be regarded as true before concrete exploration in the same way as we cannot directly confirm the truth of (2) merely by doing conceptual analysis.

However, according to Tresan, the counterexample concerning amoralists are enough to show that the proposition (4) is not plausible and thus our internalist intuitions can at most support the proposition (3) (Tresan 2006, 149). He asks us to consider the counterexamples discussed in the previous Section 2.5.2. It seems that people like Patrick—an amoralist—could be argued to exist in the actual world. According to the non-constitutional internalism, Patrick can be in the very same belief mental states as others. But as there is no relevant motivation in Patrick's

psychological make-up, it is just that we would not call those mental states moral judgments. In contrast, according to constitutional internalism, because of their very essence, moral judgments can produce motivation. This means that if an agent's moral judgments fail to produce motivation, the agent cannot be in the mental state that we call moral judgment. yet, in order to accommodate both our internalist intuitions and to accommodate the existence of amoralists, internalists should, according to Tresan, give up the proposition (4) which is the *de re* internalist claim.

2.8 A Summary of Different Types of Internalism

I have already introduced all the main forms of motivational judgment internalism that have been discussed in the literature so far. I have also briefly discussed some of the best-known counterexamples to internalism and how the defenders of different versions of internalism have responded to them by developing more sophisticated forms of internalism. When we review different versions of internalism in this way, it becomes evident that there are four key choices (strong or weak, unconditional or conditional, direct or deferred, constitutional or non-constitutional) for any internalist to make. Based on these options, sixteen different versions of internalism can be formulated as illustrated in the following map.

		Direct		Deferred	
		Constitutional	Non-constitutional	Constitutional	Non-constitutional
Strong	Unconditional	(1) Strong unconditional direct constitutional internalism	(2) Strong unconditional direct non-constitutional internalism	(3) Strong unconditional deferred constitutional internalism	(4) Strong unconditional deferred non-constitutional internalism
	Conditional	(5) Strong conditional direct constitutional internalism	(6) Strong conditional direct non-constitutional internalism	(7) Strong conditional deferred constitutional internalism	(8) Strong conditional deferred non-constitutional internalism
Weak	Unconditional	(9) Weak unconditional direct constitutional internalism	(10) Weak unconditional direct non-constitutional internalism	(11) Weak unconditional deferred constitutional internalism	(12) Weak unconditional deferred non-constitutional internalism
	Conditional	(13) Weak conditional direct constitutional internalism	(14) Weak conditional direct non-constitutional internalism	(15) Weak conditional deferred constitutional internalism	(16) Weak conditional deferred non-constitutional internalism

This table can help us to locate the different forms of internalism that have been mentioned in this chapter. For example, Stevenson's strong internalism mentioned in Section 2.4 should be located as (1) strong unconditional direct constitutional internalism. Similarly, Smith's view

can be located as (13) weak conditional constitutional internalism. We can also identify Tresan's *de dicto* internalism as (12) weak unconditional deferred non-constitutional internalism.

Additionally, we may rely on the table above also to explore new forms of internalism that have not yet been discussed. It might lead us to find new plausible forms of internalism. Take, for instance, (4) Strong unconditional deferred non-constitutional internalism, a view that has not been considered yet. Yet, even if this version of internalism seems odd and surprising at first, it might still turn out to be a plausible candidate if powerful arguments were offered in its support. This is why the internalists should systematically continue to explore, which of the options shown in the table above is the most plausible of internalism.

2.9 The Externalist Challenges

The view according to which all forms of internalism are false is, of course, called externalism. Most externalists agree with the internalists that people usually have motivation that matches their moral judgments. But they grant that the connection between moral judgments and motivation is merely contingent even if it is reliable at the same time. Accordingly, the externalists make the straightforward claim that there is not any kind of a necessary connection between moral judgments and motivation. Many of the externalist counterexamples to different versions of internalism were discussed in Section 2.5.2. It is precisely these counterexamples, which the defenders of externalism mainly rely on in their arguments for their externalism according to which moral judgments motivate only contingently. Some well-known externalists discussed in this chapter including Brink (1986), Svavarsdóttir (1999, 2006), Stocker (1979) and Mele (1996, 2003), Zangwill (2008), have all discussed these counterexamples.

Unlike internalism, which is advanced in a number of different versions, externalism can be formulated in a much simpler way on the basis of various externalist responses.

> *Externalism*: If an agent judges that it is right to φ in circumstance C, then she will have motivation to φ only if some external contingent facts—other than her own rationality—connect her moral judgment to motivation.

Yet, given that we expect that agents are usually motivated according to their moral judgments, more needs to be said about how moral judgments are thought to motivate in the externalist framework. According to the externalists, if an agent has motivation to act in accordance to her moral judgments, some external, contingent facts must obtain to explain why the agent's moral judgments motivate her in the situation she happens to be in. Different externalists then refer to different contingent external facts to explain the contingent (and yet usually reliable connection) between moral judgments and motivation. Let me firstly briefly introduce some of the typical externalist explanations of the connection between moral judgments and motivation here.

Firstly, many externalists explain the way in which moral judgments are connected to motivation with the *de dicto* desire to do whatever is right (Lillehammer, 1997; Shafer-Landau, 1998, 2003; Svavarsdóttir, 1999). Consider an agent who makes a moral judgment that, for example, it is right to fight against discrimination. Let us imagine that this agent really happens to have a *de dicto* desire to do whatever is right. At this point, according to many externalists, the agent's moral judgment and her general desire to do whatever is right will generate a more specific desire to fight against discrimination as having that desire will be instrumental to

getting what the agent ultimately wants—to do the right things. I will explain and argue against this type of externalist theories in detail in Chapter 4. The one externalist theory I will discuss here is based on the *de dicto* desire to do whatever is right.

Yet, there are also other ways in which externalists have tried to explain the reliable connection between moral judgments and motivation. Some externalists, for example, believe that a good and strong-willed person can be disposed to desire to do what she judges to be right directly (Copp 1997). Other externalists try to explain the same reliable connection in ways, for example, by virtues (Cuneo 1999), second-order desires (Dreier 2000), and reasons (Lillehammer 1997). I will explain and evaluate these proposals in detail in Chapter 5.

2.10 The Concluding Remarks

My aim in this chapter was to introduce different forms of internalism. Before doing so, I first, in Section 2.2, explained how I will use the two core concepts, 'moral judgment' and 'motivation', in the rest of this thesis. I then explained in Section 2.3, the basic crux of the term 'internalism'—it denotes generally views according to which there is a close, internal connections between two different kinds of mental states. Specifically, I explained how I will focus on 'motivational judgment internalism' which suggests that there is a certain kind of internal connection between moral judgments and motivation.

Sections 2.4-2.7 then described different forms of motivational judgment internalism, or simply internalism. Based on four theoretical choices, different forms of internalism have been categorized under four kinds: 'strong or weak', 'unconditional or conditional', 'direct or deferred' and 'constitutional or non-constitutional'. In Section 2.4, I introduced the simple form

of strong internalism. Since strong internalism can explain many of our intuitions concerning moral hypocrisy, it seems like an appealing starting-point. However, because strong internalism fails to leave room for weakness of will, internalists tend to move towards weaker forms of internalism.

In Section 2.5, I first briefly discussed the unconditional forms of internalism which today are not very popular. After that, I explained in detail many allegedly powerful counterexamples to internalism: amoralists, depressed and listless agents, and the bad people. These counterexamples are supposed to be illustrations of situations where moral judgments fail to motivate and where, as a consequence, motivational internalism turns out to be a flawed view. In response, internalists have formulated a number of different forms of conditional internalism. The conditional forms of internalism try to deal with the relevant counterexamples by arguing that the agents in them fail to satisfy the conditions in which moral judgments motivate. They can also use these specific conditions as a part of their description of their processes through which the moral judgments motivate us.

In Section 2.6, I drew a distinction between direct and deferred forms of internalism. The versions of internalism discussed in Sections 2.4 and 2.5 were direct. According to these views, each individual moral judgment must be connected to motivation at least when the agent satisfies certain conditions. For the sake of accommodating counterexamples such as amoralists, some internalists tend to argue also for deferred versions of internalism. These versions of internalism claim that not all moral judgments need to motivate themselves, but rather, it is enough if they are suitably connected to other moral judgments that are motivating.

In order to be able to accommodate amoralists and the like, some internalists have finally also introduced non-constitutional versions of internalism. In Section 2.7, I outlined how these views recommend that we should focus on the concept 'moral judgment' merely from a semantic perspective. Internalism understood in the resulting way claims only that certain ordinary beliefs can be called moral judgments, but only if the agents, to whom the beliefs belong, also have motivation to act accordingly.

In Section 2.8, I summarized the sixteen resulting versions of internalism that result from the four categories discussed in Sections 2.4-2.7. With the help of the resulting map of the logical space, internalists can better understand what alternatives there are, and which views they should focus on.

Finally, in Section 2.9, I introduced internalism's main competitor, which is of course called externalism. I suggested that, even if externalism can be formulated with one simple principle, which denies all forms of internalism, there are still many different versions of the view that I will discuss in more detail in Chapters 4 and 5. Furthermore, I want to emphasize that the externalists have tended to motivate their views precisely by focusing on the central counterexamples to different forms of internalism, which were discussed in Section 2.5.2. With these examples, the externalists urge us to focus on situations where moral judgments fail to motivate. This means that all the preliminary discussions of this chapter should help us to better understand what is at issue in the debates between internalists and externalists and also what the central differences between different forms of internalism and externalism are.

Chapter 3: The Fetishism Argument

3.1 Introduction

Chapter 2 has outlined different forms of internalism that have been discussed so far, as well as some of the arguments that have been made in their defense. In Chapters 3-5, I will begin to provide a conclusive argument against externalism by defending and developing Michael Smith's fetishism argument. Even if the fetishism argument was first introduced as an argument for a certain form of weak and conditional internalism, with it, Smith actually managed to expose a general problem that applies to externalism. This is that all externalists must explain the reliable connection between our moral judgments and motivation in an external way, and this arguably has implausible consequences.

This is why defending the fetishism argument, understood in this way, is not important because it might provide evidence for a certain specific form of internalism but rather because it can be used to rule out all forms of externalism, or so I will argue in the next three chapters. Of course, in response to Smith's argument, the externalists have made numerous objections to the fetishism argument (in addition to their objections to different forms of internalism). Therefore, it will also be fair to try to evaluate and respond to the externalist objections carefully and in a convincing way. As a consequence, it the defense of the fetishism I will put forward in the next three chapters is successful, I will be able to conclude that all forms of externalism are implausible, and we should thus accept some form of internalism. This conclusion will also pave the way for the last two chapters of this thesis in which I will then explore which specific form of internalism is the most plausible one. Now, let us, however, focus on the current topic, namely the fetishism argument, which I will introduce in this chapter.

The so-called fetishism argument is one of the main arguments against externalism and for internalism.[22] This argument begins from an observation, which can be accepted by both the internalists and the externalists. For example, if a person judges that it is no longer right for her to eat steak, she will usually not be motivated to eat steak after making the new judgment, that is, she will be motivated not to eat steak. This observation suggests that there is a reliable connection existing between our moral judgments and motivation (Section 2.3 and Section 3.3.1).

Michael Smith has argued that, since internalists believe that there is an internal connection between moral judgments and motivation, they are able to use their views about the nature of moral judgments to explain the reliable connection in question. Internalists can explain how our moral judgments enable us to have direct *de re* desires to do the right things (Section 3.3.2). But because the externalists have denied that there is an internal, modal connection between moral judgments and motivation, they need to rely on something else, for example, a *de dicto* desire to do whatever is right to explain the same reliable connection in question (Section 3.4.2). Smith then claims that the internalist account is compatible with our commonsense moral intuitions, whereas moral agents in the externalist framework would care about something that is not primarily important when it comes to moral issues. On this basis, Smith then suggests that the externalist account of how moral judgments motivate would turn a moral agent into a moral fetishist.

[22] Michael Smith's (1994, 1996b and 1997) works contain three different presentations of the argument. Generally, the first version of fetishism argument, which was given in *The Moral Problem* (Smith 1994), has been discussed the most. Most philosophers, be they internalists or externalists, respond to Smith's argument based on how he presented the original version of the fetishism argument. Although Smith's view remains roughly the same, the second version of the argument (Smith 1996b) leads to a weaker form of internalism than the first formulation. The third version of the fetishism argument (Smith 1997) which mainly makes further clarifications, and it could be thought of as a hybrid of the previous two versions. As a result, I will mainly focus on the first two versions of the fetishism argument here.

In Section 3.2, I will start from Smith's observation about a common moral phenomenon that can be accepted by both internalist and externalists. Smith notices that when an agent changes her moral judgment, she usually changes her motivation accordingly. In Section 3.3 and Section 3.4, I will consider the internalist and the externalist explanations of Smith's moral observation. Sections 3.2-3.4 thus constitute the crux of the famous fetishism argument. In Section 3.4, I will also focus on a revised version of the fetishism argument which is arguably based on less controversial premises then the original formulation. At last, in Section 3.5, I will make a summary of my discussion of the fetishism argument.

3.2 Smith's Observation

Internalists and externalists have different views about the connection between moral judgments and motivation. As already mentioned, there exists, however, at least one phenomenon about moral motivation, which should be accepted by both sides (as already mentioned in Section 2.3 and 2.9).[23] As Michael Smith has put it, we can observe 'a change in motivation follows reliably in the wake of a change in moral judgment, at least in the good and strong-willed person' (Smith 1994, 71). This phenomenon is so common that it can be observed in many situations in everyday life. In *The Moral Problem*, Smith provides us with a good example to illustrate this reliable connection between a change in moral judgment and a change in motivation (Smith 1994, 71).

[23] In section 2.5.3, I explained why Smith believes that, in so far as we are rational, we will have motivations that correspond to our moral judgments. Very roughly, this is because, according to Smith, moral judgments are about what our fully rational versions would want us to do in a given situation. In this case, having the motivation that corresponds to your own beliefs about what your ideal version desires you to do is more coherent than lacking it. This is why on Smith's view, insofar as you are rational, you will have the corresponding motivation.

Let us suppose that we, you and me, are engaged in an argument about which party we should vote for. I have already judged that we should vote for the libertarians and thus I am already motivated to vote for the libertarians accordingly. But, during an argument, you convince me that voting for the libertarians is wrong and, instead, I should vote for the social democrats. Perhaps you manage to convince me that the social democrats will better promote the values I thought could be promoted by the libertarians. You might also be able to convince me that the values I thought would be promoted by the libertarians are themselves essentially misunderstood by them. At this point, if I am a good and strong-willed person, what will happen to my motivation, since I have changed my judgment? According to Smith, it is reasonable to think that I will be motivated to vote for the social democrats following my change in the judgment. This consequence is also what we can notice in many other similar situations.

The question then is: how can we explain the reliability with which our motivation changes to match our judgments? As it has been mentioned in Section 2.9 the internalists and the externalists have different views about whether there is an internal, modal connection between moral judgments and motivation. Because of this, both sides offer different kinds of explanations of the previous reliable connection in the example. In the next two sections, I will discuss the resulting internalist explanation and externalist explanations.

3.3 The Internalist Explanation

3.3.1 The Practicality Requirement

As the internalists believe that there is at least some kind of an internal connection (for different alternatives, see Sections 2.4-2.7), between moral judgments and motivation, it could be

suggested that it must be very easy for the internalist to explain the previous observation. To see this, let us consider Smith's (1994, 61) own formulation of a weak and conditional internalism:

> *The Practicality Requirement*: [Necessarily], if an agent judges that it is right for her to φ in circumstances C, then either she is motivated to φ in C or she is practically irrational.

The practicality requirement is a formulation of weak conditional internalism. It is a form of weak internalism because it suggests that an agent who has made a genuine moral judgment need not have overriding motivation to act accordingly—having at least some motivation is sufficient on this view. Thus, when a moral agent suffers from weakness of will, she might be unable to act in accordance with her moral judgment because she has a stronger desire to do something else instead. The practicality requirement, as a form of weak internalism, leaves room for the previous type of weakness of will because that view only requires that, when an agent makes a moral judgment, she has some motivation to act accordingly.[24]

Likewise, the practicality requirement principle puts forward a form of conditional internalism because it suggests that practical rationality is a condition that must be satisfied in order for there to be a reliable connection between moral judgments and motivation. Being a practically rational moral agent here means that the agent in question should have no false beliefs, she should have all the relevant true beliefs, she should have a systematically justifiable set of

[24] That the practicality requirement principle puts forward a weak form of internalism is made more explicit when Smith introduces a revised version of the practicality requirement (see Section 3.5.2).

desires and she should not suffer from any physical or psychological disturbances (Smith 1994, 156-161; 1996a, 160; and see Section 2.5.3 for a more detailed discussion).

Let us then consider how Smith's practicality requirement could be used to explain the reliable connection between moral judgments and motivation. If the practicality requirement were true, internalists would be able to explain the reliable connection between moral judgments and motivation in the following way. We can begin from the observation that, if I am a good and strong-willed rational person and the practicality requirement is true, then when I judge that it is right to vote for the libertarians, I must at least acquire some relevant motivation to vote for the libertarians.[25] It should then be assumed that, after your compelling arguments, I no longer believe that it is right to vote for the libertarians. In this situation, in virtue of the practicality requirement, insofar as I am rational, I will cease to be motivated to vote for the libertarians (see Section 2.5.3 for an explanation of practical rationality conditional internalism). Likewise, your argument will also make me believe that it is right for me to vote for the social democrats after your argument. Since I am a good and strong-willed rational person who does not suffer from any psychological issues such as the weakness of will, I will always be able to be motivated to act in accordance with my moral judgment, assuming that the practicality requirement is true.

[25] The terminology of 'good and strong-willed' has caused some externalists (Brink 1997; Copp 1997; Miller 1996) to misunderstand Smith's original view. Due to this reason, Smith has tried to clarify this terminology in his later responses. The word 'good' indicates a class of people who possess 'the virtue of being disposed to conform their motivations to their moral beliefs in a reliable way, at least absent weakness of will and the like' (Smith 1996b, 177). Similarly, in another paper, Smith suggests that the term of 'good and strong-willed' refers to those who do not suffer from incoherence between belief and desire (Smith 1997, 111). Because of this, we can accept that the term 'good and strong-willed' means at least roughly the same as 'practical rationality' in Smith's works.

All of this means that, as long as the practicality requirement is true, there will always be a reliable connection between moral judgments and motivation in rational people. Nevertheless, we do not merely want to know that the reliable connection exists as the practicality requirement claims, but rather we also want to know why it exists—what is the mechanism that explains how our motivations change when we make new moral judgments. Because of this, the internalists also do need some further account of why a change in motivation follows a change in moral judgments. For the internalists, there are then two kinds of explanations of why something like the practicality requirement would be true. Both of these accounts explain the reliable connection between moral judgments and motivation on the basis of the nature and content of moral judgments.

One theory is that the judgments about what are the right things to do have the power to generate a corresponding reason for action. Thus, I judge that telling the truth is right, this judgment itself will have the power to generate reasons in me. If this were right, then it would not be a surprise that, if a rational agent has a reason to act in a specific way, and her moral judgment reflects those reasons, she will have a motivation to act in that way too (Korsgaard 1986; Smith 1994).

Smith's own version of this kind of a rationalist explanation relies on a certain specific requirement of rationality. This requirement is based on the idea that rational agents are a class of people whose psychology is 'maximally coherent and unified' (Smith 1995, 129). This requirement of rationality towards coherence can be argued to entail that, when an agent has acquired a belief about what the right action is, insofar as she is rational, she will have the relevant desire to act following her belief. This is because, if the agent failed to have the desire

to act in accordance with her belief, it could be pointed out that she would suffer from a sort of incoherence between her belief and desire.[26] This incoherence could then be claimed to make the agent in question irrational. This account would entail that, in the previous example, my judgment that voting for the social democrats is right itself could be thought to have the power to produce a corresponding motivation in me to vote for the social democrats insofar as I am a rational agent disposed towards mental coherence. Hence, we can see why the practicality requirement would be true in this situation.

The second, expressivist explanation of the practicality requirement is based on the claim that judging that doing a certain action is right is itself consists at least in part of being motivated to act accordingly. Expressivists agree with rationalists about the practicality requirement itself, but they do not agree that moral requirements derive from rationality *per se*. Rather, according to expressivists, moral judgments themselves consist of desire-like attitudes rather than beliefs. The desire-like attitudes are then thought to have implications for the explanation of behavior (Blackburn 1984, 1998; Gibbard, 1990, 2003). For example, if I am able to sincerely assert that going to bed early is the right thing for me to do, then these words could be thought to express my plan to go to bed early. It is easy to see why such a view would support the practicality requirement in a similar way. If an agent judges that it is right for her to φ in circumstances C, then according to the expressivists that judgment itself would consist of the agent's desire to φ in circumstances C. Thus, in the previous example, when I judge that it is right to vote for the social democrats, I am in the state of being motivated to vote for the social democrats.[27] This

[26] For an explanation of why this is the case, see Smith (1994, 151-158) and the discussion of practical rationality in Section 2.5.3 and footnote 23 above.

[27] There are also some expressivitsts who think that, even if moral judgments are desires or plans, they are still distinct from motivation. Yet, these expressivists still believe that, due to their disposition towards coherence, rational agents are motivated to act in accordance to their moral judgments. See Section 6.3 below for a discussion.

means that both rationalists and expressivists could be argued to be able to vindicate the practicality requirement and thus explain the reliable connection between moral judgments and motivation in an internalist way.

3.3.2 The *De Re* Desire to Do the Right Thing

That the practicality requirement seems able to provide a reasonable explanation of how our moral judgments affect our motivations is not the only reason to accept it. Michael Smith has argued that we should accept the practicality requirement also because of the way in which the reliable connection between moral judgments and motivation is explained internally. This argument begins from the idea that the previous internalist explanations entail that, when an agent judges that it is right to φ in C, then absent the weakness of will and the like, she will come to have a direct *de re* desire to φ in C. For example, on this view, if I believe that it is right for me to go to bed early in order to keep myself healthy, and I am not suffering from the weakness of will or the like, I would have a direct *de re* desire to go to bed early.

Literally understood, the phrase '*de re*' means 'regarding the thing'. In order to gain a better understanding of *de re*, let us consider this sentence: 'Kalista desires to do what is right' (Dreier 2000, 621). When read *de re*, we can understand this sentence to mean that Kalista desires to do specific things that are right such as helping the poor and the elderly or taking care of children. Ultimately, Kalista is in this case directly moved by the right-making features of actions: these actions, for example, make vulnerable individuals better off. Now, according to the internalist accounts of moral motivation explained above, when Kalista makes judgments about which actions are right for her to do, these judgments produce in Kalista a corresponding, intrinsic desire to perform these actions directly. No further factors or desires are involved in

how Kalista comes to acquire her desires to do the right things. This means that, here, Kalista's desire to do specific right things can be understood as a direct *de re* desire to do the right thing.

As having a *de re* desire to do the right thing derives directly from the judgment about which actions are right, the *de re* desire to do the right thing is a basic, fundamental desire rather than a derivative desire. A moral agent who is thus motivated cares non-derivatively about the right-making features of right actions. According to Smith's rationalist version of internalism, for example, the agent's judgment that to φ in C is right causes in her a non-derivative desire to φ in C because rationality disposes one towards coherence between one's judgments and motivations. Likewise, according to the expressivist version of internalism, the agent's judgment that it is right to φ in C itself at least in part just is a non-derivative desire in her to φ in C. On both views, for instance, when an agent judges that it is right for her to care for the well-being of her family, she will take care of the family's well-being itself. Because of this, the desire to care for the well-being of the agent's family does not derive from any other more basic desire. As Smith has emphasized, being motivated in this direct way is how we normally assume virtuous people are motivated (Smith 1994, 75).

3.4 An Externalist Explanation

3.4.1 The Basic Externalist Theory

As I have already mentioned in Section 2.9, the externalists refuse to accept the idea that an agent's moral judgment must necessarily motive her to act according to her moral judgments. The externalists typically reject internalism on the basis of counterexamples such as amoralists, the depressed and listless agents, and bad people (See Section 2.5.2 above). These are all supposed to be examples in which there cannot be an internal connection between moral

judgments and motivation. Although externalists still believe that 'moral judgments are in some sense invariably action-guiding' (Svavarsdóttir 1999, 162), they claim that moral judgments only contingently motivate moral agents to act accordingly (Lillehammer 1997, 187). On their views, even if a moral judgment can sometimes successfully produce a corresponding motivation in an agent, the motivational force at least in part derives from factors other than the moral judgment itself. Brink claims that these factors might be 'the content of morality…, a…theory of reasons of action, or facts about agents such as their interests or desires' (Brink 1986, 28), whereas Svavarsdóttir's suggests that a conative state, for example, the desire to be moral must be present in together with the relevant moral judgment for producing motivation (Svavarsdóttir 1999 and 2006).[28]

Let us consider Sigrún Svavarsdóttir's example of Virginia and Patrick to get clear about what many externalists think about moral judgments (Svavarsdóttir's 1999, 176; Section 2.5.2 of this thesis). Virginia believes that it is right to help a politically persecuted stranger and she does help him eventually even at the risk of losing her social position. Later, Virginia meets Patrick who could help a politically persecuted stranger in the same way almost at no cost to himself and without any risk. Yet, when faced with the situation, Patrick makes no attempt to help the stranger. During their argument, Virginia tries to persuade Patrick to accept that he should have compassion for victims. She also argues that, considering it is a moral obligation to fight for justice, Patrick needs to help the victim. In the end, Patrick comes to agree with Virginia because of her appeal to compassion and moral obligation. However, even if Patrick understands moral matters clearly, unlike Virginia, he has no inclination or a desire to help the stranger in danger.

[28] David Brink (1989, 1997) has also discussed similar ideas.

We might find it strange and even counter-intuitive that Patrick does not have any inclination to do what he judges as the right thing. The externalists, however, would argue that this consequence is actually quite plausible. According to externalism, after all, Patrick's judgment that helping a politically persecuted stranger is morally right is not enough to produce corresponding motivation in him. Rather, in this situation, he still needs an additional desire in order to be moved to perform his moral duty. On this view, a moral judgment like the one Patrick has made, only motivates when it is accompanied by a desire which has a suitable, related content. Unless such a desire is present, a moral agent can remain unmoved, even if she has already made the relevant moral judgment.

In their description of Patrick's case, the externalists rely on a basic idea that moral judgments do not independently affect motivation. Although the externalists seem to be able to discuss Patrick's case in a way that is in many ways appealing, they will have at least two difficulties when we turn to the previous voting case. Firstly, externalists will find it difficult to explain just why I cease to be motivated to vote for the libertarians when I judge that it would not be right to do so. After all, according to the externalists, even if I judge that it is wrong to vote for the libertarians, I might still be motivated to do so. This is because the judgment that to vote for the libertarians is wrong is not supposed to be able to determine on its own whether I have the corresponding motivation, and moreover, no reference to any other independent desires have been made in this case either.

Second, externalists will find it equally difficult to explain how I acquire new motivation to vote for the social democrats when I judge that it is right to do so. As already explained,

according to them, the connection between my judgment and motivation is merely contingent. In the above-mentioned voting case, I was not explicitly described to have any other additional desire. So, according to the externalists, it is quite possible that, even as a good and strong-willed person, I will still have no motivation to vote for the social democrats even when I have judged that it would be right to do so. However, this consequence clearly contradicts to the expectation that I am motivated to vote for the social democrats in the previous case.

This is why Smith assumes, on behalf of the externalists, at this point that externalists would seek to provide the explanation of why we generally tend to be motivated to act according to our moral judgments by considering more carefully the definition of the 'good and strong-willed' persons. Externalists might argue that, as a good and strong-willed person, I would be disposed to change my motivation when my moral judgments change. This kind of motivational disposition could be argued to make me a good and strong-willed rational person.[29] We should, however, ask: what could the exact content of my disposition be according to the externalists? The content of this disposition cannot, in this case, be the mere tendency towards coherence since this explanation would lead to a necessary connection between moral judgment and motivation and thus to accepting internalism and the practicality requirement. Smith has then suggested that, within the externalist framework, only the desire to do what I believe to be right would be able to explain why I would be disposed to change my motivation to vote for the social democrats when I judge that this is the right thing to do (Smith 1994, 73; Smith 1997, 112).

[29] See Section 5.4 for a discussion of David Copp's similar proposal.

Consequently, we can explain the previous examples in the externalist framework as follows. First, when I judge that it is wrong for me to vote for the libertarians, I lose my desire to vote for the libertarians. This happens because of my desire to do what is right. If I continue to desire to vote for the libertarians, this could at least in principle prevent me from doing so. This means that, as a means for satisfying my desire to do what is right, I must cease to desire to vote for the libertarians when I no longer believe that this is the right thing to do. This loss of my old desire thus results from my new judgment together with my additional desire to do what is right. Second, when I make the new judgement that it is right for me to vote for the social democrats and I have the additional desire to do what is right, I gain a new derivative desire to vote for the social democrats. It makes sense for me to acquire this new desire because by having it I am more likely to satisfy my more basic desire to do what is right. This new desire thus derives from the more fundamental desire to do what I believe is right and my judgment that the action in question is the right thing to do.

We can thus conclude that, according to Smith, externalists must think that, if an agent judges that it is right for her to φ in C and she has a desire to do what is right, then the agent will be motivated to φ in C. The agent's desire to φ in C is, on this view, a derivative desire because it derives from a moral agent's judgment and her non-derivative desire to do what she believes to be right. Thus, according to this externalist proposal, only when a moral judgment is suitably connected to a non-derivative motive to do whatever is right, a moral judgment can produce a corresponding desire. Furthermore, this motivation will be a derivative desire to do what is right. It thus might at first seem that externalists too are able to explain the reliable connection between moral judgments and motivation. But should we accept their explanation? And more importantly, is it as good as the internalist one?

3.4.2 The *De Dicto* Desire to Do Whatever Is Right

According to Smith, even though the previous externalist explanation of how moral judgments affect our motivations in a reliable way can explain the reliability of the connection, this kind of an explanation would still be unacceptable. He has argued that the way in which moral agents would be motivated externally according to the previous externalist account gives us sufficient reason to reject externalism. Let us see why Smith thinks that this is the case.

According to the previous externalist view, virtuous people are ultimately motivated by their non-derivative desires to do what is right. This non-derivative desire is also a *de dicto* desire to do what is right. Literally, the phrase '*de dicto*' means 'about what is said'. To understand why this desire to do what is right is a *de dicto* desire, let us return to the example we already discussed earlier: 'Kalista desires to do what is right' (Dreier 2000, 621 and Section 3.2.2 of this thesis). When we understand this sentence in the *de dicto* way, we think that Kalista has an abstract desire to do whatever she happens to think is right, under that description as the right thing to do. Because of this desire, Kalista may desire to help the poor and the elderly or to take care of children. Nevertheless, it is not because these actions are themselves right that moves Kalista to do these things—it is not the right-making features of these actions that she cares about directly. Rather, the reason why Kalista chooses to do these things is that she has a *de dicto* desire to do whatever is right and those actions just happen to be exactly the things she believes to be right. The desires to do specific right things such as helping the poor thus here derive from the fundamental desire which is to do whatever is right. Kalista can hold this *de dicto* desire to do whatever is right even if she has no idea about what the right thing is.

According to Smith, if we use the *de dicto* desire to do whatever is right to explain the reliable connection between moral judgments and motivation (as he thinks externalists must do), we will find it difficult to explain the virtuous people's behavior in a way that would match our intuitions. It seems that, within this externalist framework, if a moral agent chooses to be honest or to help her friends and family, the right-making features of these honest and caring actions would not be the primary source of the agent's motivation. Instead, the desire to be honest and help one's friends and family would on this view derive from the good person's more fundamental desire, which is the *de dicto* desire to do whatever is right. Smith (1994, 75; 1997, 113) has pointed out that, this explanation of why virtuous people would typically be motivated to act morally is counterintuitive. We would, of course, hope that good people care non-derivatively about honesty and the well-being of their friends and family. When asking why their moral concerns change because of their moral judgments, we would not expect them to answer: 'although I have no inclination to do these actions in themselves, I really want to do what is right'. We normally think that being motivated merely by the *de dicto* desire to do whatever is right would be too cold and inhumane—it is not the way a caring moral person would be motivated. It thus appears that people who are motivated by the relevant *de dicto* desire would actually care about something that is not primarily important in morality. Thus, it can be argued that, if an agent is motivated by the *de dicto* desire to do what she believes to be right externally, she has a moral fetish or a vice.

An example from Bernard Williams (1981) might help us to illustrate why relying on being motivated by the *de dicto* desire to do whatever is right could be claimed to be the wrong way to explain the reliable connection between moral judgments and motivation. Williams asks us to consider a man who chooses to save his wife instead of a stranger in a dangerous situation.

Some philosophers believe that, even in this sort of a situation, a virtuous person should be motivated without partiality. A moral agent should, in this situation, be moved by the moral principle according to which it is morally permissible to save one's wife together with the fact that the woman in peril is his wife. Nevertheless, Williams objects to this view. He suggests that we should consider the situation from the wife's perspective. It is, of course, normal for the wife to hope that the whole motivation of her husband would be 'I am saving my wife'. This kind of an intuition shows that we expect that good people would be moved to act directly by his or her love. If any additional motivation, for instance, in the form of the thought that saving one's wife is morally permissible were required, the moral agent would be treating his wife as a stranger and thus he would be alienated from his wife. In this case, requiring the good person to have a *de dicto* desire to do whatever is right provides him with one thought too many (Williams 1981, 17-18). Similarly, externalists could be argued to be making the same mistake in explaining the reliable connection between moral judgment and motivation with the *de dicto* desire to do whatever is right[30].

[30] Carbonell (2013) has three objections to the use of Williams's idea of 'one thought too many' in Smith's context. First, Smith thinks that when the husband decides to save his wife, he thinks that 'it is my wife [and I have a desire to save my wife]' and it is morally permissible to save my wife'. According to Carbonell (2013, 465), if the cases are analogical, then we should think that the former corresponds to the husband's *de re* desire to save his wife and the latter to his *de dicto* desire to do whatever is right. In this situation, a moral agent would be motivated by both his *de re* desire to do the right thing and the *de dicto* desire to do whatever is right.

Secondly, according to Carbonell, Smith seems to assign moral agents in both cases only one motivating desire—the relevant *de dicto* desire—in the externalist framework. Based on the assumption of a single motivating desire, Smith argues that the relevant *de dicto* desire is 'one thought too many'. As a consequence, Carbonell claims that Smith does not allow the externalists to employ any motivating desires other than the *de dicto* desire to do whatever is right in their explanations of the relevant cases.

Lastly, on Carbonell's view, Smith still confuses different kinds of phenomena when he discusses Williams' example in order to support the fetishism argument. According to Carbonell, the initial phenomenon for Smith is the reliable connection between a change in one's moral judgment and a corresponding change in one's motivation. But, in Williams's example, the man does not change his mind to save his wife. As Smith's cases are more complicated than Williams's due to the changes in moral judgment, the additional desire which is *de dicto* would not be 'one thought too many' in Smith's case.

3.5 A Revised Version of the Fetishism Argument

The fetishism argument in *The Moral Problem* (Smith, 1994) which I have just explained, was the first version of the argument. In order to respond to many of the externalist objections to internalism, Smith has discussed the fetishism argument also later on and he has also clarified the argument in many important respects (Smith 1996b). One main difference between the first and the second version of the fetishism argument is that Smith introduced a new concept of 'a moralist' in the second version of the argument to clarify his view. By relying on this concept in the second argument, Smith defends a weaker version of internalism, which is also called 'Weaker Moralist Internalism' in order to distinguish itself from other forms of internalism.

The second formulation of the fetishism argument too starts from the observation that moral judgments usually motivate us. Smith then takes for granted that the externalist definition of a person who is an amoralist is not problematic in any respect. This means that he can therefore stipulate that moralists are, by definition, those people who are not amoralists: they are the contrast class of people who are motivated by their judgments. As I will explain below in Section 3.5.2, Smith first uses the notion of 'moralists' to formulate a principle which he calls 'Weak Moralist Internalism'. He then uses this principle to make two arguments against externalism, both of which I will explain in detail below. Here too, according to the externalist explanations, the moralists' primary source of motivation to comply with his or her judgments

---

In response to Carbonell's first two objections, I will argue that being motivated by both the *de re* desires to do the right things and the *de dicto* desire to do whatever is right would not make an agent any less fetishistic (see the response to co-presence objection in Section 4.2). In response to Carbonell's last objection, I grant that Smith's and Williams's cases are different and that Smith's case seems to be more complicated. However, what Smith actually argues is that, in the externalist framework, moral agents will have the additional desire which is 'one thought too many' even in the more complicated cases.

would have to be a desire to do what is right. According to Smith, such an explanation not only gives us an inappropriate description of the moralists' psychology but rather it also commits us to an implausible conception of moral perfection. To figure out this point, let us first consider an example of the reliable connection between moral judgments and motivation.

3.5.1 An Example of Reliability

In his more recent discussion of the reliable connection between moral judgments and motivation, Smith introduces a new example in which one of his friends changes his view about utilitarianism. At first, this person (let us call him Mike) is a utilitarian. Mike strictly complies with utilitarianism and he believes that it is always right for him to maximize the total amount of happiness (and also minimize the overall extent of suffering). Basically, Mike does what he thinks to be right due to his considerations that these those actions have the following right-making features: they maximize happiness and minimize suffering. After a few years, however, Mike changes his mind because he considers different objections to utilitarianism. So, he now believes that it is right to care more about his friends and family, and sometimes to give them extra help, even when doing so is not the ideal utilitarian option. We can also assume that Mike is a moralist and changes in his moral motivation follow reliably changes in his moral judgments. According to Smith, in this situation, it seems that 'what he has acquired are new non-instrumental personal concerns, whereas before he had only one concern: a non-instrumental impersonal concern' (Smith 1996b, 181). The question then again is how are we going to explain the reliable connection between Mike's judgments and his motivations?

3.5.2 Weak Moralist Internalism

For the sake of an argument, Smith first accepts the externalists' definition of 'amoralists'. According to this definition, an amoralist is a person who recognizes the existence of moral considerations and yet remains unmoved by them (Brink 1986, 30). Externalists believe that there are amoralists (or at least there could be) and so they think that there are people who have no motivation to act according to their moral judgments, Smith then points out that we can define a concept of 'moralists' in the same way in which the externalists define 'amoralists'. A 'moralist' is, by definition, a person who is motivated to act according to his or her moral judgments, at least absent weakness of will and other psychological failures. Although moralists might sometimes have false beliefs, they still possess an executive virtue that amoralists simply do not have: the virtue of being disposed to conform their motivation to their moral beliefs in a reliable way, at least absent weakness of will and the like.

In order to distinguish it from the internalist position defended in *The Moral Problem*, Smith then calls the following view 'Weak Moralist Internalism':

> *Weak Moralist Internalism*: [Necessarily], if an agent judges it right to φ in C, and that agent is a moralist, then she is motivated to φ in C, at least absent weakness of will and the like (Smith 1996b, 176).[31]

[31] As I mentioned in Section 3.3.1above, the practicality requirement is a form of weak internalism. Although there exist some minor differences between the formulations, 'The Practicality Requirement' expresses exactly the same thought as 'Weak Internalism' does and thus, it is equivalent to 'Weak Internalism' in this text. As a formulation that appears in Smith's second version of the fetishism argument, 'Weak Moralist Internalism' also expresses almost the same thought as 'The Practicality Requirement' in Smith's first version of the fetishism argument.

As the definition of moralists is created by contrasting the moralists with the amoralists whose existence is endorsed by the externalists, the externalists should therefore acknowledge that the moralists exist too. Because the Weak Moralist Internalism is a view that is based on the concept of moralists to which the externalists are thus committed, according to Smith, externalists too are also committed to Weak Moralist Internalism.

Let me illustrate this idea by returning to the previous case of Mike. When Mike is a utilitarian, he judges that maximizing happiness and minimizing suffering is right. Let us also assume that, Mike is also motivated to do exactly that and also, at this point, Mike's motivation to maximize happiness and to minimize suffering is non-instrumental, which means that this motivation does not serve to fulfill any other motivations Mike might have. Similarly, when Mike comes to believe that it is sometimes right to give his friends and family additional help even when doing so is not utilitarian, Mike is still a moralist and he has a motivation to act in accordance with his judgment. Mike's motivation to give extra help to his friends and family is his final end, which does not aim to satisfy some other, further motivations he might have. Thus, we can see that there is a reliable connection between Mike's judgments and his desires to act accordingly.

It seems clear that *weak moralist internalism* captures the observation that a change in motivation follows a change in moral judgments at least when you are a moralist. By defending *weak moralist internalism*, internalists are then trying to demonstrate that, if a moral agent is not irrational (she is not weak-willed or anything alike), she will be motivated to φ in C as a consequence of two factors: one is that the agent judges it is right to φ in C in question and the other factor is that the agent in question is a moralist.

Again, there are two possible explanations of why this previous principle might be true (see Section 3.3.1 above). According to the rationalist explanations, the judgment that it is right to φ in C is a certain moral belief. If an agent is rational which means that the agent has a disposition towards psychological coherence, she will have a non-instrumental desire to φ in C because, according to Smith's view, it would be less coherent to have the previous belief and lack that desire. This is why the agent would suffer from at least some kind of irrationality (such as weakness of will) if she didn't come to acquire the direct desire compatible with her moral judgment. This desire would also be a non-instrumental desire because it would not serve any further desire. In contrast, according to the expressivist alternative, the judgment that it is right to φ in C is itself at least in part a non-instrumental desire to φ in C, or at least a more general plan such that rationality as coherence would require the agent to have a non-instrumental desire to φ in C insofar as she is not weak-willed.

3.5.3 The Externalist Account and Objections

As before, the externalists too would need to provide us with an explanation of the previous reliable connection between moral judgments and motivation. As we already saw above, externalists believe that amoralists either exist or at least they could exist. Therefore, in the externalist framework too, the difference between moralists and amoralists must be able to explain why the moralists are reliably motivated by their judgments and the amoralists are not. According to Smith, externalists must assume at this point that not only 'the nature of the judgment that the moralist makes', but also 'the nature of the moralist herself' contributes to the fact that only the moralists are motivated (Smith 1996b, 179). This is because, as already explained, Smith thinks that, in the externalist framework, the moralists would have to happen to desire to do what they think is right, whereas, in contrast, the amoralists happen to lack this

desire. Thus, externalists too can accept that, if a moralist believes it is right to φ in C and she desires to do what is right, then she desires to φ in C or she is weak-willed. The desire to φ in C in question would be here an instrumental desire that derives from the non-instrumental desire to do what is right and the relevant moral judgments. Smith has two objections to this externalist explanation of why changes in moral motivations tend to follow changes in moral judgments. I will elucidate these objections next through discussing the previous example of Mike.

The first objection to externalism is based on the idea that the externalists' core claim—that a moralist's desire to φ in C derives from a non-instrumental desire to do what is right—will lead to an objectionable description of the moralist's psychology. Smith argues that the externalist description of Mike's moral conversion would be entirely driven by the externalists' prior commitment to the externalist framework itself. This is because, according to Smith, the previous externalist description of the case does not fit our ordinary understanding of our agent's psychological processes during moral conversions.

According to the previous externalist description, Mike's desire to maximize happiness and to minimize suffering and his desire to give extra help to his friends and family are both in the end merely instrumental desires. These desires derive from Mike's only non-instrumental desire, which is his *de dicto* desire to do what is right. Furthermore, we must keep in mind that Mike must have this non-instrumental desire as having this desire is what makes Mike a moralist (rather than an amoralist). If Mike desires to do what is right and he thinks that it is right for him to maximize happiness and to minimize suffering, he would desire instrumentally to maximize happiness and to minimize suffering—he would desire to do so as a means to doing what is right. Likewise, if Mike desires to do what is right and he believes that it is right to give

additional help to his friends and family, he would desire instrumentally to do so– again, as a means to do what is right.

But intuitively, we would rather think that Mike cares about right things directly, instead of merely the abstract moral rightness, whatever that happens to be. We would hope that Mike would be motivated directly by right-making features of the actions which he takes to be right. Hence, it seems that the only reason why we would be inclined to accept the previous counter-intuitive description of Mike's moral conversion would be our prior commitment to the externalism itself. This is why Smith takes the previous description of Mike's mental conversion to be entirely theory-driven. More importantly, it is evident that the externalist description of Mike's moral conversion does not seem to fit our ordinary understanding of virtuous people's psychology and thus, we should reject the externalist account for Mike's moral conversion.

Smith's second objection to externalism is based on the idea that the externalists' core claim—that a moralist's desire to φ in C derives from a non-instrumental desire to do what is right—also commits the externalists to an implausible conception of moral perfection. According to Smith, externalists claim that the primary source of the moralists' motivation is the non-instrumental desire to do what is right. This means that, even if Mike believes that it is right for him to help his friends and family, he does not desire to do so non-instrumentally. As we have just seen, Mike's desire to help his friends and family is merely instrumental because it is derived from the non-instrumental desire to do what is right together with Mike's belief that it is right for him to help his friends and family.

Yet, Smith argues that this picture of morally perfect people is not plausible. We would usually assume that, if Mike is a morally good person, he is directly moved by the fact that his actions would serve the well-being of his family and friends. Here we can explain what motivates Mike without referring to any other desire, by relying on right-making features of his actions.[32] However, according to the externalists, Mike would be moved only by the fact that his actions are right, which means Mike would be motivated by the fact that certain features of his actions are right-making features. Mike would thus appear to care about the moral status of his actions much more than the qualities of his actions that are responsible for that status.

The previous two additional features of arguments by Smith support the idea that the externalist explanation of how Mike is motivated is not very plausible. The externalist account thus seems to entail that morally perfect people would care about something that is not primarily important in morality. Morally perfect people, who would be motivated in the same way as Mike is according to externalism, would intuitively turn out to be morally imperfect after all. On their view, a moral fetish or a moral vice would seem to acquire that status of the only moral virtue. This awkward consequence of externalism is what Smith's second objection objects to (Smith 1996b, 183).

[32] To see what Smith means by 'right-making features' and 'the feature of being a right-making feature', let us consider an example. Suppose that you are walking on a road and pass by a person who falls from his bike and hurts himself. You would probably try to help him by pulling him up or calling an ambulance if he has broken his legs. Here, if you are motivated by the situation where the person needs some help, it appears that you are motivated by the 'right-making features' of the right action in Smith's sense. However, if you were to accept the externalism, then you would not be motivated by the right-making features, which here seems to be the fact that the other person needs help. Rather, you would be moved by the thought that the fact that the other person needs some help is a right-making feature, which is 'the feature of being a right-making feature'.

3.6 A Summary of the Fetishism Argument

In this chapter, I have discussed in detail two versions of the fetishism argument. The fetishism argument starts from Smith's observation that an agent's motivations tend to change when there is a change in her moral judgments. According to Smith, the internalists and the externalists will provide different kinds of explanations for the previous reliable connection between moral judgments and motivation. The internalists believe that an agent will have the *de re* desires to do the things that are right as a result of judging which actions are right. In contrast, according to Smith, the externalists must believe that an agent's motivations tend to conform to her moral judgments only because the agent has the *de dicto* desire to do what she judges to be right. Smith's arguments then seem to support the idea that the internalist explanation of the previous reliable connection fits our moral intuitions better than the externalist explanations. We can then summarize the fetishism argument for internalism in the following way:

(1) A change in motivation follows reliably judging that an action is right, at least absent weakness of will and the like.

(2) We can account for how a moral agent is motivated by his judgment either in an internalist or an externalist way.

(3) Internalists are in a position to provide an explanation of the previous reliable connection, which matches our understanding of morality. According to internalism, there is an internal, modal connection between moral judgments and motivation such that a moral judgment naturally leads to the agent having a *de re* desire to act accordingly;

(4) The only explanation of the reliable connection between moral judgments and motivation the externalists could offer is that judging an action to be right together with a *de dicto* desire to do the right thing can produce new motivations in the agent.

(5) We usually assume that moral people are motivated by the right-making features of actions, whereas the externalists are forced to claim that moral people are motivated by rightness itself. Similarly, we usually believe that moral people have *de re* desires to do the right thing, whereas the externalists can only think that moral people have a *de dicto* desire to do what is right.

(6) According to our intuitions, an agent who is motivated by the *de dicto* desire to do whatever is right would have something akin to a fetish.

(7) Therefore, good moral agents cannot be motivated in the way that the externalists claim they are. Such agents thus have to be viewed as being motivated in the way described by the internalists.

In this chapter, I have already explained Smith's fetishism argument in detail. In the next two chapters, I will move on to defend this argument by responding to the most forceful objections to it. Firstly, in Chapter 4, I will critically evaluate the externalists' attempts to defend the externalist explanations of the reliable changes between moral judgments and motivation based on the relevant *de dicto* desire. Then in chapter 5, I will argue that the externalists will not be able to avoid the fetishism argument by relying on other types of externalist explanations of how our motivations tend to track our moral judgments.

Chapter 4: The Externalist Defenses of the *De Dicto* Desire

4.1 Introduction

Chapter 3 introduced the fetishism argument. It argues that, if an agent is motivated by the *de dicto* desire to do whatever is right, she will be thought of as a moral fetishist. In this situation, many externalists reject Smith's fetishism charge by explaining that it is acceptable or even commendable to have the *de dicto* desire to do whatever is right during the process of acquiring motivation. Thus, some externalists think that the relevant *de dicto* desire to do what is right does not amount to a moral fetish (Lillehammer 1997), but rather it can even be a part of the motivational structure of the good and strong-willed persons who have both the relevant *de dicto* and *de re* desires at the same time (Copp 1995 and 1997; Svavarsdóttir 1999). Some externalists, furthermore, believe that the relevant *de dicto* desires might even be preferable to the *de re* desires because the *de dicto* desires would be able to restrict unreasonable direct desires (Lillehammer 1997; Shafer-Landau 1998 and 2003). In this chapter, I will explain and respond to these types of externalist objections to the fetishism argument. I will first argue that, if the *de dicto* desire constitutes even a part of an agent's motivational structure, the agent should be deemed as a moral fetishist. I will also argue that, even if the externalists manage to escape the fetishism concerns in the previous ways, their views would have other implausible consequences.

In Section 4.2, I will explain the externalists' claim that being motivated by the *de dicto* desire to do whatever is right is not objectionable. Many externalists claim that this is because good and strong-willed people could have both the relevant *de re* desires to do right things and the relevant *de dicto* desire to do whatever is right at the same time. In response, I will argue that an agent would still be a moral fetishist because she would be in part motivated by the *de dicto*

desire to do whatever is right. If this is right, then the externalists' claim that good and strong-willed people have both different kinds of desires cannot save externalism from the fetishism argument.

In Section 4.3, I will consider two further externalists defenses of the view that an agent can have the *de dicto* desire to do whatever is right. The first claim is that the *de dicto* desire to do whatever is right needs to be invoked only in certain rare cases. Some externalists suggest that we can usually explain the reliable connection between moral judgments and motivation with an agent's standing, pre-existing *de re* desires. Because of this, only in rare cases we need to invoke the *de dicto* desire to do whatever is right to explain what motivates the agent. In response to this view, I will argue that not only the externalists' explanation of moral motivation based on standing *de re* desires is not plausible, but also the *de dicto* desire to do whatever is right cannot ground good reasons for an agent's actions.

The second externalist claim discussed in Section 4.3 is Lillehammer's idea that there are some cases where the *de dicto* desire to do whatever is right is needed. The claim here is that agents need the *de dicto* desire to do what is right as this desire will help them to resist temptations in cases in which they would otherwise act wrongly. In response to this claim, I will argue that here the externalists have moved to discuss a different question than the original one put forward by Smith. Smith's question which was discussed throughout the last chapter seems to be: how can we explain the reliable connection between moral judgments and motivation? In contrast, Lillehammer's new question could be formulated as: how are we going to explain the reliable connection between moral judgments and sufficiently strong motivation?

What Lillehammer implies is that moral judgments, as understood by the internalist, would not always be able to generate sufficiently strong motivation in moral agents sometime. Although our answers to the new question will not affect what we should think about the fetishism argument itself and thus the internalists do not need to provide such answers, I will still try to argue that the previous question can be better answered via relying on the *de re* desires to do the things that are right rather than by relying on the *de dicto* desire to do whatever is right.

Recall that, according to Smith, the externalists are also committed to an implausible theory-driven account of moral motivation and a flawed view of moral perfection. In Section 4.4, I will discuss the externalists' objections to Smith's revised version of the fetishism argument. The externalists argue that Smith's second version of the fetishism argument mistakenly attempts to commit them to an implausible theory-driven description of an agent's psychology. Svavarsdóttir tries to describe the agent's mental conversion in a different, common sense way, which she thinks is compatible with externalism. She also claims that Smith's concept of 'moral perfection' is inappropriate as Smith's understanding of moral perfection seems to apply only to agents who do not need to rely on any moral reflection. In Section 4.4, I will discuss these responses to the revised fetishism argument and, furthermore, explain why these objections fail.

If my arguments against the externalist defences of the *de dicto* desire to do whatever is right are successful, then we have good reasons to reject the relevant *de dicto* desire when we try to explain the reliable connection between moral judgments and motivation. We also have persuasive reasons to agree with the fetishism argument, at least unless the externalists are able to provide other types of more plausible explanations of the previous same reliable connection.

This will be topic of the next chapter. But for now, let us critically consider the externalists' first set of objections.

4.2 The Co-presence Objection

4.2.1 The Objection

In the fetishism argument, Smith claims that the externalists can only account for why an agent's motivations change after a new moral judgment by referring to a desire to do whatever is right, which is read *de dicto*, and not *de re* (see Section 3.4). According to him, being motivated by the *de dicto* desire to do what is right indicates that an agent does not have direct concerns for the things that really matter. So, the cost of the externalist explanation is that it would turn morally good people into moral fetishists.

The externalist objection to the fetishistic argument, which I will consider here, is that the relevant *de dicto* desire will not be a problem as Smith argues. Arguably, it is evident that, in addition to the *de dicto* desire to do whatever is right, we can also find different relevant *de re* desires from the psychological make-ups of the good and strong-willed people. Hence, it might be a feature of the morally good people that they have a wide range of both the relevant *de re* desires and the relevant *de dicto* desire. Let us consider David Copp's (1995, 212) objection as an example of a response of this type.

Imagine that Dena is a good person who cares about the well-being, fair treatment and mental health of her friends and family. Dena can thus be thought of as having the previous *de re* desires, which also motivate her to behave morally. Let us then further imagine that, Dena gains an additional *de dicto* desire to do whatever is right whilst at the same time keeping the previous

de re desires. The only difference comes about in this situation is that Dena begins to have an additional desire to do whatever is morally right. Copp thinks that, in this case, there is no reason for us to regard Dena as someone who has a moral fetish, as Smith would seem to suggest. Rather, Copp suggests that as Dena continues to have and be influenced by her *de re* desires, the additional *de dicto* desire should make no difference at all to how well we think of her as a moral agent.

In fact, following the previous line of thought, most externalists would agree with Smith that caring only about doing what is thought to be morally right would not be appropriate for a morally good person. Most externalists would also grant that, in order to be counted as a morally good person, an agent should have various direct moral concerns: he or she must directly care about honesty, kindness, loyalty, and other people. For example, Svavarsdóttir (1999, 198) still believes that an agent's characteristics direct moral cares and concerns should be assumed to be stable since they 'do not involve motivational dispositions that are engaged by distinctively moral representations of one's circumstances or behavioral alternatives'. Usually, for a moral person, the moral actions themselves are enough to yield 'comfort, relief, or encouragement', and the result of this is that the moral agent will undertake those actions due to his or her direct concerns (Svavarsdóttir 1999, 199).

What most externalists disagree with Smith about is whether they are committed to explaining the reliable connection between moral judgments and motivation with and only with the relevant *de dicto* desire. A moral fetishist could be argued to be someone whose only non-derivative desire is to do whatever is morally right. Such a person would not have any direct concerns for things such as honesty, kindness, loyalty, and compassion. In fact, all her desires

for performing moral actions would derive from her desire for doing whatever is morally right. However, according to the externalists, good and strong-willed people are not like this: they also have other direct concerns in addition to the relevant de dicto desire. The relevant *de dicto* desire could, after all, co-exist with many other direct *de re* desires in the agent's psychological make-up at the same time. Many externalists thus claim that explaining the reliable connection between moral judgments and motivation by referring to the relevant *de dicto* desire 'only commits them to maintaining the desire to be moral is *a part of* the motivational structure of the good person' (Svavarsdóttir 1999, 199).

4.2.2 The Response (1)

According to Smith's understanding of the term 'fetish', a moral fetishist is someone whose only non-derivative desire is the *de dicto* desire to do whatever is right and, also, whose derivative desires to do the right things are based on that *de dicto* desire (Section 3.4.2). Yet, as we noticed in the last section, some externalists argue that having the *de dicto* desire to do whatever is right would not be sufficient to make a moral agent a moral fetishist because they claim that morally good people can have both *de re* desires to do things that are right and the *de dicto* desire to do whatever is right under that description at the same time. Instead of rejecting Smith's definition of 'a fetish', the externalists thus deny that moral agents with both kind of the relevant desires would be moral fetishists even according to Smith's own definition of a moral fetish.

In Sections 4.2.2 and 4.2.3, I will provide an internalist response to this objection. I will argue that, even if ordinary good and strong-willed people had both the relevant *de re* desires and the *de dicto* desire to do whatever is right as the externalists believe, these people could still be

claimed to be moral fetishists. In section 4.2.2, I will first introduce and illustrate a slightly different, and yet still equally plausible definition of a 'fetish', which allows us to understand what constitutes fetishism more generally, beyond the context in which the *de dicto* and *de re* desires are discussed. The new definition should thus be more acceptable for both the internalists and the externalists. Then, in Section 4.2.3, I will explain why externalism could still be argued to turn ordinary moral agents into moral fetishists, given the fact that an action is right cannot be a plausible moral reason for an action.

Here, I want to begin from a definition of a 'fetish', which was first introduced by R. Jay Wallace. According to him, fetishism in this context should be understood as 'the investment of interest and attention in objects that are not intrinsically worthy of such responses' (Wallace 2006, 195). This definition suggests that something is a fetish for an agent if the agent treats the object as more valuable than the object really is.

To illustrate why the previous general definition of fetishism is plausible, let us consider shoe fetishism. Consider an example. Bob is a huge fan of shoes and one of his most favourite things is to collect different kinds of shoes. Many of the shoes in Bob's collection are limited editions that are worth a lot of money and all of them are fashionable, high-quality shoes. Even a person who does not consider shoes a lot would be impressed by Bob's collection. One slightly odd thing about Bob, however, is that, when being asked why he loves collecting different types of shoes so much, he cannot give a plausible reason for his hobby. Bob explains that he treats the shoes in the collection almost as a companion and he even develops an erotic interest in some of them.

The reason why we call Bob a shoe fetishist then is that he devotes attachment, love, and sexual interest in objects that do not deserve such responses. Usually, we believe that only humans—rather than shoes—deserve such a great devotion. Nonetheless, Bob chooses to love and to have an intimate relationship with his shoes instead of humans, which makes him a shoe fetishist. As Wallace's definition puts it, he is investing a certain kind of attention to objects that do not deserve it. Yet, notice that, in this case, we would continue to regard Bob as a shoe fetishist merely on the basis that he is trying to have an intimate relationship with his shoes. Even if at the same time, Bob also cared about other things, this fact would not make Bob any less of a fetishist.

Here, it is then essential to notice that Wallace's suggestion of what fetishism amounts to makes it irrelevant whether a moral agent has both the *de dicto* desire to do whatever is right and the relevant *de re* desires to do the right things at the same time. According to my reconstruction, the crucial point of the fetishism argument is not whether virtuous people act out of their direct concerns (such as their concerns for honesty and the well-being of their family). Wallace's definition of fetishism enables us to see that the decisive aspect of the fetishism argument is whether an agent cares about the property of rightness, which some actions happen to have. If an agent who has both types of desires counted as a moral fetishist according to Wallace's definition merely in virtue of having the *de dicto* desire to do whatever is right, she would not be any less fetishistic because of her additional *de re* desires.

One advantage of this new definition is that it is more general than Smith's—the introduced definition applies to both moral and other fetishists. With the help of Wallace's definition, we can not only explain why externalism would make moral agents moral fetishists, but individuals

such as Bob count as shoe fetishists. As Wallace's definition is more general than Smith's, the externalists no longer need to explain the reliable connection between moral judgments and motivation by relying solely on the relevant *de dicto* desire to do whatever is right. Because of this, we can grant the externalists that good and strong-willed people have both the relevant *de re* desires and the relevant *de dicto* desire as the externalists argue. Thus, everyone in the debate, including the externalists, should be able to accept Wallace's definition of a fetish.

Additionally, the new definition is also more explanatory than Smith's. According to Smith, the reason why moral agents in the externalist framework could be argued to be moral fetishists is that these agents are motivated by the *de dicto* desire to do whatever is right in a counter-intuitive way. This criticism that is based on our moral intuitions is more descriptive than explanatory. It does not say too much about why we should think the externalist account to be counterintuitive. This gives the externalists an opportunity to claim that their explanations of the reliable connection between moral judgments and motivation are actually just as plausible as the internalist's. In contrast, our new definition tells us more about the reason why the externalist account of how moral agents are motivated is fetishistic and thus, Wallace's definition explains more why the externalist explanation of moral motivation is counterintuitive. When a moral agent is motivated by the *de dicto* desire to do whatever is right, he cares too much about the rightness itself, or so I will argue in the next sub-section. That is, I will argue next that, when an agent cares too much about the rightness itself, he invests attention in objects that do not deserve it and thus becomes a moral fetishist.

4.2.3 The Response (2)

In this section, I will argue that, if an agent has the *de dicto* desire to do whatever happens to be right (under that description), she cares directly about something that is not intrinsically worth caring about. So, in the light of the new definition of fetishism discussed in the last section, having the *de dicto* desire to do whatever is right can be argued to be fetishistic. Thus, even if the externalists were able to argue that the *de re* desires to do right things co-exist with the *de dicto* desire to do what is right among good and strong-willed people, this argument would not be of any help to the externalists.

Let us then consider a very fundamental, intuitive view of what happens when we do moral deliberation. When we decide to act, at least at that moment, we believe that our investment and attention is needed for some good reason. We therefore usually assume that the things that motivate us to act are intrinsically worthwhile and thereby such that they give us practical reasons to act in the considered way. Even the externalists cannot deny these simple observations as they too would have to grant that acting for the sake of things that you do not take to be intrinsically worthwhile is unusual and odd.

As the externalists always claim that an agent should at least in part be led by the *de dicto* desire to do whatever is right, an agent who has this desire should be assumed to care about the rightness of actions itself in some abstract sense. The agent should take rightness itself to be reason-providing and thus something that is intrinsically worthwhile to devote time and energy for. Yet, unfortunately, at this point, the externalist view becomes less plausible. If the rightness of actions itself were reason-providing, then it could be argued that a virtuous person who acts

for the right reasons would need to often choose to do certain actions because those actions are right. The problem is that the previous claim just does not seem to be acceptable.

To see why the rightness of actions itself is not reason-providing, we can consider the following basic idea about the reasons behind moral agents' actions. Most philosophers accept the following claim: the fact that certain actions are right is the same as the fact that properties of those actions provide us with good reasons to undertake those exact actions (Dancy 2000; Stratton-Lake 2002; Suikkanen 2005). Thus, 'the reason why a good-willed person does an action, and the reason why the action is right, are the same' (Korsgaard 1996, 60; Stratton-Lake 2000, 16). The previous claim is based on the very natural thought that we are able to find out the reason why a given action is right by finding out why a virtuous person would choose to do that action. The reason that makes an action right should be identical to the reason that counts in favor of an agent performing that action. For example, when a kind and warm virtuous person helps a stranger in need, we will think that the person does a commendable action. If we then know why the virtuous person helps the person in need, then we will also know at the same time why the action in question is right.

If the previous compelling idea is true, then what the externalists need here—that the rightness of actions itself is reason-providing—leads to an awkward result. Let us begin from the previous simple observation that a virtuous agent usually does a given action because she thinks that there are some good reasons for her to act in that way. If the externalists were right, then one of these reasons that would convince a virtuous person to act in a given case would have to be the fact that the act in question is right. Given the previous plausible thesis that the reasons why actions are right and why virtuous agents do them go hand-in-hand, it would in this situation

follow that the fact that a given action is right would be one of the qualities of the action that would make it right. Yet, the fact that some action is right cannot make the action right—the right-making features of an action must be something different, and more basic than the rightness of the action itself. They must be considerations that explain why the action in question is right (for example, because the action saves lives or does not harm anyone), whereas the rightness of the action itself cannot provide such an explanation.

I have thus argued, in this section, that we cannot explain why an action is right simply by referring to the fact that the action is right. We then have reason to believe that the rightness of an action itself cannot be an appropriate reason for that right action. Since the rightness itself is unable to give us a proper reason for the action in question, it could be argued that the aim of doing whatever is right itself cannot be an intrinsically worthwhile goal itself. We know that, if an agent has a *de dicto* desire to do whatever is right, she must care about rightness itself—she must regard rightness itself as a fundamental value. Thus, an agent who is motivated, even in part, by the *de dicto* desire to do what is right actually cares about something that is not intrinsically worthwhile, given the argument above.

This further entails that, according to Wallace's definition of a 'fetish' which was discussed in section 4.2.2, being motivated by the relevant *de dicto* desire would be a moral fetish, as an agent who would be motivated by this desire would care about something that is not intrinsically worthwhile. This is the case independently of whether the agent also has other moral *de re* desires in addition to her *de dicto* desire to do whatever is right. An agent can still be argued to be a moral fetishist if she is motivated by the *de dicto* desire to do whatever is right independently of whether she also has a number of other *de re* desires at the same time.

To make my preceding responses to the externalist co-presence objection clearer, here is a brief summary of my arguments:

1) Based on Wallace's definition of a 'fetish', people should not devote interest and attention to objects that do not deserve it. In other words, we should devote interest and attention to objects that are worthwhile and worthwhile things are reason-providing properties.
2) It is possible that virtuous agents act for good reasons, this means that what they care about and what reasons they have are the same.
3) It is also possible that, what reasons there are to do actions and what makes the action right are the same. 2) and 3) taken together means that what virtuous agents care about and right-makers are one and the same.
4) According to the externalist view under discussion, the virtuous agents care about rightness itself. Yet, given 3), this would entail that rightness itself would be one right-making feature of an action.
5) However, 4) is an absurd result as the fact that an action is right cannot make the same action right, but rather the action must be right for some other reasons. And moreover, the rightness of an action itself cannot explain why the action is right (whereas the right-makers must be able to do this);
6) If an agent cares about the rightness itself and treats it as a fundamental value in her moral deliberation and motivation, she thus devotes interest and attention to a goal that is not worthwhile.

7) Thus, if an agent's moral motivation relies on the *de dicto* desire to do whatever is right, i.e., she cares about the rightness of the actions itself, the agent is a moral fetishist according to the definition in 1).

4.3 The Significance of the *De Dicto* Desire Response

4.3.1 The Objections

As discussed above, the externalist co-presence objection relies on the idea that the mere presence of the *de dicto* desire to do whatever is right is itself a harmless element of an agent's motivational structure. Yet, many externalists further defend their view by attempting to argue that the *de dicto* desire to do whatever is right is also necessarily required at least in two situations. Thus, by defending the efficacy and the appropriateness of the *de dicto* desire, many externalists would hope to prove that a plausible explanation of the reliable connection between moral judgments and motivation cannot be established without the relevant *de dicto* desire.

The externalists make two claims about the situations in which the *de dicto* desire to do whatever is right is needed. One claim is that this *de dicto* desire is needed only in rare cases and hence it is harmless as such. In the case of this first claim, the externalists argue that usually the relevant standing *de re* desires are enough for explaining why agents are motivated to act in accordance with their moral judgments. Because of this, only in very rare cases in which an agent cannot rely on her standing *de re* desires, the *de dicto* desire to do what is right must be invoked for providing a motivating desire. The other claim implies that the *de dicto* desire to do what is right is actually required in certain rare cases. In the case of the second claim, the externalists argue that the relevant *de dicto* desire is necessary when we have to ensure that an

agent is able to resist her temptations to do wrong things and that she thus has sufficiently strong motivation to follow her moral judgments.

Let us consider the first claim first. As we saw in the previous section, many externalists believe that people have a number of different non-derivative desires, i.e., they care for the well-being of their family and they desire to promote the justice and the flourishing of the society. These kinds of desires are fundamental, general and standing and, therefore, they can even survive changes in our moral judgments. This is why many externalists think that these standing desires are also able to explain the reliable connection between our moral judgments (both new moral judgments and old ones) and our motivation. This is because, in most cases, we give up our old moral judgments because we come to believe that the new ones will better satisfy our standing desires (or, put in another way: serve our fundamental moral values). Because of this, in the externalist framework, agents can usually be understood to be motivated by their standing *de re* desires in conjunction with their moral judgments (the old ones before and the new ones after).

We can illustrate this externalist proposal with a simple example. Let us suppose that Miya cares very much about her sister and would like to do everything she could to help her. At first, Miya thus judges that it is right for her to do everything that her sister asks, as doing so will best reflect her standing desires. But gradually, Miya comes to realize that doing whatever she could to help her sister to get what she wants probably does not help her sister and at times doing so could even be morally bad. Even if she still holds a desire to take care of her sister, Miya now judges that she should not satisfy her sister's every request, especially those that are inappropriate. Given this new moral judgment, plus her standing *de re* desire, Miya will not be

motivated to do everything her sister asks anymore. In this case, the externalists could explain the change in Miya's motivation by claiming that it follows from a change in moral judgment via the combination of Miya's standing non-derivative desires and her new moral judgment.

The problem of the previous proposal, however, is that there are situations where the standing *de re* desires cannot play the role of motivating us to act in accordance with our moral judgments. In these situations, one's new moral judgments arguably generate new motivation that does not consist of one's antecedent *de re* desires. In some cases, moral agents can even abandon all their previous fundamental moral cares and concerns and then become motivated as a consequence of their new moral judgments. If the new moral judgments in this type of cases do not lead to new desires themselves and we can no longer rely on the antecedent *de re* desires, where would the new motivation to act in accordance to the new moral judgment come from?

To answer this question, many externalists believe that it is necessary to introduce the relevant *de dicto* desire for doing whatever is right, that is, the motive of duty. This desire is needed to explain how we can acquire new desires and motivation even without the antecedent *de re* desires. The new desire to act in accordance with the new moral judgment is thus supposed to come from 'a standing commitment to do what is right, understood *de dicto*' (Shafer-Landau 1998, 356; 2003, 157).

Yet, this of course does not mean that we have to invoke the motive of duty in every case to explain the change in motivation following new moral judgments. Shafer-Landau (1998, 356; 2003, 158) believes that only in rare cases where our fundamental values (i.e., cares and

concerns) have changed, we need and should invoke the relevant *de dicto* desire to explain the reliable connection between moral judgments and motivation.

For example, let us consider a person who has never thought that animals have a moral standing and who, consequently, does not have an intrinsic desire to treat animals well. Let us then assume that this person then begins to assign intrinsic moral value to animals and, as a result of his change of this person's mind, he acquires new motivation to treat animals well. Where would this new desire to treat animals well come from if the person's new moral belief were unable to produce a new desire itself? According to Shafer-Landau (1998, 356), if the person in our example has assigned intrinsic value to animals, his new desire to treat animals morally must come from his new belief that animals should be treated well together with the motive of duty (which is the *de dicto* desire to do whatever is right). According to some externalists, it is thus not only appropriate but also sometimes necessary to invoke the *de dicto* desire to do whatever is right both here and in other similar cases.

Now, let us consider the second claim according to which the *de dicto* desire to do what is right is needed sometimes, a claim made by Lillehammer. He considers another scenario in which the relevant *de dicto* desire to do whatever is right seems to work better than the alleged *de re* desires and thus the former turns out to be necessary. Let us consider Lillehammer's (1997, 192) example to see his point.

Lillehammer describes a case of a woman who is tired of her husband as well as their marriage and thus goes to a party to have fun. During the party, she comes across a very charming person with whom she is tempted to have an affair. The woman judges that it would be wrong to have

an affair because she is considered about her husband's feelings and their marriage, even if she does not care about her husband's feelings too much at that moment.

However, even if the woman does not have a strong *de re* desire to do the right thing in this situation, the wife luckily has a standing *de dicto* desire to do whatever is right. Together with her moral judgment, this *de dicto* desire then causes the woman to do the right thing. In this situation, the woman's *de re* desire to do the right thing is not strong enough and so it is surpassed by her *de re* desire to do the wrong thing. According to Lillehammer, in this type of cases, only the relevant *de dicto* desire to do whatever is right can play a role in restricting people from doing something they regard as immoral. Therefore, it seems that the externalists' explanation of moral motivation can be motivated with the claim that the relevant *de dicto* desire can turn out to be irreplaceable for us in certain circumstances.

4.3.2 The First Response (1)

As it became evident during the previous discussion, many externalists agree with the internalists about the fact that moral agents have non-derivative *de re* desires to do things that are right.[33] They care, for example, about the well-being of their family and desire to promote justice and the flourishing of the society. Yet, the externalists have a different understanding of the function of these non-derivative desires. As I explained in the discussion of the previous objection (Section 4.3.1), Shafer-Landau thinks that the previous kind of non-derivative desires are 'fundamental, general and standing' desires, and, as a result, these non-derivative desires can survive changes in moral judgments. As we just saw, according to the externalists, moral

[33] See e.g. Shafer-Landau (1998, 356; 2003, 158), Cuneo (1999, 371-373), and Svavarsdóttir (1999, 201).

agents often give up their old moral judgments and make new ones because they think that the new judgments will better serve their fundamental moral values (i.e., the standing non-derivative desires). It is then the combination of standing non-derivative desires and new moral judgments that leads to changes in motivation (see the case of Miya in Section 4.3.1).

In response to this objection, I do not deny that good and strong-willed people can have some standing non-derivative desires. I can also agree with the idea that those standing non-derivative desires can sometimes have an influence on our moral judgments. However, where I think the objection goes wrong is the externalists' attempt to use these standing non-derivative desires, together with the new moral judgments, to explain the reliable connection between moral judgments and motivation. To see why this attempt fails, we can consider an example about an agent who is motivated by his standing desires to follow his moral judgments

Let us suppose that Green used to be a vegetarian. As a vegetarian, Green had a standing desire to be healthy and he also used to judge that eating meat is wrong. However, after some time, Green was also found to lack protein, because, due to his vegetarian diet, he did not get enough of the kind of protein we get from meat. According to his doctor, the best thing for Green to do is to change his diet and eat beef, pork, lamb and other meats which contain plenty of protein. Now, based on his standing desire to be healthy, Green comes to change his mind and judged that it would be morally permissible to eat meat after all. As a consequence, Green begins to eat meat. Here, what seems to cause Green to change his moral judgment is his desire to be healthy. Because of his standing desire to stay healthy, Green begins to judge that eating meat is right after all, even if he used to think before that the very same action is wrong.

The previous description of Green's mental transition from his old moral judgment to a new one seems to illustrate the externalists' explanation of the reliable connection between moral judgments and motivation explained above (see Section 4.3.1). Green abandons his old moral judgments just because the new judgment will better serve his standing desire to stay healthy. However, even if the reliable connection between beliefs and motivation in Green's case could be explained with the standing desire in the previous way, this explanation of the correlation still does not seem to be very plausible. The problem is that, according to this externalist account, Green would be making his judgments about right and wrong on the basis of his standing desires. Unfortunately, this way of thinking would make Green guilty of an objectionable form of 'wishful thinking'.

We can see the problem of wishful thinking, if we start from a plausible view of how desires should cohere with beliefs. Cian Dorr, for example, explained what the exact meaning of wishful thinking is in the following way. Suppose that there are conflicts between your empirical beliefs about how the world it is. Normally, it would be rational to resolve the conflicts by changing your views about one part of the world to cohere with the rest of your beliefs about how the world is. Yet, ordinarily we think that it is irrational to change our beliefs about how the world is simply so that those beliefs will match our desires and feelings (Dorr 2002, 99). We should not think that things are in certain way just because we would want them to be in that way. Personal desires and feelings just are mistaken grounds for us to form plausible beliefs of how the world is. This means that, equally, giving up beliefs on the basis of your desires is wishful thinking, something that is widely acknowledged to be irrational.

When the externalists attempt to argue that moral judgments change in accordance in terms of

standing *de re* desires, the problem is that this would make ordinary good agents guilty of committing the mistake of wishful thinking. The externalists seem to recommend that we, like Green in the previous case, should change our moral judgments so that beliefs about what is right can match our desires and feelings.

The externalists could reject this charge by arguing that it is harmless and ordinary to form our moral judgments in order to satisfy our desires given that there exist many cases in which we do form our beliefs on the basis of our desires. For example, suppose that you live in Nottingham and you have decided to travel to another city, perhaps London. In this situation, your desire to reach London may prompt you to form a belief that taking a train is the easiest and fastest way to get to London. And so, the externalists can argue that the way in which the standing desires prompt us to make new moral judgments cannot be claimed to be wishful thinking—it is not any more irrational than what happens in the previous case.

But this explanation will not help the externalists to avoid being committed to wishful thinking in their account of the reliable connection between moral judgments and motivation. A clear distinction can be made between different cases to show why the externalist account is implausible. In the previous case, your desire to reach London only prompts you to form a new belief but it does not determine content of that belief: in itself it does not make you believe that taking a train is the best way for you to satisfy your desire. Instead, your desire to get to London only prompts you to form a belief about the best means to get to London—after this, you need to come to the conclusion that taking the train is the best way on the basis of the evidence you have concerning how good the different ways to get to London are. Thus, we need to be aware that in some cases, even if our desires do prompt us to form corresponding beliefs, those beliefs

are still based on evidence.

Furthermore, the externalists are not merely committed to previous type of harmless and ordinary cases. Rather, the externalists are also committed to the more problematic form of wishful thinking way of making moral judgments, merely on the basis of one's desires. For example, in the previous case of Green, in order to have a reliable connection between moral judgments and motivation, Green must make the new moral judgment that eating meat is not wrong only on the basis of his desire to be healthy. It is not just that Green must make some moral judgments concerning meat eating but rather he must make a new moral judgment with that specific moral content.

We can then draw a brief conclusion about the externalists' claim that standing desires could explain the reliable connection between our moral judgments and desires. The externalists' theory seems to be objectionable especially when it comes to the claim that we can and should abandon our old moral judgments when the new moral judgments will serve our standing desires better. Even if we agree with the externalists that our standing desires can prompt us to form the relevant moral beliefs, we should not agree with them about whether our standing desires should also be able to determine the content of these beliefs. When we judge that an action is right, we should make such judgments on the basis of solid evidence rather than on the basis of our standing desires. If an agent changes her judgment about whether a certain action is right or wrong simply because of her existing standing desires, then she actually modifies her views about the world to cohere with her desires and thus commits the mistake of wishful thinking.

4.3.3 The First Response (2)

Furthermore, we may also have additional doubts about the externalist account of how the relevant *de dicto* desire will be required in the situations in which agents abandon their fundamental moral values completely (Section 4.3.1). Since the agents in these situations have lost their standing non-derivative desires, we cannot rely on the assistance of those antecedent desires to explain how the relevant agents are motivated by their moral judgments. In this situation, if the new moral judgments are unable to produce new corresponding desires directly as the externalists believe, where would the new desires to do things that are thought to be right derive from? The externalists believe that, under these unusual circumstances, we need and should rely again on the de dicto desire to do whatever is right.[34] The relevant *de dicto* desire to do whatever is right is necessary in this situation to ensure an agent would act according to her new moral judgments.

Let us return to an example, which we have already discussed in the section which outlined the externalists' objection to the fetishism argument (Section 4.3.1). In this example, there is a person who suddenly begins to assign intrinsic values to animals even if, before this point, she never thought that animals would have a moral standing. In this case, there does not seem to be an antecedent standing desire for animal welfare such that it could explain how the new moral judgment would motivate the agent to act. An externalist explanation will have to employ the *de dicto* desire to do whatever is right in order to explain how the agent forms a new desire to treat animals in humanistic way. Otherwise, without that standing desire, the externalism would be unable to explain why the agent suddenly is motivated to treat animals with respect. So, the

[34] See e.g. Lillehammer (1997, 191-192), Shafer-Landau (1998, 356-357; 2003, 158-159), and Svavarsdóttir (1999, 199&210).

involvement of 'the relevant *de dicto* desire' seems to be required in this case so as to enable the externalist framework to explain the agent's change in motivation.

But, in my mind, to cite the relevant *de dicto* desire is not a plausible way to explain the above-described case as the externalists claims. As I have already argued in Section 4.2.2, if an agent is motivated by the *de dicto* desire to do whatever is right, then the *de dicto* desire to do whatever is right itself is sufficient to make the agent a moral fetishist. This is because, if an agent is motivated by the *de dicto* desire to do whatever is right, she cares about something that is not worthwhile of her interest and attention. And as we have already seen, according to our new definition of fetishism, caring about something that does not deserve our interest and attention appears to be fetishistic. Our previous conclusion from Section 4.2.2 thus entails that, if the externalists think that the *de dicto* desire to do whatever is right is necessary to motivate moral agents at least in some cases, their view would make ordinary agents moral fetishists in an objectionable way. The idea that ordinary moral agents would be moral fetishists still creates a problem for the externalists.

4.3.4 The Second Response

As explained above, some externalists have also argued that the relevant *de dicto* desire to do whatever is right is in fact preferable and needed in certain circumstances (Section 4.3.1). They have argued that, in addition to the fact that the relevant *de re* desires can co-exist with the relevant *de dicto* desire, the *de dicto* desire to do whatever is right can also play a central role in good and strong-willed people's motivational structure when there are motivational conflicts.

According to the externalists in questions, only the *de dicto* desire to do whatever is right can help us overcome temptations and to prevent us from being motivated to do the things that are regarded as immoral.

Recall Lillehammer's example (1997, 192). In his case, the woman who is tempted to have an affair with the charming person at the party is claimed to need the *de dicto* desire to do whatever is right to prevent her from following her *de re* desire to do the wrong thing. Lillehammer argued that, since the woman is overwhelmed by her *de re* desires to do things that are indeed wrong, the co-existing *de re* desires to do things that are right (such as her concern for her husband's feelings) would not be sufficiently strong. As a consequence, the relevant *de dicto* desire to do whatever is right seems to be needed—it is not clear how the woman would be able to resist temptations to do what is wrong in this case without that desire, or so Lillehammer argues.

We can see that Lillehammer believes that the *de dicto* desire to do whatever is right has to be invoked in order to guarantee that in certain situations moral agents are able to perform right actions. What Lillehammer thus ultimately cares about seems to be what is required for the generation of *sufficient* motivation to act when an agent has judged that she ought to act in a certain way. At this point, the externalists seem to have changed the question into a new one that is obviously different from the one discussed by Smith.[35]

[35] Many externalists tend to defend their view by considering questions that are slightly different from Smith's (and Lillehammer's new question is just one of them). In line with Lillehammer, Vanessa Carnbonell (2013) shares the view that the relevant *de dicto* desire is supposed to play a role when there are conflicts between the relevant *de re* desires. Ron Aboodi (2015, 2016) discusses situations where the relevant *de dicto* desire can resolve an agent's uncertainty regarding underived moral values.

What Smith considers is the question of how we are going to explain the reliable connection between moral judgments and motivation when an agent has made a moral judgment. However, Lillehammer has quietly changed that question into a new one which could be summarized as: how are we going to explain the reliable connection between moral judgments and sufficiently strong motivation that would enable an agent to do what she thinks is right? Lillehammer's question clearly has some relation to Smith's question. Yet, even if these two questions are related, they are still different questions. I grant that Lillehammer's question is also an important question and something we need to consider, but before we do so, we need to focus on Smith's question and on whether the externalists have provided a satisfactory response to it. The main crux of the whole chapter is to argue that the externalists are not able to do so.

Additionally, even if we were concerned about how to guarantee that agents have sufficiently strong motivation to act according to their moral judgments, it does not seem like relying on the externalists' *de dicto* desire to do whatever is right would be the best choice. The internalists can be argued to offer a better answer than the externalists to the question of how a sufficiently strong motivation can be acquired following a moral judgment. Let us return to Lillehammer's case of the woman who is tempted to have an affair and see what internalists would say about it.

The internalists can argue that, even if the woman in question is fascinated by the charming person she meets and even if she does not care too much about her husband's feelings, she gradually realizes that she should not have an affair with the stranger because such behaviour would hurt her marriage and her husband as well. The more the woman thinks about her husband, the more she would remember the happy times they have spent together. In this

situation, the woman could come to realize that it is much more important to be faithful to her husband and their marriage. Instead of having an affair with the charming person, she now wants to maintain her relationship and being faithful to her husband.

At this moment, it could well be that the agent's *de re* desire to do the right thing—the way in which she cares about her husband and their marriage—is able to outweigh her *de re* desire to do the wrong thing. With this in mind, it is not obvious that we do need to invoke the relevant *de dicto* desire to ensure sufficient strong motivation to do the right thing. The externalists' account thus makes a 'simplistic and misleading contrast being made between morality and rightness on the one hand, and personal feelings and the like on the other' (Toppinen 2004, 311). When being faced with conflicts between temptations and moral requirements, the best an agent can do is to seek more reasons for performing the actions which she judges to be right as such considerations can strengthen her motivation. Furthermore, it is not more preferable but rather much less preferable to think 'even more furiously about the rightness of the act in itself' (Toppinen 2004, 312). As I have already argued, that is merely a way of trying to be motivated by something that is not intrinsically worthwhile.

4.4 An Externalist Objection to the Revised Fetishism Argument

In his revised version of the fetishism argument, Smith discusses the moral conversion that happens to one of his friends (we called him Mike). Through discussing the example, Smith presents two reason for why we should reject externalism (Section 3.4.3). He first argues that the externalist accounts of Mike's moral conversion seem to be merely theory-driven. When Smith makes this objection, what he has in mind is that the only reason to think that Mike's new desire to give extra benefits to his family and friends is an instrumental desire would be

based on our prior commitment to externalism. In other words, if we had not accepted externalism, we would not have any reason to accept the externalist counterintuitive description of Mike's moral conversion. Secondly, Smith also argues that, if the externalists believe that morally good agents' primary source of moral motivation is a non-instrumental desire to do whatever is right, this commits the externalists to an implausible picture of moral perfection.

In Section 4.4.1, I will first discuss Svavarsdóttir's objection to Smith's theory-drivenness accusation. According to Svavarsdóttir, even in the externalist framework, morally good people's desires to act in accordance with their moral judgments do not always have to be instrumental desires that are derivative from the non-instrumental desire to do whatever is right. Based on her understanding of 'instrumental and non-instrumental' desires, Svavarsdóttir provides a new description of how Mike's mental states are before and after his moral conversion. The motivation for this description is that it is supposed to be both compatible with externalism and also fit our common-sense intuitions about the case and hence the claim is that it will not be merely theory-driven. Following Svavarsdóttir's argument, I will give an internalist response to this argument in Section 4.4.2. I will argue that, if Svavarsdóttir's view were true, then agents would in many cases still desire to do what they have already judged to be wrong.

After this, in Section 4.4.3, I will consider Svavarsdóttir's objection to Smith's moral perfection criticism. Following Svavarsdóttir's argument, in Section 4.4.4, I will argue against her objections. I will first demonstrate that, if morally perfect people should be motivated by the feature of being a right-making feature and the right-making features themselves at the same time, then such morally perfect people would not be any less fetishistic. I will then argue that

the feature of being a right-making feature itself cannot help us to conduct moral deliberation effectively in the way that Svavarsdóttir suggests.

4.4.1 Externalists on Theory-drivenness

As already explained in Section 3.4, in his revised version of the fetishism argument, Smith first argues that the externalist account of Mike's moral conversion would give us an implausible entirely theory-driven picture of the process. After all, the externalist description of Mike's moral conversion suggests that Mike would be moved to follow the changes in his beliefs by acquiring new instrumental desires to benefit his family and friends as such desires would serve his non-instrumental desire to do whatever is right. If Mike were to be motivated in this way instead of being motivated by his moral judgment in some more direct way, he appears to care about something that is not of primary moral importance. Yet, ordinarily we assume that virtuous people are motivated by their moral beliefs directly and have non-instrumental desires to do the right things. Smith thus claims that the only reason anyone could have for accepting the externalists' instrumental description of Mike's motivation would be theory-driven—motivated by their prior commitment to externalism.

In order to illustrate the previous responses, Svavarsdóttir offers a different description of the psychological transition during Mike's change of his convictions, which she claims to be a more plausible and intuitive account than the internalist alternatives (Svavarsdóttir 1999, 209). At the first stage, even if Mike is thoroughly convinced of utilitarianism, according to Svavarsdóttir he still has some inclinations to give additional benefits to his family and friends. It is just that, because he is so convinced of the correctness of utilitarianism, this prevents Mike from favouring his family and friends very often. After all, utilitarianism, as an ethical theory, states

that doing so does not always meet the criterion for moral rightness. Given that he accepts utilitarianism, Mike considers each person's happiness equally in order to maximize the total amount of happiness, as utilitarianism requires. Given that his dominant desire is to maximize happiness and minimize suffering, some might think that Mike is a cold and calculating utilitarian monster. Yet, Mike is about to change his mind. Through discussing utilitarianism with others, Mike gradually realizes that he applies the utilitarian theory in a very strict way, which means that he has to sacrifice his family and friends. As a consequence, Mike's belief in utilitarianism erodes slowly and he then comes to prefer favouring his family and friends even when this behaviour cannot be justified on utilitarian grounds.

Svavarsdóttir (1999, 210) argues that it is not necessary to assume that Mike's desire to do whatever he judges to be right needs to be invoked all the time, especially when it comes to his motivation to give extra benefits to his family and friends. Rather, Svavarsdóttir believes that Mike does not have to develop an additional desire to give extra benefits to his family and friends during the second stage of his moral conversion. Instead, having a disposition to be partial towards friends and family should be considered to be a standing desire, which is held by most ordinary people. It is just that Mike's standing desire to give special benefits to his family and friends was initially outweighed because his commitment to utilitarianism is so strong. Yet, when Mike's commitment to utilitarianism erodes, his standing desire to give more benefits to his family and friends will again be strong enough to prompt him to act.

Svavarsdóttir (1999, 210-212) admits that Mike's desire to maximize happiness and minimize suffering actually derives from the *de dicto* desire to do whatever is right in conjunction with his utilitarian moral judgment. However, Svavarsdóttir does not take the relevant derivative

desire to be an instrumental desire in the same way as Smith does. Instead, she argues that Smith is mistaken in treating the desire to maximize happiness as an instrumental desire, which serves the non-instrumental desire to do whatever is right.

To support her view, Svavarsdóttir tries to explain what kinds of a desire she considers to be instrumental. Svavarsdóttir illustrates the idea by describing a case about someone who intends to buy sandals. When an agent enters a store, which she believes to sell them, her desire to enter the store is a part of her means to satisfy the desire to buy the sandals. Svavarsdóttir thinks that the agent's desire to enter the store is an instrumental desire that we assume to be because the desire to enter the store clearly aims to satisfy the agent's further desire which is to buy sandals.

Svavarsdóttir, however, argues that Mike's desire to maximize happiness is not supposed to be in the same way a part of Mike's means to satisfy his non-instrumental desire to do whatever is right. She thinks that the alternative externalist revised description of Mike's mental conversion has shown that Mike's desire to maximize happiness does not satisfy a more fundamental desire, such as, the non-instrumental desire to do whatever is right. According to Svavarsdóttir, it is more plausible to think that Mike's desire to maximize happiness should be construed as a response to Mike's relevant moral judgment, and more importantly, Mike's desire to maximize happiness is a non-instrumental desire to do the actions that are right.

Furthermore, Svavarsdóttir indicates that even if the desire to maximize happiness were a result of the combination of Mike's acceptance of utilitarianism and his non-instrumental desire to do whatever is right, the desire to maximize happiness would not always need to be subservient to the non-instrumental desire to do whatever is right. It could be suggested that the desire to

maximize happiness could also function on its own even when Mike abandons utilitarianism and revises his moral values. At this point, the desire to maximize happiness would have its own independent status without serving any other desire anymore and, thus, this desire to maximize happiness should not be treated as an instrumental desire.

This new externalist description could also be claimed to be both compatible with externalism and our common-sense intuitions about how Mike's motivations are non-instrumental and hence, contrary to Smith, it could be argued that the description is not wholly theory-driven but rather also independently motivated. It seems that, according to Svavarsdóttir, Smith has not described fairly Mike's mental conversion on behalf of the externalists. As a consequence, Svavarsdóttir's new externalist description of Mike's moral conversion would not make Mike a moral fetishist as Smith argues.

4.4.2 The Response

Svavarsdóttir's objection to Smith thus consists of two steps. First, Svavarsdóttir believes that, after his conversion, Mike would not need to accept any special moral theory which would permit partiality towards family and friends as this kind of behaviour is already widely thought to be permissible. Rather, according to her, Mike's desire to favour his family and friends is a standing desire, which he must have held in some form or another even during the time when others considered him to be a utilitarian monster. After his mental conversion, Mike just loses the relevant restrictions on his desire to favour family and friends and, as a consequence, his desire to give extra benefits to family and friends becomes more effective.

I agree that Mike does not need to develop an instrumental desire to help his family and friends.

Yet, the problem is that Svavarsdóttir then goes on to defend the less plausible claim that our standing desires explain the reliable connection between moral judgments and motivation. Svavarsdóttir indeed suggests that Mike's standing desires to care about family and friends are stable and can survive even the changes in Mike's moral judgments. Furthermore, these standing desires can eventually be sufficient to motivate Mike to act accordingly.

I have already discussed a very similar thought proposed by Russ Shafer-Landau in Section 4.3.2, the conclusion of which equally applies here. The externalists' theory seems to be objectionable especially when it comes to the claim that we can and should abandon our old moral judgments when the new moral judgments will serve our standing desires better. Even if we agree with the externalists that our standing desires can prompt us to form the relevant moral beliefs, we should not agree with them on whether our standing desires should also be able to determine the content of these beliefs. If an agent changes her judgment about whether a certain action is right or wrong simply because of her existing standing desires, then she actually modifies her views about the world to cohere with her desires and thus commits the mistake of wishful thinking.

The second, and more serious objection, made by Svavarsdóttir concerns the meaning of 'instrumental desires'. She argues that in Mike's case, his desire to maximize happiness does not have to be an instrumental desire that serves the non-instrumental *de dicto* desire to do whatever is right. Svavarsdóttir also claims that, even if Mike's desire to maximize happiness functioned as an instrumental desire at first, this desire will be an independent desire later. Later, it will not need to aim at the satisfaction of any other, more fundamental desires. If this description of Mike's conversion were along the right lines, then Mike could not be accused of

having an instrumental desire to maximize happiness. In that case, Smith would be wrong to claim that the externalist must provide a theory-driven description of the case.

In response to this claim, I will argue that the externalists will find it hard to explain Mike's case plausibly if they rely on the idea that Mike's initial desire to maximize happiness could become a non-instrumental desire later on. Before pointing out what the difficulty for the externalists really is, let us consider an example of a racist who develops an independent desire of wanting to treat people of other races with prejudice.

This racist used to believe that her race is superior to other races because an unhappy personal experience convinced her of that view. She used to judge that it was right for her to treat others badly because of her superiority. Let us imagine that, based on her racist moral judgment and her desire to do whatever is right, the racist develops an instrumental desire to act in the racist way. Although this desire at first merely satisfies the racist's non-instrumental desire to do whatever is right, we can imagine that the racist's desire to act in the discriminating way begins to have a life of its own. Let us then assume that years later and through having more experiences about people of other races the racist comes to realize that it is wrong to be a racist. She comes to think that we should not judge people simply on the basis of racial stereotypes because, for one, doing so prevents her from seeing many other, more important differences and similarities between people. The racist in question has thus changed her mind.

If we were to accept the externalist framework according to which instrumental desires can gain a status of their own and hence function independently later on, what should we think would happen to the racist as we have described her in the externalist framework? Presumably, even

if she doesn't think anymore that she should be a racist, she would undoubtedly still not treat members of other races kindly and politely. As the racist has developed an independent desire to act like a racist, even though she has changed her judgment about such behaviour and now believes that such behaviour is wrong, the desire to act as a racist could still function independently if the externalists are right. The 'racist' would thus probably still treat people of other races with superiority. It seems that the agent in question would still be motivated by her bad desires even though she no longer holds the corresponding moral judgments. If we were to accept the externalist account of how instrumental desires can gain a life of their own, then we would also have to accept this absurd result, which derives from externalism.

The same conclusion also applies to Mike's case. If Mike's initial desire to maximize happiness becomes a non-instrumental desire in the way that the externalists have suggested, then Mike could still desire to maximize happiness even after his moral conversion because this desire could exist and function independently. If Mike would still be influenced by his old utilitarian desire even when he has lost the corresponding moral judgment, it seems that he would not necessarily be more motivated by his standing desire to give extra benefits to his family and friends. Because of this, Svavarsdóttir does not seem to be able to offer a plausible explanation of Mike's conversion from being a strict utilitarian practitioner to becoming an ordinary human being who cares more about his family and friends. If this is right, then Svavarsdóttir cannot explain the reliable connection between moral judgments and motivation in a compelling way. This means that the only reason to accept the previous implausible description of Mike's psychological conversion would again be a prior commitment to externalism, which means that even Svavarsdóttir's new description of Mike's conversion is not any less theory-driven than the previous externalist description sketched by Smith himself.

4.4.3 Externalists on Moral Perfection

In his second version of the fetishism argument, Smith also argues that the externalist description of Mike's moral conversion is not compatible with how we ordinarily understand morally perfect people. As a consequence, the concern is that the externalists set an implausible standard of moral perfection (for a more thorough discussion, see Section 3.4.3 above). Smith begins from the claim that, in the externalist framework, morally perfect people must be motivated by 'the feature of being a right-making feature' (Smith, 1996b, p.182). However, if this were what the externalists thought about moral perfection, their view would make morally perfect people alienated from our normal understanding of morality, or so Smith argues. We usually assume that morally perfect people should be motivated by the features of actions that make them the right actions to do. As Smith puts it, if a morally perfect person were motivated by the feature of being a right-making feature as the externalists seem to suggest, then she would seem to desire something that does not have primary moral importance.

Here too, Svavarsdóttir's first response is the co-presence objection (see Section 4.2.2 above). She argues that, even if externalism were true, moral agents could still be motivated by the right-making features of actions most of the time (Svavarsdóttir 1999, 214). In most cases, the moralists will, on the basis of their standing *de re* desires, desire to act in ways that track the right-making features of their actions. As a consequence, the moralists care about the right-making features themselves rather than about the abstract feature of being a right-making feature. Furthermore, at least at times, good moral agents can in this way be motivated by the right-making features of their actions without realizing that those features are right-making

features. For example, most people help their family and friends without even realizing that it is the fact that those people are their family and friends that makes doing so right.

Secondly, Svavarsdóttir also argues that being motivated by the feature of being a right-making feature is sometimes necessary for moral agents. This is because, according to Svavarsdóttir, what the relevant right-making features are in a given situation is not always so obvious to us as moral agents (Svavarsdóttir 1999, 214). Hence, as we are not always clearly aware of the specific rightness-making features of our alternatives, we occasionally have to rely on moral rightness itself to guide our actions[36]. On this view thus, when an agent considers 'the feature of being a right-making feature'—that is, the fact that certain qualities of an action make the action right—she is just conducting ordinary moral reflection, which is a process that helps her to identify what she ought to do. Svavarsdóttir assumes that this type of moral reflection is quite normal and necessary for us as moral agents. As moral agents cannot always recognize the right-making features directly, we need to rely on 'the feature of being a right-making feature' for guidance every now and then. If this is right, it can be argued that it cannot make us any less morally perfect that the feature of being a right-making feature motivates us in some cases.

In her third objection, Svavarsdóttir argues that Smith is committed to the idea that morally perfect people do not morally deliberate which would make moral perfection an unattainable ideal (Svavarsdóttir 1999, 214). This is because, as Smith tries to exclude 'the feature of being a right-making feature' from being salient for morally perfect people in their deliberation, it seems that he is also suggesting that insofar as we ordinary agents try to pursue moral perfection we should not be reflecting morally. Svavarsdóttir claims that, if this is true, then Smith sets

[36] For a similar view, see Carbonell (2013). In her paper, Carbonell argues that the *de dicto* desire is usually a way of 'mediating our unreflective motivations' (Carbonell 2013, 466).

the standard of moral perfection very high. Morally perfect people seem to become a group of people that are very different from ordinary good and strong-willed people who do deliberate morally. After all, it seems like morally perfect people would have to be able to recognize the right-making features of their actions without moral reflection. These moral agents would seem to be able to always make right judgments and do morally right actions without any deliberation. According to Svavarsdóttir, this picture of morally perfect agents does not describe how moral agents are usually motivated to act as we normally understand the process. As we humans are imperfect and as we are unable to identify the right-making features all the time in all cases, we need to consider the feature of being a right-making feature in our deliberation at least sometimes. If Smith excluded the necessity of relying on 'the feature of being a right-making feature' in deliberation, he would also preclude the imperfections of human beings, or so Svavarsdóttir argues.

In the same way as Smith accuses the externalists of an implausible account of moral perfection, Svavarsdóttir too thus claims that Smiths sets a too demanding standard for moral perfection. She believes that, on Smith's view, moral perfection would be particularly demanding for ordinary good people because we normally assume that good people make mistakes and sometimes resort to considering the feature of being a right-making feature in moral reflection. Moreover, Svavarsdóttir also claims that, if there were a group of moral agents who satisfied Smith's conception of moral perfection, those people would appear to be moral fetishists. According to Svavarsdóttir, moral fetishism should be most appropriately thought of as 'the phenomenon of holding oneself and others to rigorous moral standards, while being completely unwilling to entertain any reflective question about their nature or grounds' (Svavarsdóttir,

1999, p.213). She claims that morally perfect people in the way Smith understands them would be exactly this kind of moral fetishists.

4.4.4 The Response

In response to Svavarsdóttir's first objection to Smith's moral perfection argument, we should recall that I have already argued that the externalists' co-presence objection fails. That same conclusion also applies to Svavarsdóttir's first objection here. Caring about both 'the right-making features' themselves directly and also about the general abstract 'feature of being a right-making feature' will not make a moral agent any less fetishistic given that having the latter concern itself is sufficient to make an agent a moral fetishist (see the previous Section 4.2). In response to Svavarsdóttir's second and third objections to Smith's moral perfection argument, I will argue that (i) considering 'the feature of being a right-making feature' itself will not help moral agents in their moral deliberation and also that (ii) Smith is not committed to the idea that morally perfect people do not need relying on moral reflection.

Recall that Svavarsdóttir's second objection to Smith's moral perfection argument was that we are sometimes uncertain about the right-making features of our actions, which means that we need to think about the feature of being a right-making feature. However, instead of being a successful objection, Svavarsdóttir's objection leads to a dilemma. To see one horn of the dilemma, let us imagine a situation in which we have no idea of what the right-making features of our actions are. In this case, it would be impossible to think about the feature of being a right-making feature as we are unaware of what features of actions in question have that exact feature—the feature of being a right-making feature. To see the other horn of the dilemma, let us imagine another situation in which we already know what the right-making features are. In

this situation, we are able to think about the feature of being a right-making feature, which some of the features of our actions have. Nevertheless, here we do not need to think about the feature of being a right-making feature as we already know what makes actions right. Thus, we do not need to accept Svavarsdóttir's second objection that we need to rely on the feature of being a right-making feature for moral deliberation sometimes.

After considering Svavarsdóttir's second objection to Smith's moral perfection argument, let us discuss Svavarsdóttir's last objection to Smith's argument. Smith would unquestionably not argue that morally perfect people do not need moral reflection as Svavarsdóttir claims. It is probable that, during his discussion of how moral judgments motivate, Smith just wants to consider simple cases in which the moralists are able to make right moral judgments without too much of a need to deliberate (Smith 1996b, 181-182). Yet, it is obvious that even virtuous people cannot always have a clear idea of what is right and wrong and making right moral judgments is not always as easy as in Smith's examples suggest. What this implies is that, with his example, Smith does not intend to define morally perfect people as agents who are so impeccable in nature with respect to morality so as not to need any moral reflection.

The morally perfect people on Smith's view are far from being akin to the moral fetishists that Svavarsdóttir has in mind. In fact, the internalists can also provide an account of moral deliberation. In moral deliberation, we usually try to consider what features of actions make them right or wrong. On the basis of considering right-making features of an action, we then come to conclude that the action is either right or wrong overall. The only thing that the internalists insist on is this final judgment is able to produce motivation directly in the agent.

So far, in Sections 4.4.1-4.4.4, I have already considered and discussed Svavarsdóttir's objections to Smith's revised version of the fetishism argument. In her response to Smith's first theory-driven objection, Svavarsdóttir argues that in order to be motivated in a way that matches their moral judgments, virtuous people do not need to be motivated by instrumental desires grounded in their non-instrumental desire to do whatever is right. Even if moral agents have acquired their instrumental desires to act in accordance with their moral judgments, these instrumental desires could function on their own later on. Based on her understanding of instrumental and non-instrumental desires, Svavarsdóttir thus provides a new description of how Mike is motivated by his moral judgments. In response, I have argued that, if we accepted Svavarsdóttir's theory of instrumental and non-instrumental desires, this would have absurd consequences. It could be argued that an agent would desire to do what she used to judge to be wrong even after she has changed her mind, as her previously instrumental desires would have acquired at this point their own stable and direct standing and could thus function independently. For the same reason, Svavarsdóttir's re-description of Mike's moral conversion fails, too.

In her second externalist objection to Smith's argument based on 'moral perfection', Svavarsdóttir argues that morally perfect people in Smith's theory would not need to rely on the feature of being a right-making feature in their moral reflection at all on Smith's view. She then claims that Smith has set an implausibly high and rigorous standard for who would count as morally perfect people. If so, according to Svavarsdóttir, these morally perfect people would furthermore seem to be a kind of moral fetishists. In response, I argued that, unlike specific right-making features, the feature of being a right-making feature would not provide any new information for moral agents and so it cannot help moral agents to make their moral judgments. More importantly, in Smith's framework, morally perfect people could deliberate morally even

if their reflection would be focused on the right-making features of actions themselves rather than on these features' additional feature of being a right-making feature.

4.5 Conclusion

This chapter has discussed some of the main externalist objections to (both versions of) Smith's fetishism argument. My aim has been both to deal with the direct externalist objections and to respond to the externalist challenges based on those objections. Each section in this chapter has focused on different externalist objections, as well as provided internalist responses to them.

In Section 4.2.1, I considered the externalists' common claim according to which the *de dicto* desire to do whatever is right can exist at the same time with the relevant *de re* desires in good and strong-willed people. on the basis of this claim, many externalists then go onto claim no one should be thought to be a moral fetishist merely because she has the *de dicto* desire to do whatever is right. After all, she can also have the required direct *de re* moral concerns at the same time.

In response to this objection, I first introduced a more general definition of fetishism. This suggested understanding of fetishism claims that caring about something that does not deserve our interest and attention is fetishistic. Then, in Section 4.2.3, I argued that being motivated by the *de dicto* desire to do whatever is right is identical to caring about an abstract property—the rightness of actions itself. I then went onto argue that *rightness* itself is never a good reason for an action and for this reason we can conclude that being motivated even in part by the *de dicto* desire to do whatever is right is still fetishistic even if the agent has other direct *de re* concerns.

Section 4.3 then discussed two externalist attempts to avoid the fetishism objection. In Section 4.3.1, I first introduced how some externalists have tried to appeal to standing *de re* desires in order to explain the reliable connection between moral judgments and motivation. Many externalists then go on to add that, in the situations where the standing *de re* desires are not present, the *de dicto* desire to do whatever is right will be required. I then introduced the second, more radical externalist objection, which argues that moral agents always need the relevant *de dicto* desire in order to be able to resolve conflicts between their *de re* desires, and finally also in order to gain sufficiently strong motivation to act in accordance with their moral judgments.

I criticized the first aspect of the externalist objection in Section 4.3.2. I suggested that the externalists who give this response are committed to the idea that moral agents should change their moral judgments to match their standing *de re* desires. This would mean that moral agents would fix the content of their moral judgments on the basis of their standing *de re* desires. We, however, should not accept this idea as it would endorse an objectionable kind of 'wishful thinking'. I also emphasized and clarified in Section 4.3.3 that the employment of the relevant *de dicto* desire, even if as a supplement theory, would still make ordinary agents fetishists as explained earlier in Section 4.2.

In Section 4.3.4, I argued that the externalists also make a mistake when they claim that the *de dicto* desire to do whatever is right could help us to overcome temptations and prevent us from being motivated to do things that are regarded as immoral. At this point, the externalists are actually trying to answer a new, different question: How could we explain the reliable connection between moral judgments and sufficiently strong motivation that would enable an agent to do what she thinks is right? This new question is different from Smith's original

question concerning the reliable connection between moral judgments and motivation. Because of this, the externalists cannot defend their view merely by raising the new question unless they are also able to answer the original question too. Furthermore, the internalists seem to have an answer to the new questions too: the best an agent can do in order to be sufficiently motivated to follow her moral judgment is to seek more reasons for performing the actions which she judges to be right as such considerations can strengthen her motivation. It is much less useful to think furiously about the rightness of the actions as the externalists suggest, as that will be of little help.

In Section 4.4, I went through Svavarsdóttir's objections to Smith's revised fetishism argument. The first objection was to the 'theory-driven' charge. Svavarsdóttir tried to explain away this objection by again relying on standing *de re* desires to explain the way in which Mike is motivated. She also tried to argue that Smith misunderstands the meaning of instrumental desires. According to Svavarsdóttir, the fact that certain desires derive from the non-instrumental desire to do whatever is right does not necessarily mean that these derivative desires are instrumental. Instead, she claims that the initially derivative desires should be assumed to gain an independent status and therefore serve for their own purposes. With this new understanding of 'instrumental' and 'non-instrumental' desires, the externalists hope that their redescription of Mike's mental conversion is not as theory-driven as Smith suggests.

In Section 4.4.2, I objected to Svavarsdóttir's new description of Mike's mental conversion as well as to her analysis of instrumental desires. I argued that, if the externalists' claim that the initially derivative desires should later on be thought to be self-standing, non-instrumental desires were true, then an agent would be moved by her old desires even when she has changed

her mind—when she has begun to believe that being motivated by those desires is wrong. Obviously, given that this is an absurd result, we should reject the externalist account of this type of cases.

In Section 4.4.3, I responded to the externalist objection to Smith's new argument based on moral perfection. According to Svavarsdóttir, though a moral agent is motivated by the right-making features of her actions most of the time, she can also be motivated by the feature of being a right-making feature when it is not clear what the real right-making features are in a given case. In this situation, Svavarsdóttir suggests that moral agents should rely on the feature of being a right-making feature in their moral deliberation. She also claims that morally perfect people in Smith's theory would not need to do much moral reflection at all. Because of this, the externalists tend to imply that Smith has set a too demanding standard for who would count as morally perfect and they also think that acting morally without any moral reflection itself appears to be a kind of a moral fetish.

In Section 4.4.4, I responded to these externalist objections to Smith's argument against externalism based on moral perfection. I first argued that, even if a moral agent were motivated by the right-making features for most time, the fact that she would also be at least in part motivated by the feature of being a right-making feature would still make her a fetishist. I then argued that the feature of being a right-making feature cannot help a moral agent to deliberate as effectively as Svavarsdóttir suggests. Finally, I responded that morally perfect people within Smith's framework not only conduct moral reflection but also will deliberate better than the externalists suppose.

Overall, if my responses to the externalist objections to the (both versions of) Smith's fetishism argument are compelling, I will have at least shown the implausibility of accepting the forms of externalism that rely on *de dicto* desires to explain the reliable connection between moral judgments and motivation. My defences and developments of the fetishism argument will also in this situation make internalism a more compelling way to understand the previous reliable connection. The externalists might, of course, then try to find other ways of defending their views, for example, versions of externalism that do not rely on the *de dicto* desire to do whatever is right at all. If those attempts succeeded, the externalists could potentially bypass the fetishism altogether. This will be the topic of the next chapter.

Chapter 5: The Externalists' Alternative Explanations

5.1 Introduction

Chapter 4 focused on one way in which the externalists have tried to respond to the fetishism argument. The externalists, who have pursued the strategy I explored in the previous chapter, are still committed to relying on the *de dicto* desire to do whatever is right—which I argued to be problematic—to explain the reliable connection between moral judgments and motivation. Yet, many externalists have also done more than that. Many of them intend to avoid Smith's fetishistic objection all together by providing explanations of the previous reliable connection that do not rely on the relevant *de dicto* desire at all (Copp 1995,1997; Cuneo 1999; Dreier 2000; Lillehammer 1997). In this chapter, I will consider these other ways in which the externalists have tried to explain the reliable connection between our moral judgments and motivation. I will also argue that these externalist alternatives are all implausible in one way or another and thus should be rejected

In Section 5.2, I will discuss the so-called practicality option explanation and its problems (Lillehammer 1997, 194). Hallvard Lillehammer develops this explanation based on the idea that motivation connects to our moral judgments by corresponding to the normative reasons, which our moral judgments track. Lillehammer then suggests that there are certain exceptional cases in which what were motivated to do should not be supposed to correspond to the normative reasons connected to our moral judgments. For example, when a moral judgment is defective, corrupted or even wrong, there are no reasons to act in accordance to that judgment.

In response to Lillehammer's proposal, I will argue that, if the practicality option were true, we would be unable to make sense of what happens in many actual moral disagreements. Normally

we expect that both sides in ordinary cases of moral disagreement are still motivated by their respective moral judgments. But the practicality option view, as defended by Lillehammer, would entail that we should expect only one side of the moral disagreement to be motivated by her judgment—the person whose judgment is correct.

In Section 5.3, I will discuss Terence Cuneo's explanation of the reliable connection between moral judgments and motivation based on the notion of genuinely virtuous people (Cuneo, 1999, 369). Cuneo believes that, if we can figure out how virtuous people are motivated by their moral judgments, we will be able to explain the reliable connection between moral judgments and motivation more generally too. According to him, as an indication of their virtue, virtuous people have a variety of substantial concerns that consist of different desires, aversions, attachments, interests, cares and the like. He then suggests that these concerns motivate virtuous people to act in accordance with their moral judgments. To illustrate his idea, Cuneo discusses how the notion of virtue can explain how different people are motivated in Smith's voting case (Smith 1994, 71; Section 3.2 of this thesis). In response to Cuneo, I will first argue that, if a moral agent's motivation always had to be based on her pre-existing virtues as Cuneo seems to claim, it would be difficult to explain the reliable connection between moral judgments and motivation in those who are less than fully virtuous.

In Section 5.4, I will discuss the notion of 'morally suggestible people' provided by David Copp (1997, 50) and James Dreier (2000, 623-624). This proposal begins from Copp's idea of a *disposition* to desire to do what you judge to be right. He then suggests that this disposition can be used to explain the reliable connection between moral judgments and motivation. According to Copp's view, a rational agent who has the previous disposition will be disposed to do what

she judges to be right directly without the assistance of any other desires. In response, I first introduce Dreier's tracking test and then show that morally suggestible people would fail to pass it and therefore it fails to be a plausible account of the reliable connection between moral judgments and motivation. The problem of the morally suggestible people model is that agents understood within this framework would sometimes be afraid of any future changes to their moral views as a consequence of their morally suggestible disposition.

Finally, in Section 5.5, I will focus on James Dreier's (2000) second-order desire model. According to Dreier, when an agent makes a moral judgment, her second-order desire to desire to do what she judges to be right can produce a first-order desire in the agent and this explains why the agent is thereby motivated to act in accordance with her moral judgment. The second-order desire model is presumably more plausible than the previous view because the second-order desire would not make an agent scared of any prospective changes to her views. I will, however, argue that the relevant second-order desire can be argued to be constitutive of rationality, which is why the view collapses into a form of internalism.

5.2 The Practicality Option and Its Problems

5.2.1 The Practicality Option

The first externalist account of the connection between moral judgments and motivation I will discuss here is Lillehammer's 'practicality option'. Lillehammer's practicality option principle could be formulated as follows: 'if an agent judges that it is right for her to *f* in circumstances *c*, then if she has a normative reason to *f* in circumstances *c* she will be motivated to *f* in *c* unless she is practically irrational' (Lillehammer 1997, 194). At face value, the practicality option

principle is thus quite similar to the practicality requirement principle that was discussed in Section 3.3.1.

At very general level, Lillehammer endorses Smith's view of how moral judgments can motivate good and strong-willed people. According to Lillehammer's account, it is plausible that good and strong-willed people are usually motivated by their moral judgments and thus, in them, there is a reliable connection between a change in moral judgment and a change in moral motivation. Moral agents act according to their moral judgments because they are aware of what is morally required of them. At the same time, moral agents are also aware of the fact that failing to be motivated in accordance with their moral judgments will make them irrational. Up to this point, the consequences of the practicality option are almost the same as those of the practicality requirement.

However, Lillehammer's proposal can also accommodate exceptional situations in which a rational agent does not have to be motivated to do what she judges to be right, which makes the practicality option different from the practicality requirement. In this way, Lillehammer's claim is far weaker than Smith's assertion that a rational agent is always irrational when she is not motivated by her moral judgment. In contrast to the practicality requirement, the practicality option suggests that 'it is irrational not to be so motivated [only] when one has a reason to be so motivated' (Lillehammer 1997, 194). Lillehammer believes that, under certain circumstances, moral agents have no reasons to act in accordance to their moral judgments. Furthermore, he thinks that, in these situations, agents should not be considered to be irrational which is compatible with the externalist practicality option principle but not with the internalist

practicality requirement principle. There are at least two occasions, Lillehammer claims, where an agent may have no reason to follow her moral judgments about what the right thing to do is.

First, Lillehammer suggests that many moral judgments can be defective, corrupted, or even wrong. It seems that in such cases, it is plausible to think that an agent has no reason to be motivated reliably by any of the previous kind of moral judgments. Lillehammer's illustration of this idea is extreme. He thinks that, for example, if someone has judged that it is morally right to 'drown all handicapped people at birth', it is obviously rational for the agent who has made this judgment not to be motivated to act accordingly (Lillehammer 1997, 194).

And, secondly, Lillehammer also thinks that rationality does not always require moral agents to fulfil their moral obligations. According to some ethical theories, we have obligations to donate a significant amount of wealth to help people who live in extreme poverty. Let us assume that one of these views is right, and an agent comes to correctly judge what she ought to do. Lillehammer would suggest that in such a case, considering the demandingness the moral requirement, it may be rational for the agent not to be motivated by her moral judgment (Lillehammer 1997, 195). Based on the previous two cases, Lillehammer argues that a moral agent should not be claimed to be irrational in all cases in which her judgments do not motivate her.

From the previous discussion, we can extract a view of how moral judgments are supposed to motivate agents in Lillehammer's framework. According to Lillehammer, as rational people, we are usually motivated to do what there are good reasons (that match the content of moral judgments) to do. The motivations of rational individuals, after all, tend to be sensitive to what

reasons they have. There are some exceptional situations where the content of our moral judgments does not correspond to good reasons for actions (such as when our moral judgments are defective, corrupted, wrong or irrational). In these exceptional cases, moral agents do not count as irrational if they fail to be motivated by their moral judgments.

5.2.2 An Objection

In this section, I will critically evaluate Lillehammer's practicality option view. I intend to show that, if we accepted the view as true, we would be unable to explain how both sides in moral disagreements can be expected to be motivated by their moral judgments. I will suggest that, when two agents have made conflicting moral judgments and only one of them has made a correct moral judgment, both agents will presumably be motivated equally by their respective judgments. This, however, cannot be explained by the practicality option principle.

Let us begin from an ordinary case of moral disagreement. It is well known that female genital mutilation is practiced in some communities. According to an estimation of the United Nations Children's Fund made in 2016, millions of young women were still victims of this horrible practice in the countries in which the female genital mutilation continues to be a tradition. Female genital mutilation is widely thought to be harmful because it causes long-term health problems (for example, difficulties in urination and menstruation) for the women who are forced to undergo the procedure when they are young. In addition, female genital mutilation does not seem to be based on any genuine religious requirements, but rather it just violates women's human rights (Nussbaum 1999, 125 and 129). Due to these reasons, female genital mutilation should obviously be considered to be wrong by the public. Despite this, this wrong practice is still supported by some people as they believe that it is a legitimate part of their

special culture. It seems that there thus exists a moral disagreement about whether it is wrong to practice female genital mutilation.

The previous example of a moral disagreement leads naturally to the following question. If two agents make conflicting moral judgments about whether it is right to practice female genital mutilation, will both sides have corresponding motivation to act in accordance with their own moral judgments? Let us suppose that Adele and Elisa have come to make conflicting moral judgments about the rightness of the female genital mutilation: Adele believes that this practice is clearly wrong whereas Elisa believes that this practice is morally permissible when practiced within a culture the tradition of which the practice belongs. Will both Adele and Elisa be motivated by their own moral judgments?

Presumably in this situation, Adele would be motivated to try to stop the practice. Yet, in the same way, intuitively we would also expect Elisa to be motivated by her own judgment, too. At this point, it does not seem plausible to assume that Elisa's judgment is any less effective motivationally than Adele's. It would thus make sense, for example, to expect that Elisa would be motivated to protect the practice against any attempt to stop it. We can then consider whether the practicality option would be able to explain why both Adele and Elisa should acquire motivation in this case.

The practicality options principle will be able to explain why Adele will be motivated to act in accordance with her judgment. When Adele correctly judges that it is wrong to practice female genital mutilation, this judgment presumably corresponds to many good reasons to stop the practice because the practice causes different kinds of harm to women without any

compensating benefits. The practicality option then suggests that, assuming that she is rational, Adele's motivation will track what she has good reasons to do. Consequently, in this situation, Adele will be motivated to try to stop the female genital mutilation according to the practicality option principle.

In contrast, it seems that we would expect Elisa to be equally motivated by her moral judgment even if it is difficult to see why this would be the case if the practicality option principle were true. We can assume that Elisa's judgment is incorrect, corrupt in the relevant sense given that there are no good reasons to preserve the objectionable practice. If following the practical option view, we then accept that the motivations of rational agents track good practical reasons, we should not expect Elisa to have any motivation to try to protect the practice of female genital mutilation.

Yet, this consequence contradicts with what we intuitively think: that there should also be a reliable connection between Elisa's moral judgment and her motivation in the same way as there is between Adele's judgment and her motivation. In a moral disagreement where two agents have made conflicting moral judgments and only one of them could be right, we still assume that both agents will be motivated equally by their respective moral judgments. There is no reason to expect that only the right judgment will motivate the agent who makes it. This is why Lillehammer's version of externalism based on the practicality option principle is so implausible—in order to make sense of the observation that usually both sides in a moral disagreement are motivated by their moral judgments, we should reject this view.

5.3 The Explanation Based on Virtuous People and Its Problems

5.3.1 The Explanation Based on Virtuous People

In the previous section, I discussed Lillehammer's explanation of the reliable connection between moral judgments and motivation and explained why it fails to make sense of how agents are motivated in moral disagreements. In this section, I will discuss Cuneo's more distinctive, sophisticated account of the reliable connection between moral judgments and motivation. Cuneo tries to account for that reliable connection based on how morally virtuous people are motivated. The basic idea of his theory is that, if we could understand how virtuous people are motivated, we would be able to understand how moral agents are motivated to act in accordance with their moral judgments more generally.

Let us first consider how truly virtuous people are motivated. We usually believe that a genuinely virtuous person possesses a number of different virtues. Cuneo then introduces the concept of concerns to further explain how we should understand virtues in the context. According to him, concerns are collections of 'desires, aversions, attachments, interests, cares and the like' (Cuneo 1999, 369). Each virtue than could be claimed to be intimately tied to a different kind of a concern. Thus, if a virtuous person possesses and exhibits the virtues of honesty, compassion, justice, equality and the like, we will know that she will also have concerns that are needed for fulfilling the previous virtues. When an agent has the virtue of honesty, she will desire to tell the truth herself, she cares if other people tell the truth to her and so forth. As desires, aversions, attachments, interests, cares and the like are motivational mental states, the relevant concerns that they constitute can do more than merely grounding the different virtues—they can move the virtuous agents to act in accordance with their moral judgments.

With the help of this picture, the externalists could explain how moral people are motivated without appealing to the *de dicto* desire to do whatever is right. If a moral agent is a virtuous person, when she judges that she should be honest to her friends in a given situation, she will be motivated to perform the relevant honest action due to her concerns that constitute her virtuous character and its honest elements. The combination of a given concern and a moral judgment thus causes the agent to have the corresponding motivation. Likewise, we can expect an honest person to have the desire to tell the truth because her virtue consists of caring about truth or a just person to be moved to promote justice within her society because she cares about treating others fairly.

Based on the previous explanation of how moral motivation is formed, Cuneo can further offer an externalist explanation of why our moral motivations reliably track changes in moral judgments. To achieve this goal, Cuneo returns to Smith's original voting case (Section 3.2). Cuneo (1999, 365) assumes that Sarah is a genuinely virtuous person who initially votes for the libertarians and then changes her mind and votes for the social democrats instead. According to him, there are two different kinds of scenarios that are relevant for evaluating externalism: in one type of cases the relevant motivation is derivative, whereas in another type of cases the relevant motivation is non-derivative (Cuneo 1999, 371-373).

Let us consider derivative motivation first. Let us assume that Sarah has the virtue of benevolence and thus she has a non-derivative concern to 'promote the flourishing of others'. In Sarah's mind, independence plays an important role as a constituent of flourishing life. If Sarah believes that libertarians will promote flourishing by cultivating people's independence,

she presumably will judge that it is right to vote for the libertarians. This judgment, together with the previous concerns and considerations, will then presumably motivate Sarah to vote for them. Now, after her friends manage to convince Sarah that independence will only dissolve the unity of communities, she will no longer believe that independence would contribute to flourishing life. Her new belief, combined with her aforementioned concerns, will cause Sarah to vote for the social democrats instead of the libertarians. Here, it is explicit that Sarah's motivation to support communal ties by voting for the social democrats derives from a non-derivative concern to promote the flourishing of others. Thus, the externalists believe that they can explain the reliable connection between new moral judgments and new motivation by relying on motivations that derive from a more fundamental desire to promote flourishing that is a part of the set of concerns that are constitutive of the virtue of beneficence.

There is also another explanation of what could happen in Sarah's conversion based on non-derivative concerns. In this explanation, we still need to suppose that Sarah is a genuinely virtuous person who cares about fostering friendships, enhancing community relations, and helping people who are badly off. Now, if Sarah is convinced that independence will dissolve communal bonds and this idea conflicts with and is outweighed by Sarah's other values and the qualities which make her a virtuous person, Sarah will not believe that it is right to promote independence, which is the aim of the libertarians. Together with Sarah's judgment that promoting independence will lead to the dissolution of communal bonds, her concern to foster communal bonds and the like is going to motivate Sarah more than her concern for independence. It seems that the content of Sarah's motivational state would in this case be enough to 'defeat her non-derivative concern to promote independence' (Cuneo 1999, 373).

Thus, externalists believe that they can also explain the reliable change between moral judgments and motivation with non-derivative motivations that constitute different virtues.

5.3.2 An Objection

In the previous section, I discussed in detail Cuneo's virtue-based explanation of the reliable connection between moral judgments and motivation. According to Cuneo, virtues and moral judgments together motivate virtuous people to act in accordance with their moral judgments. This claim is that a virtuous agent's virtuous character-traits will produce desires to bring about different states of affairs in accordance with the agent's virtues, which consist of different concerns, and her corresponding moral judgments. The core of Cuneo's view are the relevant *concerns* that I assume to be both emotional and motivational and thus capable of motivating agents to act. It is the different concerns, for example, that enable Cuneo to explain Smith's voting case in terms of either derivative or non-derivative motivation. In this section, I will argue that Cuneo's alternative does not seem to be a compelling alternative because it can explain the reliable connection between moral judgments and motivation only in certain types of virtuous agents. I will show that Cuneo's solution, for example, fails to explain cases where agents who are relatively but not fully virtuous are motivated by their moral judgments.

In order to illustrate the problem, let us reconsider Cuneo's case of a virtuous person—Sarah. For the sake of the argument, we can temporarily grant that Sarah is motivated by her desires based on her virtues and moral judgments as Cuneo has suggested. It is common to think that ordinary virtuous people cannot be completely perfect and thus even Sarah is unlikely to have all virtues to their fullest extent. We can then suppose that, even if Sarah has many different virtues, maybe she still lacks the virtue of generosity, or at least she is not very generous.

Imagine that Sarah then comes across a situation in which some of her relatives owe her a small sum of money and, because of experiencing financial problems, these relatives wish that Sarah could just forget about the debt. Even if Sarah did so, this would not affect her life, as the amount of debt is quite small. In this situation, her relatives would, of course, praise Sarah if she decided to help them. After thinking about it for a while and talking to some of her friends who are more generous than she is, Sarah, perhaps uncharacteristically, judges that it would be right for her to forget about the debt. In this kind of cases that are quite common in our daily life, we would also then expect Sarah to be motivated by her moral judgment. The question then is: can Cuneo's account explain the way in which Sarah presumably would be motivated to act in accordance to her judgment even if we assume that she lacks the virtue of generosity at least in its fullest form?

According to Cuneo's view, a virtuous person's moral motivation relies on the combination of her antecedent virtues and her moral judgments. This means that, neither the virtuous person's antecedent virtues alone nor the moral judgments alone can motivate her. It also means that, in the case just introduced, Sarah cannot be motivated solely by her judgment that it would be right and generous not to require her relatives to pay the debt back since it has to be the virtue of generosity in conjunction with Sarah's moral judgment that will motivate her. The fact that Sarah lacks the relevant virtue of generosity then is a clear problem for Cuneo's account. This is because, on that view, it was exactly that virtue and the concerns of which it consists that were supposed to motivate Sarah to act in accordance to her moral judgment. Without the required virtue, there is no reason why we should expect Sarah to be motivated to let her

relatives not to pay back their debt to her. And yet, we do expect her to be motivated in the previous case.

This is why the revised Sarah's case that I have just described raises a problem for Cuneo's account based how virtuous people are motivated. For example, an ordinary person who helps a stranger in need might have the intention to be kind whilst lacking the virtue of kindness. If we tried to understand cases like this by relying on Cuneo's model, the consequences would be counter-intuitive: many people who we ordinarily think are motivated by their moral judgments should not be expected to do what they think is right. To sum up my response to Cuneo's account briefly, Cuneo's externalist alternative will be unable to explain some cases in which relatively virtuous moral agents who do not have all the virtues to the fullest extent, still have some motivation to act in accordance with their moral judgments.

5.4 The Explanation Based on the Suggestible Disposition and Its Problems

5.4.1 The Explanation Based on the Suggestible Disposition

The third externalist alternative that I will consider in this chapter is David Copp's and James Dreier's. David Copp (1997) has suggested that we can explain the reliable connection between moral judgments and motivation by relying on a disposition to be motivated by moral judgments. Likewise, James Dreier (2000) also has provided a similar account of how morally good people could be motivated as a result of having a disposition which morally good people could be assumed to have. The idea is that, if morally good people have a disposition to be motivated by their moral judgments, such people could be expected to desire to do what they judge to be right.

As an illustration of this view, Dreier (2000, 623) introduces an example of a 'list of foods' in order to explain the *disposition* in question, which according to him morally suggestible people have. Suppose that you want to eat something from a menu which I happen to have. Let us imagine that I will only show you the list occasionally even if most of time, I will keep the list to myself and even change the items on the list whilst offering you hints of what the items on the list are.

Dreier offers two explanations of how your desires could track the items on my list in this situation. Firstly, it might be a mere coincidence that I have just listed what you would like to eat or maybe I knew in advance what your favourites are, which enabled me to put them on my list. In this case, you would have a *de re* desire to eat the foods that appear on my food list. Secondly, it might alternatively be that you have a *de dicto* desire to eat whatever is on my list. If you had such a desire and I gave you hints concerning what might be on the list, you would presumably be able to form derivative desires to eat many of the foods that are listed. As we can notice, the previous two ways of explaining the match between what is on the list and what you want to eat correspond to the internalist and externalist explanations of the match between moral judgments and motivation in Smith's framework.

According to Dreier (2000, 623-624), however, neither of the previous explanations is very plausible. The first explanation seems to be unrealistic since you cannot always happen to desire to eat the items on the list especially if we assume that I am changing the list frequently.
It is obvious that, at times, I will list some foods that you do not desire to eat. Consequently, your *de re* desires to eat the foods on my list will fail to track the change in the content of the list. In contrast, even if the second explanation meets the previous tracking condition (it can

explain why the items on the list and your desires match), that explanation would arguably make you a list fetishist. After all, it would be very odd to desire to eat any foods that just happened to be on the list. As neither of the previous ways can meet the tracking condition and avoid fetishism at the same time, it seems that a new kind of an account is needed for explaining the reliable connection between how my list of the relevant foods changes and what you want to eat.

After evaluating the previous two proposals, Dreier proposes a third explanation which is grounded on the notion of a disposition to desire to eat the items that you believe to be on the list (Dreier 2000, 624). Dreier calls this disposition in his discussion the 'suggestible disposition' for short. This disposition can also be illustrated with avocados. Perhaps you do not like avocados because you cannot stand their strange taste. Yet, if you get a sense that I have added 'avocado' on the menu, the previous disposition would make you inclined to form a desire to taste avocados and have a craving for them. According to Dreier, holding the previous disposition would thus make you a 'suggestible person'. This third explanation which is based on a suggestible disposition seems to meet the tracking condition and additionally, it can also avoid the fetishism objection as the desires that are formed on the basis of the relevant disposition are not derived from any more fundamental desire.

On the basis of the previous discussion, Dreier then develops an account of a morally suggestible person. Just as the suggestible person in the previous example is disposed to desire to eat what she believes to be on the list, a morally suggestible person is disposed to desire to do what she believes to be morally right. On this view, if an agent has a disposition to desire to do what she thinks to be right, this disposition can connect the agent's beliefs and her motivation

in a reliable way. David Copp too has explained why a change in moral judgment generates a change in motivation in a similar way. He suggests that a good and strong-willed person has 'a disposition to desire straightaway to do what [she] believes to be right' (Copp 1997, 50). As a consequence of this disposition, for any action that a moral agent believes that she is required to do, she will be inclined to desire to perform that action, without deliberation or inference, consciousness or not (Copp 1997, 50).

Let us consider Copp's explanation in a little bit more detail. Copp assumes that Smith would agree with him that strong-willed people desire 'as they judge valuable'. He thinks that it is plausible to read this remark to be about good people 'having a disposition to desire straightaway as they judge valuable' (Copp 1997, 50). In this way, if a good and strong-willed person judges that it is valuable to perform an action, he will thus desire directly to perform the action. It would be unnecessary to postulate that a desire to do what the agent judges to be valuable derives from a *de dicto* desire to do whatever she judges to be valuable.

Therefore, from the above discussion we can conclude that some externalists, such as Copp and Dreier, believe that the concept of good and strong-willed person is sufficient on its own to explain the connection between a change in moral judgment and a corresponding change in motivation. Good and strong-willed characters have a disposition due to which they will come to desire to do what they believe to be the right thing to do directly. It therefore seems that some externalists think that they do not need to give up externalism or to accept the relevant *de dicto* desire to explain the reliable connection between moral judgments and motivation. [37]

[37] Some may find it difficult to distinguish Copp's view from Smith's in this case. It seems that Smith also endorses the idea that rational agents have this kind of disposition as he suggests that rational agents have a disposition towards coherence when he explains what grounds the practicality requirement. Yet, in Smith's works, it is always the moral judgments rather than the disposition that generate the relevant

5.4.2 An Objection

In the last section, I introduced Copp's and Dreier's account based on 'morally suggestible disposition' that attempts to provide a plausible externalist explanation of how changes in moral motivation can reliably track changes in moral judgments. This externalist alternative, however, has its own problems as well. According to Dreier himself, the model of morally suggestible people fails to pass a more sophisticated test, which any acceptable explanation of the reliable connection between moral judgments and motivation would need to be able to pass. In order to address this problem, Dreier asks us to consider an example concerning Ursula who is both a utilitarian and a morally suggestible person (Dreier 2000, 631).

As a utilitarian, Ursula cares deeply about the happiness and suffering of everyone equally. Yet, meanwhile, Ursula does not care about the abstract quantities of happiness or suffering, which makes her different from many other extreme utilitarians (Dreier 2000, 631). Furthermore, she is aware of many of the objections to utilitarianism raised by the supporters of the rights-based moral theories. As a consequence, knowing that the utilitarianism is not a perfect theory, Ursula hesitantly agrees with the idea that some individual rights might potentially be needed as side-constraints against the utilitarian pursuit of the maximum amount of happiness. As a morally suggestible person, Ursula is also aware that she would come to be motivated to respect the relevant side-constraints if she came to believe that they existed, even if doing so would weaken her adherence to utilitarianism.

---

motivation in agents who have the rational disposition. Although Copp also tries to rely on the same disposition of rational agents, his view is still a version of externalism. This is because, for Copp, an agent could lack the disposition to desire to do what she judges to be right and yet be fully rational. On Copp's view, the reliable connection between moral judgments and motivation even in rational people is contingent, which makes his view externalist.

Under these circumstances, Ursula needs to find a way to respond to her own uncertainty: she needs a way of thinking about the conflict between her utilitarian view and her tendency to accept some elements of the rights-based theories. Unfortunately, the model based on the morally suggestible disposition is unable to offer a plausible account of Ursula's uncertainty. The problem is that, as a morally suggestible person, Ursula knows that her *disposition to desire to do what she judges to be right* might prompt her to desire to do things which she has currently no desire to do. From Ursula's perspective then, she would feel fearful and be threatened by the prospect of a change in her motivations when she considers the situation in which she comes to accept a rights-base ethical theory in the future.

Let us consider in more detail why Ursula would be scared of the previous prospect. Before Ursula is affected by her morally suggestible disposition, she does not care about respecting other individuals' rights as the rights-based ethical theory requires. At the moment, Ursula cares only about the amount of general happiness and suffering as directed by utilitarianism. Ursula knows that, due to her morally suggestible disposition, her concerns will change if she comes to accept a rights-based ethical theory instead of utilitarianism. If that happened, she would be no longer be motivated to pursue general happiness in the same way as she did before, even if this pursuit is still something that Ursula values greatly. This prospect is what Ursula would be afraid about, and more importantly, she might refuse to investigate the alternative—the rights-based theory so as to avoid any change of her mind.

As a consequence, it seems that Ursula's uncertainty and her potential change of mind from utilitarianism to the rights-based theories cannot be properly made sense of insofar as we think

that she is a morally suggestible person. Because of her disposition, she now knows that she will be disposed to do in the future what she does not want to do then, even if doing so will align with her prospective future moral judgment. This, however, makes the prospect of changing one's moral views an unintuitively scary prospect. If this is right, then we should not believe that an agent's moral motivations track changes in her moral judgment successfully due to the previous kind of suggestible dispositions. This is why the morally suggestible person model does not appear to be a plausible alternative for the externalists.

5.5 The Higher Order Desire Explanation and Its Problems

5.5.1 The Higher-order Desire Explanation

According to Smith, the previous model of the morally suggestible disposition actually requires that rational agents would have a desire to acquire fundamental desires to take right actions (Smith 1997, 115). He also suggests that the previous kind of a desire could at least in principle ground the agents' disposition to desire to do what they judge to be right. Dreier (2000) assumes that Smith has misunderstood the proposal of morally suggestible disposition, which is based on a disposition rather than a desire. However, it seems that Smith's interpretation of the morally suggestible disposition matches another potential externalist account of explaining the reliable connection between moral judgments and motivation. The fourth and the last externalist model I will discuss in this chapter is called the second-order desire view. I will first explain the concept of maieutic ends (Schmitz, 1994, p.228) that will then enable me to outline the externalists' second-order desire account. Then, after this, I will describe Dreier's discussion of the differences between the morally suggestible people model and the second-order desire model. I will also explain why Dreier believes that the latter model is not threatened by Smith's fetishism objection.

First, the maieutic ends. By definition, a maieutic end is an end that is 'achieved through the process of coming to have other ends' (Schmitz 1994, 228; cf. Dreier 2000, 630). In order to explain what this definition means, Schmitz provides an example of choosing a career. Suppose that you want to have a rewarding career and because of this, you want to pursue a career in medicine. Pursuing a career in medicine necessarily requires adopting other ends, such as, relieving the patients' suffering, saving their lives, etc. Effectively, the end of having a rewarding career in this case is also an end to have other ends in the professional life, all of which make the career you end up choosing rewarding. Here, the end of having a career in medicine is a maieutic end because it can only be pursued through having other ends.

We may think that pursuing the maieutic end of having a career in medicine is in this case instrumental to your desire to have a rewarding career. Even if this were the case, the non-instrumental pursuit of relieving the patients and saving their lives would constitute ends that are acquired through coming to have the previous instrumental maieutic end. As a consequence of adopting these necessary ends for pursuing a career in medicine, you no longer pursue a career in medicine instrumentally, but rather you come to care about the career in medicine for its own sake non-instrumentally.

The previous discussion suggests that having a maieutic end requires having some other ends. In this way, a maieutic end resembles a second-order desire that requires holders to have first-order desires. Based on this insight, Dreier thinks that we should be able to explain the reliable connection between moral judgments and motivation by assuming that rational agents have a higher-order desire to desire to do what they believe to be right.

To see how the second-order desire model works, let us consider return to an example I already discussed in Section 3.2. In the voting case, I came to judge that it is no longer right to vote for the libertarians and so I should vote for the social democrats instead. Given we are assuming that there is a reliable connection between moral judgments and motivation, as a consequence of the change in my judgment I should acquire new motivation to vote for the social democrats in accordance with my new judgment. The second order-desire model is then supposed to explain how changes in motivation follow changes in moral judgments in this way. According to this model, under these circumstances, when I judge that it is right to vote for the social democrats, my second-order desire to desire what I judge to be right will produce a first-order desire to vote for the social democrats in me.

At this point, you might doubt that the second-order desire model is just another formulation of the morally suggestible disposition account discussed in the last section. To see why this is not the case, it is important to emphasize that the second-order desire view can avoid the problems of the morally suggestible disposition model. In Section 5.4.2, I explained how the morally suggestible disposition model predicts that potential changes in our moral views would be a scary prospect for us if the latter model were true. By comparison, it can be argued that the second-order desire model is able to explain why such changes should not appear to be something to be afraid of.

To see this, we need to return to the previous example of Ursula (see Section 5.4.2). We can assume that Ursula is still a utilitarian and aware of the objections to utilitarianism raised by the supporters of the rights-based theories. This case allows us to consider if the second-order

desire model too would make Ursula feel afraid of the prospect that she might come to accept the rights-based theory instead of the utilitarianism. According to the second-order desire model, two changes will happen. First, when Ursula comes to believe that the rights-based theories are right in valuing individuals' rights, her second-order desire to desire to do what she judges to be right will generate a first-order desire to respect the rights of others. Second, we can also expect that, because Ursula will no longer accept utilitarianism, her second-order desire will no longer lead to a first-order desire to act in accordance with utilitarianism. Given the fact that changes in her views all happen due to her own desires, Ursula should not feel scared of the prospect of such changes. She will know that, whether or not she changes her mind, she will always have desires she desires to have. This is why, Ursula's case illustrates that the second-order desire model is a more plausible view than the morally suggestible disposition model.

Dreier (2000, 636-637) also suggests that the second-order desire model should be accepted because it can explain how our motivations track changes in our moral judgments in a non-fetishistic way.[38] He first argues that the second-order desire itself is not fetishistic. He claims that the role of the second-order desire to do what one judges to be right is to generate first-order desires to do the right things. Here, the resulting first-order desires to do the right things will be *de re* desires being related to the genuine right-making features of the relevant actions. As I explained earlier in Section 4.2.2, a fetishistic desire is a desire for something that is not worthwhile to be desired and pursued. Here, it is clear that neither the higher-order desire nor

[38] In one paper, Smith (1997, 115-116) argues that, if an agent acquires her motivation through a second-order desire and the relevant moral judgments, the agent is not motivated by the right-making features of her actions. Rather, the agent seems to be motivated by the features she believes to be right-making features (and by those features under that description). Smith thus claims that the externalist second-order desire model is still vulnerable to the fetishism objection. I will respond to this concern below. Smith claims that the externalist alternative of the second-order desire model thus still commits the mistake of a moral fetish. I will respond to this concern below.

the consequent first-order desire is fetishistic on this criterion. This is because it is worthwhile to have desires that correspond to one's moral judgments and also the first-order desires are for things that are worthwhile to desire even according to Smith himself.

Dreier also gives another reason why we should not think that having the relevant second-order desires would be fetishistic. The reason he gives is based on the same idea as Svavarsdóttir's account of instrumental and non-instrumental desires in Section 4.4.1. Thus, according to Dreier, the relevant second-order desires are not always needed to maintain their generated first-order desires they will generate. Once produced, the first-order desires to do the things one judges to be right can function on their own.

We can consider Dreier's own illustration of this point (Dreier 2000, 635 and 637).[39] Imagine that David judges that it is right to stop using chimps in medical research, David's second-order desire will in this situation generate a first-order desire to stop doing so. After this point, David's first-order desire can play a motivating role by itself, and it can even produce other first-order desires. For example, the first order-desire to end using chimps in medical research can generate a new first-order desire to use other substitutes, or a first-order desire to stop other researchers who continue to use chimps in their medical research. That said, all of David's first-order desires in this case are *de re* desires that are not derivative of any other first-order desires and so they cannot be accused of being fetishistic.

[39] According to Dreier, within the second-order desire model, rightness and our thoughts about it do not play a big role in explaining why agents are motivated and continue to be motivated to act in accordance with their moral judgments. For example, in the previous example of David, David's newly acquired first-order desire to end using chimps in research is not conditional on the rightness itself. Also, David could continue to desire to end using chimps in medical research because he cares about the chimps' feelings and suffering rather than because he accepts the rightness of that action.

5.5.2 An Objection

In the previous section, I discussed Dreier's new proposal, which Dreier thinks can explain why changes in moral motivation track changes in moral judgments in a non-fetishist way. In response, I will argue that an argument can be given to the conclusion that rationality itself requires us to have the second-order desire to desire to do what one judges to be right. So, having the relevant second-order desires would be a feature constitutive of being a fully rational agent and thus any agent who failed to have the relevant second-order desire would thereby be irrational. If this is right, then it looks like Dreier's proposal collapses into a form of internalism. After all, in this case there would be a necessary connection between moral judgments and motivation in all rational agents via their second-order desire that would be constitutive of their rationality.

My discussion will proceed in three steps. First, I will begin by explaining Michael Smith's claim that the rationality of a given set of motivations can generally be thought to consist at least in part of how unified and coherent the motivations in that set are. If the previous claim is true, then a rational agent's disposition towards unity and coherence will in many cases generate more general desires that will unify the agent's existing set of desires. Secondly, I will introduce Sayre-McCord's objection to Smith's view. According to him, adding more general desires to an existing set of desires will not always make the set of desires more rationally preferable and thus an agent cannot be made more rational in this way. I will briefly explain why Sayre-McCord's objection is not plausible.

Thirdly and finally, I will argue that a rational agent's tendency towards unity and coherence will generate the relevant higher-order desire relevant in this context in her—the second-order

desire to desire to do what one judges to be right—to ensure that the agent will be motivated by her moral judgment. It thus seems that the second-order desire model will entail that there is a necessary connection between a rational agent's moral judgments and her motivations. At this point, this result should be enough to show that Dreier's second-order desire model indeed collapses into a version of internalism.

Let us take the first step and consider Smith's view of rationality. Smith believes that if an agent is fully rational, she must have a systematically justifiable set of desires. Such a set of desires should be informed, coherent and unified (Smith 1994, 156-161; Smith 1996a, 160).[40]
To clarify this view, I will explain these three features in turn next.

Here, informedness means that a rational agent has only desires that are based on her knowing all the relevant facts about the situation she is in. A quick example might be helpful. Consider a rational agent who is driving a car and feels thirsty. In Bernard Williams's famous case (Williams 1981, 102), there is a thirsty agent who has a glass full of transparent liquid in front of him. In this case, the actual agent will desire to drink from the glass. Williams argues, however, it would not be rational for the agent to do so. This is because, if the agent knew all the facts about the situation, including the fact that there is gin in the glass instead of water, the agent would no longer desire to drink from the glass. This illustrates how rational desires, need to be informed, they are the desires we would have if we know all relevant facts.

[40] In addition to having the maximally coherent and unified set of desires, a fully rational agent must also meet two other requirements of rationality: 1) they must not suffer from any physical or psychological disturbances, and 2) they must have all relevant true beliefs and no false beliefs. As these two conditions are not relevant to the present topic, I have set them aside here.

Let us then consider the second feature. The coherence of a set of desires that can be thought to consist of the fact that the desires that belong to the set do not pull the agent to different, incompatible directions. For example, a rational agent does not suppose to get cooler by opening the window, and warmer turning the heater higher at the same time. Likewise, the unity of desires consists of the fact that the agent's different desires support one another (see the discussion below).

Smith then argues that a fully rational agent's disposition towards coherence and unity will under some circumstances change her desires. The rational disposition towards coherence and unity can, for example, produce general desires that will support more specific desires and also these new general desires will in some cases also destroy some of the previous specific desires that do not fit them. For example, you may have a set of desires concerning which methods of transport you would like to use for travelling. This set can include a desire to take a bus to work, a train when travelling to other cities nearby, and a desire to fly when going abroad. At least initially, these specific desires need not be derived from any more general desires.

Smith then guides you to ask yourself whether the previous specific desires would be more systematically justifiable if a more general desire which could justify and explain those specific desires were added to your psychological make-up. For example, you could add a general desire for choosing the most affordable and convenient means to go to where you want to go you in your set of desires. This general desire could justify the previous set of desires by explaining why you would not want to travel to a faraway country by bus, for example, as it is obvious that traveling by plane to another country is often more convenient and more economical. With

the new added general desire, the relevant set of desires will be more systematically justifiable and thus also more unified and hence also rationally preferable, according to Smith.

let us then consider Sayre-McCord's objection to what I have just explained. He casts doubt on the view that, in many cases, a more general desire will make a given set of desires more coherent, unified, and therefore also more rational (Sayre-McCord 1997, 75). To see this, he asks us to consider a case of choosing an ice cream. If we suppose that Smith's view is true, then, if I have a desire for coffee ice cream, my set of desires could be argued to exhibit more coherence and unity if a more general unconditional desire for ice cream were added to my current desiderative profile. My set of desires could be more coherent and unified because the newly added general desire would be able to explain why I desire to enjoy coffee ice cream.

In this situation, eating coffee ice cream will satisfy both my desire for coffee ice cream and my general unconditional desire for ice cream. Sayre-McCord then raises an objection by claiming that it is not plausible to think that satisfying the previous two desires would make me any more rational than how rational I would have been had I only satisfied my original desire. So, he thinks that adding more desires, for example, a more general desire to a desiderative profile cannot itself enhance an agent's rationality as Smith suggests.

It seems that the crucial dispute between Sayre-McCord and Smith is over whether adding a more general desire to an agent's desiderative profile can make the agent more rational. I think that Sayre-McCord is right in claiming that merely satisfying more desires cannot itself make an agent more rational. Yet, the number of satisfied desires is not what Smith's view of rationality is based on. The key point of his view is that sometimes adding a more general desire

to an existing set of desires can make the set more coherent and unified. This is the real reason why Smith would think that adding a more general desire to eat ice cream can in the previous case make my desire set more rationally preferable.

In the previous case, it is supposed that I initially have a desire to have coffee ice cream. Usually, my desire to have coffee ice cream will move me to get it when it is practically available. Despite this, if I only had this one desire, I would presumably often ask myself: why do I choose to have coffee ice cream rather than other flavors, or even other kinds of dessert (Smith 1997, 94)? The desire to have coffee ice cream itself does not seem to be able to answer this question. Yet, if a general, unconditional desire to eat what I enjoy eating, for example, were added to my desire set, this more general desire would be able to explain my specific desire to have coffee ice cream. The desire to eat coffee ice cream would no longer appear to be arbitrary, but rather it would be well-supported by the more general desire. In this way, my desire set has turned out to be more coherent, unified, and thus more rationally preferable.

Analogously, we can argue that the second-order desire to desire to do what one judges to be right would be required by rationality, exactly in the same way as the general desire is required in the case above. Consider, for example, an agent who has various moral desires, desires to treat her friends very well, to keep her promises, to not cause physical harm to anyone and etc. These things are all distinct from one another because they are all about different matters. However, a second-order desire to desire to do what one judges to be right would in this case justify and explain why the agent has the previous desires to do all the different things that she

also judges to be the right things to do. As a result, it could be argued that the desiderative set also becomes more rationally preferable as a consequence.[41]

To wrap this section up, I have argued that a relevant second-order desire discussed by Dreier would be required by the fundamental constituents of rationality—coherence and unity. For a rational agent, having a second-order desire to do what one judges to be right (that will produce a first-order desire to do such right things) should thus be a matter of fulfilling a constitutive requirement of rationality. Without that desire, the agent would not really count as a rational agent in the first place. This means that, insofar as one is a rational agent who has the relevant higher-order desire, there will be necessary connection between one's moral judgments and motivation. As a result, it seems that what Dreier proposes (see Section 5.5.1 above) is not an entirely new externalist solution. Rather, as we were led to believe via a further explanation of Smith's understanding of rationality, the idea of second-order desire is as something that explains the reliable connection between your moral judgments and motivation is actually compatible with Smith's version of conditional internalism based on practical rationality. The second-order desire model has actually collapsed into a form of internalism. This is why there still is not a plausible externalist account of the reliable connection between our moral judgments and motivation.

[41] As in the ice cream case above, without the higher-order desire the agent could ask herself just why she should desire the things she desires. That is, without the second-order desire, the relevant first-order desires might appear to be arbitrary to herself. Yet, with the second-order desire, the agent can make sense of the first-order desires as the desires she desires to have.

5.6 Conclusion

In Chapter 5, I have discussed four representative externalist alternative explanations of the reliable connection between moral judgments and motivation. I have also tried to argue in response that none of these explanations are very plausible.

In Section 5.2.1, I discussed Lillehammer's practicality option account which suggests that moral agents are reliably motivated to act in accordance with their moral judgments because they are sensitive to good normative reasons for action. On this view, an agent can still be rational if she is not motivated by corrupt and wrong moral judgments that do not reflect good reasons for action. In response, I argued in Section 5.2.2 that the practicality option cannot explain the way in which both sides in moral disagreements are supposed to be motivated by their moral judgments.

In Section 5.3.1, I discussed Cuneo's proposal based on the notion of genuinely virtuous people. Cuneo claims that virtuous people are motivated to act in accordance with their moral judgments because of their virtues that consist of various substantial concerns. In Section 5.3.2. I have argued that, though seemingly promising, Cuneo's alternative is unable to explain certain situations where an agent who is not wholly virtuous is motivated to act in accordance with what she judges is right.

In Section 5.4.1, I discussed the externalist model of morally suggestible people that was first introduced by Copp and discussed further by Dreier. The morally suggestible person model suggests that rational agents have a disposition to desire to do what they judge to be right. This disposition can then be used to explain why motivation and moral judgments are reliably

connected. In Section 5.4.2, I argued, following Dreier, that the previous model would make the prospect of changing one's moral view something to be afraid of in an unintuitive way.

In Section 5.5.1, I discussed Dreier's own second-order desire model. This model suggests that we can explain the reliable connection between moral judgments and motivation by referring to a second-order desire to desire to do what one judges to be right. When an agent judges that something is the right thing for her to do, her second-order desire to do what she judges to be right produces a first-order desire to do the right thing in her. Yet, in Section 5.5.2, I argued that having the relevant second-order desire to desire to do what one judges to be right could be thought to be required by rationality itself. If this were right, Dreier's second-order desire model would collapse into a form of internalism, which would not be acceptable for the externalists.

Chapter 5 has been the last part of the main argument that started in Chapter 3. Chapters 3-5 constitute my defence and development of the fetishism argument. I first introduced the fetishism argument in Chapter 3, then responded to the externalist objections based on defending the *de dicto* desire to do whatever is right in Chapter 4, and finally critically explored the other externalist ways of explaining the reliable connection between moral judgments and motivation in this Chapter 5. All these three chapters have served a more fundamental role, which is to prove that all forms of externalism are implausible because they fail to provide a plausible explanation of the reliable connection between moral judgments and motivation. This is why I believe that we should ultimately reject externalism and consider some form of internalism to be true instead. This leads to the question I will explore in the next two chapters:

if we should accept some form of internalism, which form would be the most plausible one? I will begin my exploration of this topic next.

Chapter 6: Defenses of Non-constitutional Internalism and Unconditional Internalism

6.1 Introduction

As I just mentioned, Chapters 3-5 were my exploration of whether we should accept externalism or some form of internalism as the more plausible view. In those chapters, I argued that we should reject externalism because all the externalist objections to the fetishism argument fail. As internalism and externalism are the only two alternatives, the consequence of the previous chapters is that some form of internalism must be true. This thought then naturally leads to another equally important question: which form of internalism is then the true one?

That question will be answered in Chapters 6-7. Beginning from this chapter, in the rest of this thesis I will focus on investigating which form of internalism is the most plausible view. Since I have already introduced all the existing forms of internalism in Chapter 2, I will take each theoretic 'choice-point' (strong/weak, unconditional/conditional, direct/deferred and constitutional/non-constitutional) and consider which one of the two alternatives can be eliminated. So, I will first discuss unconditional/conditional internalism in Chapter 6 and constitutional/non-constitutional internalism in Chapter 6. I will then examine strong/weak internalism and direct/deferred internalism in Chapter 7. I believe that, in the end of the whole discussion, we will be able to see what the most plausible version of internalism is.

In this chapter, I focus on two pairs of the internalist alternatives: constitutional/non-constitutional and unconditional/conditional. It will first explain how the constitutional (*de re*) forms of internalism and externalism and non-constitutional (*de dicto*) forms of internalism and

externalism are about different subject-matters. The former views are about the nature of certain mental states, whereas the latter views are about how the words 'moral judgments' are used. I will then argue that already the arguments of Chapters 3-5 rule out the constitutional (*de re*) forms of externalism.

After this, in Section 6.3, I will introduce a new formulation of unconditional internalism based on the notion of dispositional desires. I will argue that this new formulation is not vulnerable to all the traditional objections to unconditional internalism. The defenders of the new view can accept that there are agents who have no motivation to act in accordance to their moral judgments as long as they have a desire to do so, i.e. they are in a dispositional desire-like state that in the standard conditions produces motivation. Thus, we can outline a new, more plausible version of unconditional internalism in terms of dispositional desires, or so I will argue later.

6.2 Reconsidering Non-constitutional Internalism

In this section, I will reconsider and evaluate non-constitutional internalism, the view that was already introduced earlier in Section 2.7. In Section 6.2.1, I will first remind my readers of some of the key elements of non-constitutional internalism. Then in Section 6.2.2, I will discuss four combinations that can be formed from the *de re and de dicto* forms of internalism and externalism. These combinations of views include 1) *de dicto* internalism and *de re* externalism; 2) *de dicto* externalism and *de re* externalism; 3) *de dicto* internalism and *de re* internalism; 4) *de dicto* externalism and *de re* internalism. It will eventually turn out that we should reject the combinations of 1) *de dicto* internalism and *de re* externalism and 2) *de dicto* externalism and *de re* externalism on the basis of the arguments already presented in Chapters 3-5 as these two combinations will be fetishist as pointed out in the previous chapters. I will then remain neutral

about whether we should accept *de re* internalism with *de dicto* internalism or externalism.

6.2.1 A Re-examination of Non-constitutional Internalism

In Section 2.7, I outlined non-constitutional internalism and how it differs from the more traditional constitutional forms of internalism. Constitutional internalists traditionally treat moral judgments as certain unique kind of psychological states that are connected to motivation in a certain internal way due to what kind of psychological states they are essentially. The constitutional externalists, by comparison, have traditionally denied that the nature of moral judgments would be in any special way connected to motivation. Despite having different views of how exactly moral judgments lead to motivation, philosophers in this debate have thus all assumed that it is the nature of moral judgments that determines whether the corresponding motivation will exist necessarily in the agents who make these judgments. This is why constitutional internalism can also be called *de re* internalism—it is a view of what certain mental states are like. In contrast, because it has nothing to do with the nature of moral judgments, *de dicto* internalism does not provide any insight of what the mental states of an agent are like or what these mental states' essence is. Instead, *de dicto* internalism is merely a view of when the term 'moral judgments' can be applied to a mental state of an agent (Tresan, 2006, 2009a, 2009b).

For example, consider what the word 'planet' means: a celestial body that is accompanied by a star. Here, the concept 'planet' seems to tell us nothing about the nature of the objects we call 'planets'. All we can know from this definition is that, if a celestial body is accompanied by a star, we can call it 'a planet'. Likewise, according to *de dicto* internalism, the words 'moral judgment' can apply to a mental state only if that mental state is accompanied by motivation.

Thus, although *de dicto* internalism is still neutral about the nature of moral judgments, we can still infer from it the following internalist claim in which the modal necessity operator has a wide scope: necessarily, if an agent has made a genuine moral judgment, she will have some corresponding motivation.

One important observation to make about *de dicto* internalism is that this view is completely neutral about the nature of the mental states that are called 'normative judgments'. We have seen that *de re* views and *de dicto* views are about different subject-matters. *De re* views are about certain kind of mental states, whereas *de dicto* views are about how the words 'moral judgments' are to be used. *De dicto* views can thus leave any controversies about the nature of moral judgments untouched. Moreover, it turns out that *de dicto* internalism would be true even if the mental states we call 'moral judgments' were always accompanied by motivation only via certain external, contingent mechanisms.[42] After all, when these mental states are not accompanied by the relevant motivation, it is just that we would not call them 'moral judgments' according to *de dicto* internalism.

It is true, of course, that *de dicto* internalist, most notably Tresan (2006, 149), generally tend to also reject different forms of *de re* internalism. Despite this, Tresan and other non-constitutional internalists have not really made direct arguments against *de re* internalism but rather they only

[42] In Section 2.7 where I introduced *de dicto* internalism, I mentioned that, *de dicto* internalism is not vulnerable to the objections based on amoralism. This actually further implies that *de dicto* internalists can accept the externalist explanations of how moral judgments are connected to motivation. In Chapters 3-4, I mentioned multiple times that the externalists believe that moral judgments motivate via external mechanisms that involve an additional desire to do whatever is right and the like. It is thus no wonder that many *de dicto* internalists are often committed to that type of externalist mechanisms. I will explain this view in detail in Section 6.2.2 below.

tend to argue against them indirectly. For example, Tresan first begins from the simple internalist intuitions that can be accepted by both *de dicto* and *de re* internalists.

These shared internalist intuitions suggest that amoralists do not make genuine moral judgments precisely because they lack the relevant motivation. On the basis of this, Tresan first suggests that there is enough evidence to support *de dicto* internalism (Tresan 2006, 149). As a view of whether we would call Patrick's mental states moral judgments in the actual world, *de dicto* internalism only takes into account whether Patrick has motivation here.[43] Our intuitions about whether we would call the mental state Patrick is in a 'moral judgment' in the actual world seems enough to enable us to know whether that term applies only to mental states that are accompanied by motivation. Otherwise, we would not be able to tell the difference between moral judgments and other mental states in the actual world.

Tresan then suggests that, in contrast, the *de re* internalists have not yet provided enough evidence for their view. This is because, according to him, in order to establish *de re* internalism, we would first have to find a mental state that can be correctly called a 'moral judgment' from the actual world. This, however, is not enough. After this, we would also have to investigate this very mental state and its nature not only in our world but also in all other worlds too so as to see whether the nature of that mental state is such that it is always accompanied by motivation across different possibilities.

This investigation would enable us to determine whether a part of the essence of that very mental state is to provide motivation. *De re* internalists, however, have according to Tresan not

[43] See Svavarsdóttir (1999, 176-177) and Section 2.5.2 of this thesis for Patrick's case.

carried out such an investigation and so he argues that currently we have no good reasons to believe that *de re* internalism is true. In this situation, Tresan seems to believe that our internalist intuition at most supports *de dicto* internalism since it is a less controversial and a weaker view.

6.2.2 Implications of Non-constitutional Internalism

As I discussed in the previous section, non-constitutional (*de dicto)* and constitutional (*de re)* internalism are views about different subject-matters. This entails that it is hard to find a common ground on the basis of which we could compare and then justify either *de dicto* internalism or *de re* internalism. Yet, we may even not need to do so because Tresan himself has suggested that *de dicto* internalism is in principle compatible with both *de re* internalism and externalism. This implies that *de re* and *de dicto* internalism are not mutually exclusive views. So, I suggest that we can assume that *de dicto* internalism and *de re* internalism could be at least in principle both true at the same time. This would allow us to consider different combinations of views based on *de re* and *de dicto* forms of internalism and externalism. There are at least initially four possible combinations of view including 1) *de dicto* internalism and *de re* externalism; 2) *de dicto* externalism and *de re* externalism; 3) *de dicto* internalism and *de re* internalism; 4) *de dicto* externalism and *de re* internalism. In this way, if there would be unacceptable consequences to any of the combination, of course, we would doubt whether the constituent views would be problematic as well.

As these combinations have close connection with the previous discussions of internalism and externalism, I can disclose my conclusion before engaging in more detailed discussions. Below, I will first argue that we should reject 1), the combination of *de dicto* internalism and *de re* externalism. This is because this combination would commit us to the externalist explanations

of the reliable connection between moral judgments and motivation that were already criticized in Chapters 3-5. For the very same reason, I will also reject 2), the combination of *de dicto* externalism and *de re* externalism. I will remain neutral between the other two combinations, 3) and 4) that are both committed to *de re* internalism.

In this section, I will discuss the plausibility and the implications of the previous combinations. The first combination I will discuss combines *de dicto* internalism with *de re* externalism. Even if *de dicto* internalism says nothing about what kind of mental states moral judgments are, *de re* externalists usually assume that moral judgments are ordinary belief-like mental states. This means that the defenders of this combination of views must also grant that moral judgments do not produce any motivation unless they do so with the help of some external mechanism, contingently. Then we can combine to this idea *de dicto* internalism about how to apply the term 'moral judgment' correctly—that is, that description can only be applied to a given belief if it is accompanied by motivation. This first combination of views can thus be thought to claim that, when ordinary belief-like mental states are accompanied by motivation produced by the relevant external mechanisms in the agent, we call these belief-like mental states 'moral judgments'.

But if we accept this combination of *de dicto* internalism and *de re* externalism as recommended by Tresan (2009b, 194), this would make it difficult to explain the reliable connection between moral judgments and motivation in a plausible way. At this point, it is important to notice that in Chapters 3-5, I already argued that all externalist attempts to explain the reliable connection fail. Since I have discussed *de re* externalism carefully in the previous chapters, I will only go through my conclusions briefly here. In Chapter 4, I argued that the externalists' attempts to

defend the *de dicto* desire to do whatever is right failed to avoid the fetishism objection. Likewise, in Chapter 5, I argued that all other externalism-friendly explanations of the reliable connection between moral judgments and motivation fail too for different reasons. If these arguments were sound, they equally show that the previous combination of *de dicto* internalism and *de re* externalism must fail too.

Since the externalist ways of explaining how moral judgments motivate have to be adopted by those who endorse *de re* externalism, any combination of views that is committed to *de re* externalism will subsequently inherit all the problems of *de re* externalism. Thus, given the arguments already presented in this thesis, we should conclude that the 1), the combination of *de dicto* internalism and *de re* externalism, should not be accepted. Furthermore, for the very same reason, the 2) combination of *de dicto* externalism and *de re* externalism should be rejected, as well.

We can then consider another possible combination of views, which is 3), the combination of *de dicto* internalism and *de re* internalism. Here it is important to notice that *de dicto* internalism is not a view that conflicts with *de re* internalism because these views are about different subject-matters. Rather, if we assume that both *de dicto* internalism and *de re* internalism are true at the same time, then the truth of the latter view seems to be able to provide a nice explanation of why the former view would also be true. This is because *de re* internalism about the nature of moral judgment itself can naturally be used to explain why we would use the term 'moral judgment' in a certain *de dicto* internalist way. Consider how we understand the term 'water' and its relation to the substance water. Long time ago, we interacted with a transparent, tasteless and odorless liquid substance with a certain essence and we then wanted to give that

liquid a name. The fact that we wanted to name a certain substance we interacted with 'water' explains why we do not call anything that is not H_2O water (Kripke 1980, 128). In the very same way, it could be argued that the fact we wanted to give the name 'moral judgment' to a certain kind of mental states that can motivate us explains why we do not call mental states that are not accompanied by motivation 'moral judgments'.

However, it might also be possible that *de dicto* internalism is false, even if *de re* internalism were true as I have argued in this thesis. If *de dicto* externalism were true in this way, this would lead us to the last combination of views which is 4), the *de dicto* externalism and *de re* internalism. This combination would entail that we do call at least some mental states moral judgments even when they are not accompanied by relevant motivation and yet the nature of moral judgments would be such that they are internally related to motivation. In order to see whether this combination would be coherent, we can consider, for example, conditional forms of *de re* internalism (see Section 2.5.3 for details).

According to conditional forms of *de re* internalism, moral judgments produce motivation only when certain conditions are satisfied. It further entails that agents can hold genuine moral judgments without having any motivation—this can be the case when the relevant conditions are not satisfied. Take the case of a depressed person who is in a mental state, the essence of which is such that it produces motivation in the agent when she satisfies the relevant conditions. It is at least a possibility that the term 'moral judgment' could be used in such a way that we would apply this term to the depressed agent's mental state even when she has no motivation. This possibility shows that there are certain forms of *de re* internalism that are compatible even with *de dicto* externalism. Hence, even if *de dicto* externalism were true, this would still not

necessarily mean that *de re* forms of internalism would have to be false. Since *de dicto* externalism is compatible with *de re* internalism, it can also be concluded that the combination 4) is at least a possibility that should not be ruled out.

After this brief discussion of the four combinations that we get from the *de re* and *de dicto* and versions of internalism and externalism, it is time to sum up conclusions of my discussions so far in this chapter. I have claimed that the first two combinations—1) *de dicto* internalism and *de re* externalism and 2) *de dicto* externalism and *de re* externalism—fail for the same reason, which is that they are vulnerable to the same general fetishism objections that I made to externalism in Chapters 3-5

I then proceeded to discuss the other two combinations 3) and 4). I suggested that one advantage of 3), the combination of *de dicto* internalism and *de re* internalism, is that the truth of *de re* internalism would at least partially explain why *de dicto* internalism might be true as well. In this way combination 3) seems unproblematic to me.

The last combination—4), i.e. *de dicto* externalism and *de re* internalism—is based on the assumption towards the combination 3) assuming that *de dicto* internalism is false in the way suggested by some of the famous counterexamples to internalism. Even if this were the case, this would not be a problem for my defence of *de re* internalism as it turned out that there are forms of *de re* internalism that are compatible with *de dicto* externalism, too. Since combinations 3) and 4) appear to be equally plausible at this point, I believe that I should remain neutral between them. To put this bluntly, my main interest in this thesis is to defend a form of *de re* internalism. If that view is compatible with both *de dicto* internalism and *de dicto*

externalism, I can remain neutral about which one of those views is true in addition to *de re* internalism. Therefore, I will move on to continue my investigation into other forms of (*de re*) internalism.

6.3 A Re-evaluation of Unconditional Internalism

The previous section argued that, assuming that the arguments against externalism in the Chapter 3-5 were successful, then we should accept some form of constitutional internalism. Furthermore, in Section 6.2, I also argued that, if we accept a form of constitutional internalism, we can then remain neutral about whether or not some form of non-constitutional internalism about the way in which the term 'moral judgment' is used is also true. We can then proceed to the next choice-point, where we can again continue to consider which form of internalism is the most plausible one.

Here, in this section, I will focus on answering that question by exploring two other contrasting forms of internalism: unconditional internalism and conditional internalism. In Section 6.3.1, I will first outline the key differences between these two views, and I will also remind the reader of the reasons that have moved many philosophers from unconditional internalism to conditional internalism. In Section 6.3.2, I will then explain away the externalist counterexamples which are supposed to be the main reasons for rejecting unconditional internalism. After that, I will provide a new version of unconditional internalism formulated in terms of dispositional desires. In favor of this form of unconditional internalism, I will then put forward two thought experiments in Section 6.3.3. I will argue that, when depressed people and amoralists make genuine moral judgments, we have sound reasons to think that they have at least the relevant dispositional desires to act accordingly. Likewise, when the depressed agents

and amoralists lack these dispositional desires, we have similarly good reasons to doubt whether they have made genuine moral judgments in the first place. In Section 6.3.4, in order to respond to an important objection to the resulting view, I will consider the question of how to distinguish occurrent moral judgments from dispositional ones. As a response to this challenge, I will outline two ways in which occurrent moral judgments can manifest themselves without the requiring corresponding motivation.

6.3.1 From Unconditional Internalism to Conditional Internalism

In Sections 2.5.1 and 2.5.3 above, I already introduced unconditional and conditional internalism. Unconditional internalism is a straightforward view that is meant to capture ordinary people's internalist intuitions at face value. As its name suggests, unconditional internalism entails that moral judgments must always lead to at least some motivation whenever agents make such judgments. In the same section, I formulated a typical version of unconditional (weak) internalism.[44]

> *Unconditional (Weak) Internalism*: Necessarily, if an agent judges that it is right to φ in circumstances C, she will always have at least some motivation to φ in C.

Traditionally, unconditional internalism has been thought to be a problematic view because it seems unable to respond to the objection based on the possibility of amoralism and the other alleged counterexamples. Take the depressed mother again as an example (see Section 2.5.2 and Mele 2003, 111). The depressed mother used to make genuine moral judgments, which can

[44] It needs to be pointed out that here, and in the rest of this thesis, I understand different formulations of internalism in a constitutive way.

be proved by the fact that she helped her ailing uncle for many years. Yet, as the mother suffers a loss of her husband and her children, Mele also seems to be right in supposing that the mother at this point ceases to have any motivation to act according to that very same moral judgment. Despite this, there are no good reasons to think that the depressed mother has lost her previous moral judgment according to which aiding her uncle is the morally right thing to do. This is why our intuitions about the depressed mother's case seem to show that there must be something wrong with the previous form of unconditional internalism.

Internalists have, of course, always been aware of the previous problem, which has led them to pursue two main strategies in response. The first strategy has been to attempt to argue that individuals like the depressed mother do not really make genuine moral judgments. Here, the internalists have, for example, adopted Hare's idea of the inverted commas moral judgments to explain away the fact that the depressed mother does not seem to be motivated by her moral judgments. Hare originally used the idea of inverted commas moral judgments to refer to non-moral judgments that only resemble moral judgments (Hare 1952, 124). Though the agents who make inverted commas moral judgments use the very same words to express these judgments, the judgments they make are not genuine ones with the standard moral content. Rather, the inverted commas judgments reflect that an agent accepts what other people believe as right or wrong. They are only about what is 'right' and 'wrong' in this different sense.

It might then be argued that the depressed mother makes merely inverted comma moral judgments when she is depressed, which would, from the internalist perspective, be compatible with the idea that she is not motivated accordingly. Although her judgments appear to be exactly the same both before and after the depression, on this view their content has changed due to

depression from moral to non-moral content. It could be argued that the depressed mother's inverted commas judgment that aiding her sick uncle is the 'right' thing to do only resembles her previous thought (she used to believe that this is the case). Thus, if Hare's view is right, the depressed mother does not make a genuine moral judgment, which is compatible with her not having any motivation and so a successful counterexample has not been put forward, or so it could be argued.

Yet, many internalists have taken the previous objection to be a sufficient reason to reject unconditional internalism and so they have pursued other internalist ways to avoid the problem. The second strategy grants that individuals like the depressed mother make genuine moral judgments even if they have no motivation to act accordingly. This approach then argues that this is possible because moral judgments only produce motivation under certain conditions. A number of internalists have then tried to formulate what these conditions are in a way that would exclude the relevant counterexamples such as the depressed people. In contrast to unconditional internalism, conditional internalism can be formulated, for example, in the following way:

> *Conditional (Weak) Internalism*: Necessarily, if an agent judges that it is right to φ in circumstances C, either she has at least some motivation to φ in the circumstances C or she fails to satisfy certain conditions D.

The contrast between conditional internalism and unconditional internalism thus reveals that these views rely on very different strategies when dealing with alleged externalist counterexamples to internalism. The unconditional internalists tend to deny that the individuals in these cases, such as the depressed mother, have made genuine moral judgments, whereas the

conditional internalists try to find internalism-friendly ways of explaining how these individuals continue to make genuine moral judgments even if they are not motivated to act accordingly. Yet, the latter strategy is indeed a more popular one than the former because it does not require accepting that a person's moral judgments change from being genuine moral judgments to non-moral judgments, for example, when she becomes depressed. This means that conditional internalism seems like a more promising and flexible response for dealing with the alleged counterexamples.

6.3.2 A New Version of Unconditional Internalism

The previous section explained how the counterexamples have been the crucial reason for why most internalists have defended conditional internalism. This is to say that most philosophers have preferred conditional internalism because they have believed that unconditional internalism is unable to offer a plausible enough explanation of the allegedly problematic cases. Whether this conclusion is true, however, almost entirely depends on how we understand 'motivation' in the relevant formulations of internalism. If we were to reconsider the concepts of 'desire' and 'motivation' that are used when formulating unconditional internalism, new conclusions might be drawn accordingly, or so I will argue next.

So far, I have not explicitly distinguished between motivation and desires as their minor subtle differences between the two would have not influenced the previous discussions. What I have assumed so far is that the motivation we have been discussing consists of a desire-like state to act in accordance with a given moral judgment. I have thus also assumed that the relevant desire-like states are occurrent in the sense that they move an agent to act at least when the desire is not outweighed by other, even stronger desires to perform other actions. As a

consequence, I have supposed that being motivated to act in a certain way thus is the same thing as desiring to do that action. Consider an example that will illustrate this view. Imagine that I have a desire to help a stranger in need after judging that this is the right thing to do. Given the previous assumption, in this case unless I have another outweighing desire to do something else instead, saying that I desire to help the stranger is equivalent to saying that I am in a state of being motivated to help the stranger.

What I want to explore next is whether drawing a clearer distinction between desires and motivation would help us to avoid the seemingly powerful counterexamples to unconditional internalism. When in Section 2.2, I first discussed what moral motivation is, I mentioned that desire-like states have the world-to-mind direction of fit, which means that it is a part of the functional role of these states that they aim at changing the world in order to make it fit what the agents desire the world to be like. Desires thus explicitly have a functional role in issuing actions, even if this function may not always be invoked instantly. So, in order to respond to the externalist objections, one option would be to formulate unconditional internalism, not in terms of motivation, but rather in terms of desires. This might be useful even if motivation can always be supposed to be an occurrent state, desires are not necessarily occurrent. The crucial thought is that desires can also be understood in dispositional terms and so the idea is that perhaps this quality they have could be used to formulate a more plausible version of unconditional internalism.

As it happens, many internalists have already understood desires in dispositional terms. For example, according to Blackburn (1998, 67), holding a desire that is expressed by a sentence that contains moral vocabulary typically reflects the fact that the agent has a certain set of stable

dispositions to be motivated to act in certain ways. Likewise, Smith takes the relation between desires and dispositions to be moved to be a closed one. On his account, desiring to do a certain action is equivalent to having a set of dispositions (Smith 1994, 113). He believes that, if we desire certain outcomes to obtain, we are disposed to have motivation to act in a way that would lead to those outcomes. As Smith puts it, such dispositions include 'dispositions to act, dispositions to feel pleased or disappointed,' dispositions to conduct deliberation, dispositions to respond to questions and so on (Smith 2004, 97).

If the previous line of thought is along the right lines and we can emphasize the desires' ability to invoke motivation to act when possible, then we can formulate a slightly different form of unconditional internalism by relying on desires understood as dispositions to be motivated:

> *Unconditional (Weak) Internalism with Dispositional Desires*: Necessarily, if an agent judges that it is right to φ in circumstances C, she will always have at least some desire (i.e., a disposition to have motivation) to φ in circumstances C.

This form of unconditional internalism with dispositional desires can still explain the reliable connection between moral judgments and motivation that obtains usually. The relevant dispositions introduced above, after all, can explain why moral motivation tends to steadily align with moral judgments. To see why this would be the case, we can consider a simple example. As a movie-lover, I make the judgment that to watch an old movie, for example, Sean Connery's *From Russia with Love* is the right thing for me to do because watching the movie will be enjoyable. Internalists can argue that when I make this judgment, I must be in a mental state that consists of a desire to see the movie. Since this desire is a disposition to be motivated

in a certain way, I will then acquire the relevant motivation, the motivation to see the movie that I judge I should see.

I do not deny that there are situations in which an agent's desires, i.e., dispositions to have certain motivations may be blocked and thereby unable to produce motivation due to different emotional disturbances such as depression. In fact, these exceptional cases are exactly the same cases as the ones which the externalists always try to present as an objection to traditional unconditional internalism (formulated in terms of motivation). The reason why, for example, the depressed mother grounds an objection to unconditional internalism is that this form of internalism does not leave room for the moral judgments of depressed people and the traditional attempts to explain the phenomenon away have been somewhat artificial (see Section 2.5.2). Yet, the possibility that the depressed mother can still make genuine moral judgments without obtaining any motivation to act accordingly can now be reasonably accounted for within the framework of unconditional internalism by relying on the dispositional desires introduced above. Unconditional internalism with dispositional desires can be argued to be a form of unconditional internalism that is invulnerable to the externalist counterexamples.

Let us discuss the case about the depressed mother again in order to find out how the proposal handles this example. The strategy is fairly simple: I will argue that, assuming the depressed mother has made a genuine moral judgment, she can still have the corresponding dispositional desires to act accordingly even if she has no motivation to do anything. We can endorse Mele's view that the depressed mother still retains her moral judgment that helping her ailing uncle is the right thing to do. With unconditional internalism with dispositional desires, we can argue that the depressed mother reacts to her moral judgment by coming to have the corresponding

desire-like mental states—i.e., dispositions for being motivated. The only effect caused by her depression is that the depressed mother's desire fails to be able to produce motivation to act in accordance with her moral judgment. Thus, on this view, even if the externalists could be right in claiming that the depressed mother has no motivation, the depressed mother could still be argued to be disposed to have the relevant motivation and so still she has a desire to act in accordance with her moral judgment. As a result, we can conclude that there is a form of unconditional internalism that is not defeated by the typical externalist counterexamples.

6.3.3 Responses to an Objection against the Dispositional Desires

The externalists might, of course, object to the previous formulation of unconditional internalism in terms of dispositional desires. Besides suggesting that amoralists lack motivation, the externalists might further claim that the depressed people and especially the alleged amoralists also lack the relevant dispositional desires—they are not even disposed to have the relevant motivation. The critics of the previous proposal could thus argue that, when there is no sign of motivation, the burden is on me to show that the agents in the alleged counterexample cases have at least the relevant desires, i.e. dispositions that link moral judgments to motivation. They then may argue that this cannot be shown in any non-question begging way. In response to this objection, I will outline a thought experiment which could be used as one test for whether the relevant amoralists should be thought to have at least the required dispositions for being motivated.

In order to investigate whether the depressed mother would have the relevant desires, we need to consider situations in which her dispositional desire—which is currently unable to produce motivation in her—could manifest itself in other ways. This is because, for any disposition, it

must be possible to specify at least some manifestation conditions where the disposition is effective. Take a wine glass as an example. We normally assume that a thin wine glass is fragile, which means that it has a disposition to break easily. If we then want to test whether a given wine glass has the disposition to break easily, we need to establish some conditions in which the wine glass break easily. Similarly, if the depressed mother in our case does have a disposition to be motivated, we should be able to specify at least some conditions where that disposition can manifest itself.

To specify the relevant manifestation conditions, let us imagine that a magic button is invented. This magic button is such that, when you press it, it brings about what you think you should do without any effort from you whatsoever. The whole process is made incredibly easy—all you need to do what you judge to be the right thing is to press the touch-sensitive magic button on the panel of the machine and the magic machine will take care of the rest. We can then imagine that the depressed mother is given the magic button. One advantage of considering this scenario is that whether or not the mother will press the button can provide a good indication of whether she really has a dispositional desire to help her uncle. Suppose that the depressed mother knows what pressing the button will lead to. Will she press the touch-sensitive button or not?

One possibility is that the depressed mother will press the magic button. I believe that, if this were the depressed mother's reaction, it would show that she has a dispositional desire to help her uncle corresponding to her moral judgments that this is the right thing to do. That she would press the button if she had one reveals the depressed mother's dispositional desire that remains masked in the ordinary circumstances when she does not have the button.

Yet, it is highly possible that the externalists would answer the previous question in a different way. They would be more likely to claim that, even if the depressed mother could do what she judges to be right by merely pressing the button, she would not do so. On the basis of this, the externalists could then claim that the depressed mother does not need to have even the relevant dispositional desire in this situation and so the unconditional form of internalism outlined above must be false.

However, in this case, the externalist objection to unconditional internalism with dispositional desires is less plausible than the corresponding objection to unconditional internalism with motivation. If the depression eliminates even the mother's dispositional desire to help her uncle so that the mother would not even have sufficient motivation to press the magic button, we would be tempted to say that in this case the depression has also affected her moral judgments. In this situation, many of us would doubt whether the depressed mother really is still making the relevant moral judgment in the first place.

After arguing that even the depressed mother would in all likelihood have the relevant dispositional desire to be motivated, we can consider an even harder case—the amoralist Patrick who was claimed to be making genuine moral judgments even if he has no motivation to act accordingly (Section 2.5.2). Let us imagine that we give Patrick too the magic button that would enable him to do what he judges to be the right thing merely by pressing the button. We can also imagine that Patrick hears about a group of people who live in a faraway country where an evil dictator is threatening to kill them. Fortunately, Patrick could save the innocent people by simply touching the magic button. If he presses the button, the danger will disappear. Now suppose that Virginia then again asks Patrick: do you think it is right to save the innocent group

of people in this case? Patrick again answers ‘yes’ to this question. He explains that he believes that it is morally right to save the lives of other people when you can easily and safely do so. The question then again is: will Patrick press the magic button in order to do what he claims is right?

If Patrick were to press the touch-sensitive magic button, then the innocent residents in the remote island would be saved. In this situation, from Patrick’s behavior, we could naturally deduce that he has at least some dispositional desires to save the innocent people in a way that fits his moral judgment, which in this case could be claimed to be a genuine one. But what if Patrick would not press the magic button even if he were aware of how it works and even if he claimed that he has made a genuine moral judgment that saving the people would be the right thing to do in this case?

My intuition in this case is that, if Patrick does not press the magic button, this is a sufficient reason to think that he has not made a genuine moral judgment. Let me explain what makes me think that this is the case. The previous thought experiment is designed so that, if Patrick had the relevant dispositional desire, it would be able to manifest itself as easily as possible. If Patrick then does not press the button in this situation, this gives us a reason to think that he has no dispositional desire to act in accordance to his moral judgment at all. After all, it seems that there would be no other possible situations where Patrick’s dispositional desire could manifest itself easily. As a consequence, because Patrick arguably does not even have a dispositional

desire to act according to his alleged moral judgment, this makes me think that he has not made a genuine moral judgment in the first place.[45]

Nevertheless, in this argument, I do not want to rely merely on my own intuitions about the previous case. Rather, I want to suggest that there is empirical evidence suggesting that a similar intuition is actually quite widespread. Recently, in a survey aiming to test folk intuitions, a large number of subjects were presented with various scenarios concerning the behavior of psychopaths (Björnsson et al. 2015).[46] The psychopaths in these cases are described as amoralists—the described agents display two distinguishing features. Firstly, the way in which they classify some actions as 'morally right' and others as 'morally wrong' follows exactly the way in which other ordinary people tend to classify the very same actions. Secondly, the amoralists are stipulated not to have any motivation or desire to do what they seem to think is right (or refrain from doing what they think is wrong).

[45] At this point, the externalist position also faces other challenges, e.g., how to distinguish between sincere and insincere moral judgments. Those who defend unconditional internalism with dispositional desires can use the presence of the dispositional desires as a criterion of when an agent has made a genuine moral judgment and likewise for when her moral utterances are sincere. The externalists, however, cannot use the presence of the dispositional desires for these purposes. This means that would need some other ways of distinguishing between sincere and insincere moral utterances. Furthermore, it is not clear what such an alternative criterion could be.

[46] Psychopaths are usually thought of as being unable to distinguish moral standards from conventional ones and they are also assumed to be less concerned about the welfare of others (Blair 1995, 25). Many assume that this is because psychopaths have emotional deficits, and as a consequence, they lack empathy, remorse, and even guilt when acting wrongly (Kennett & Fine 2008, 189; Prinz 2006, 32). There has been a debate about psychopaths between the internalists and the externalists because both sides have drawn very different conclusions from the cases concerning them. On the basis of reviewing empirical experiments and observing patients in medical contexts, the externalists claim that psychopaths do really make moral judgments (Nichols 2002, 2004; Roskies 2003, 2006). In contrast, the internalists have described the same cases in ways that have suggested that psychopaths cannot make genuine moral judgments in the first place (Prinz 2006). For the other internalist responses see also Cholbi (2006a, 2006b, 2011), Kennett & Fine (2008) and Smith (2008). For a general description of the debate on psychopaths between internalists and externalists, please refer to Tiberius (2015, 79-84).

Consider Anna, a recognized psychopath, who has found the mobile phone she intends to buy (Björnsson et al. 2015, 728).[47] The sales assistant introduces two phones of the same model to Anna. Both phones are identical both in how they function and in their retail price. The only difference is that, if Anna buys the phone on the left, 20$ will be donated to starving children in Sudan. Anna takes herself to believe that not choosing this phone would be morally wrong, exactly in the same way as all ordinary people believe as well. Yet, in contrast to the most ordinary people's reactions, Anna does not show even the slightest desire or impulse to buy the phone on the left. She even tells the sales assistant that she does not care at all which phone she ends up buying.

When asked whether Anna believes that it is wrong not to choose the phone on the left, only 36% of the participants in the study attributed such a belief to Anna. This means that the majority of the responders, 64% of the participants, had the intuition that Anna does not hold the belief that failing to choose the phone on the left is morally wrong. This result suggests that ordinary people do not think that a person has made a genuine moral judgment unless she has at least some desire to act accordingly, as shown by the previous kind of cases where the relevant desires can manifest itself incredibly easily.

The previous case of Anna is similar to the case of Patrick because both Anna and Patrick show complete indifference to what they take to be their moral judgments. Let us return to the case

[47] In an experiment testing whether Smith's conditional internalism was true, Nichols (2004) designed a scenario in which a psychopathic criminal killed other people for their money. Compared with Nichols' experiment, the current one is better (Björnsson et al. 2015, 722). The described scenario rules out factors that could override the agent's weak moral motivation. This leaves only one explanation to Anna's lack of the relevant motivation, which is her lack of moral judgment.

of Patrick. Patrick was described as remaining completely unmoved by his thought that not pressing the magic button would be wrong, which, if true, makes him an amoralist. Because he has no motivation to press the button, I suggested that this case gives us a reason to think that he does not have the relevant dispositional desire either, which further suggests that he has not made a genuine moral judgment. Anna's situation is almost exactly the same as the case of Patrick. Anna too remains completely unmoved by the prospect of buying the phone on the left even if she takes herself to judge that this would be the right thing to do. Similarly, Anna does not display any emotional remorse when she says that she does not care about which mobile phone she will buy. This means that neither Anna nor Patrick displays any signs of the relevant dispositional desires (that would create motivation in them at least in the relevant manifestation conditions). Arguably, this consequence makes us hesitate to ascribe genuine moral judgments to them. If the majority of subjects in the study believe that Anna has not made a genuine moral judgment considering her behavior, presumably they would also think that Patrick has not made a genuine moral judgment if he does not press the button in my thought experiment.

The previous similarity between the two cases is the reason why the results from Björnsson et al.'s study nicely support my proposal according to which making a genuine moral judgment requires having a dispositional desire that must be able to manifest itself by producing motivation at least in some cases. The study reveals that ordinary people do not accept that a person holds a genuine moral belief unless that belief manifests itself via a dispositional desire that can produce some motivation to act accordingly at least in some cases where this can happen very easily. If this is the case, then we should accept that it is necessary that, when an agent has made a moral judgment, she must have a corresponding dispositional desire to act

accordingly. Thus, the empirical study supports the unconditional internalism that I introduced and defended in Section 6.3.2.

6.3.4 Responses to Strandberg's Objection

I will further defend my view by considering another recent objection that targets the kind of unconditional internalism introduced above. This objection comes from Caj Strandberg (2012). Strandberg's argument consists of two steps. First, Strandberg draws a distinction between two kinds of mental states which he calls 'dispositional' mental states and 'occurrent' mental states. The distinction between two general kinds of mental states means that there are also both occurrent and dispositional moral judgments and motivational states. Strandberg then claims that even the expressivists have to grant that when agents make occurrent moral judgments they do not always also have occurrent motivation to act accordingly. In the depressed mother's case, the depressed mother obviously makes an occurrent moral judgment whereas she lacks a corresponding occurrent desire. Yet, at this point, it becomes mysterious how we should understand her occurrent moral judgment—it cannot be an occurrent moral judgment because it itself is either an occurrent desire-like state or it has a power to produce such states. Furthermore, this also means that we need to be able to explain what the difference between occurrent and dispositional moral judgments is given that the difference cannot be the kind of motivation these states produce as both at best can only entail dispositional desires.

In response to this challenge, I will provide two methods with which we can distinguish occurrent moral judgments from dispositional ones. The first will be that an occurrent moral judgment is more transparent to the agent who makes it—when an agent makes an occurrent moral judgments, she is thereby aware of that judgment in a way that is not the case when it

comes to dispositional moral judgments. I will then adopt Toppinen's (2015) view that occurrent moral judgments can also manifest themselves also in many other ways than merely by producing motivation.

Let us first see how Strandberg understands dispositional mental states. Take dispositional desires as an example. Very roughly, a dispositional desire consists of an agent's tendency to do certain kinds of actions. An agent can hold a dispositional desire over a long period of time without ever being motivated to perform the desired actions because she is never in the right kind of circumstances to do those actions and so the dispositional desire is never activated. When the circumstances are right however, the dispositional desire will be activated, and it will manifest itself. At this point, the desire acquires a new, different vivid status in which it can exert influence on the agent's behavior, and eventually issue actions (Goldman 1970, 86-88; Mele 2003, 31, Strandberg's 2012, 83). Consider a very simple case. I may have a standing background dispositional desire to eat fried chicken. Yet, this desire only provides pressure on what I do when I see a fast food store where fried chicken is available.

In contrast, an 'occurrent' desire often is what derives from a 'dispositional' desire. According to Strandberg, an occurrent desire 'takes the form of an episodic mental event … at a particular moment' (Strandberg 2012, 83). As I just mentioned, when the conditions are right, a dispositional desire can be activated and so it turns into an occurrent desire with the very same content. An occurrent desire is the desire that has been aroused from its dormant status, and thus it can exert influence on the agent's behavior directly (Goldman 1970, 86-88; Mele 2003, 31; Toppinen 2015, 155). Because an occurrent desire can exercise influence on an agent's

action, Strandberg (2012, 83) suggests that an occurrent desire can in part explain an agent's behavior because such a desire exercises influence on her actions.

An example might help us to understand the relation between dispositional and occurrent states. As a movie-lover, I always want to watch Mission Impossible films even if I have watched all of them several times. Every time when I watch one of these films again, I am still excited about the adventures of Tom Cruise and his team. This is why there is a sense in which I already had a desire to watch the *Mission Impossible 6: Fallout* even before it was shot and released. Yet, before its release, my desire to watch this film had to be a dispositional desire. For one, my dispositional desire to watch the film continued to exist for a long time until I finally got a chance to see the film. Furthermore, and more importantly, the dispositional desire to watch the film was capable of being activated even when it had not yet become active so as to exert influence on my actions. However, when I came to know that the film, I desired to watch now runs in the Cineworld, my dispositional desire became active and I acquired an occurrent desire to watch that movie.

There is also a similar distinction between occurrent moral judgments and dispositional ones, which Strandberg uses as foundation for his objection to the kinds of unconditional internalism that have been defended in this chapter. The objection proceeds in three stages.

First, Strandberg invites us to recall what the expressivists claim. The key part of the expressivist views generally is that moral judgments consist of desire-like states. He suggests that this is why the expressivists are committed to unconditional internalism, which claims that whenever an agent makes an occurrent moral judgment, she must also have an occurrent desire

to act accordingly. This is presumably because the occurrent moral judgment is an occurrent desire (in the same way as a dispositional moral judgment would consist of a dispositional desire).

Secondly, Strandberg then argues that cases of depressed people are strong counterexamples against the previous form of unconditional internalism. Consider a depressed mother who utters that she really should help her son to stop using drugs. Normally when someone makes an utterance of this kind, we are inclined to ascribe an occurrent moral judgment to that person. Yet, if we are also aware that the mother is deeply depressed due to her son's serious drug problem, we accept that she might not have an occurrent desire to act according to her moral judgment. A case like this therefore means that unconditional internalism relying merely on occurrent mental states is false as occurrent moral judgments do not always entail occurrent desires.

At this point it is easy for the internalists to point out that the previous claim does not jeopardize unconditional internalism with dispositional desires, the view that I have outlined and defended so far in this chapter. This is because that view only requires that an agent who has made an occurrent moral judgment has a dispositional desire to act accordingly. This is finally where we get to the crux of Strandberg's objection. The problem is that at this point we cannot draw the distinction between occurrent and dispositional moral judgments in terms of what kind of motivation they entail. After all, we now think that both occurrent and dispositional moral judgments require that an agent who makes these judgments has at least some dispositional desires to act accordingly. The challenge which Strandberg presents for the internalists is then the demand that they should be able to differentiate between occurrent moral judgments and

dispositional moral judgments in some other way and at least initially it is not clear how that could be done.

One way to respond to this objection is to claim that occurrent moral judgments are transparent to agents, whereas dispositional moral judgments are less transparent. Yet, we should notice that this explanation is not available for the expressivists. Let me unpack this point a little. The expressivists cannot think in this situation that occurrent moral judgments are occurrent desires as they have now granted that agents who make occurrent moral judgments do not always have corresponding occurrent desires. Instead, they must now think that a given occurrent moral judgment just is identical with some dispositional desires. In this case, the expressivists cannot claim that we are easily aware of our occurrent moral judgments but not necessarily as easily aware of our dispositional desires. The expressivists cannot make this claim because they are now thinking that the two are one and the same mental state.

However, the response that occurrent moral judgments are more transparent is available for the internalist cognitivists, who are not committed to expressivism. As always, internalist cognitivists suggest that (occurrent/dispositional) moral judgments and (dispositional) desires are two different mental states and it is the case merely that the former tend to produce the latter. Even if occurrent and dispositional moral judgments both produce dispositional desires in this situation, internalist cognitivists can still argue that they are different kinds of mental states. They can, for example, claim that one difference between the two kinds of moral judgments that produce the same kind of desire is that the occurrent moral judgments are more transparent to us—we are more easily aware of them than of our dispositional moral judgments. The fact that both kinds of moral judgments entail the very same kind of desires in this situation is not

reason to think that there cannot be differences like this between occurrent and dispositional moral judgments given that neither kind of judgments are identical with the dispositional desires they are able to produce.

We can see this if we return to the depressed mother's case. When the depressed mother claims that it is right for her to help her son, she is clearly aware of the occurrent moral judgment, which her utterance expresses. Yet, if the depressed mother were only in a state of a dispositional moral judgment, she probably would not realize that she was in such a state even if that judgment already existed. It thus appears plausible that occurrent moral judgments are transparent to the agent because of her awareness of those judgments. In contrast, dispositional moral judgments lack the same kind of transparency to the agents whose judgments they are. In this way, occurrent moral judgments can be distinguished from dispositional ones by what kind of awareness we have of them.

There is also another way in which we can distinguish occurrent moral judgments from the dispositional ones—through various desire-like mental states that are connected to those two types of judgments.[48] For example, Toppinen (2015, 156) develops a response to Strandberg based on the idea of multi-track dispositional desires. According to him, an occurrent desire can manifest itself in many ways other than by merely motivating an agent, such as by causing feelings of guilt in the agent. Consider the depressed mother, for example. It could be true, as Strandberg suggests, that the depressed mother cannot form an occurrent desire to help her son with the drug issue because she is depressed. But it is at least equally plausible to believe that

[48] It is important to note that I merely describe how dispositional desires differ from occurrent ones here according to Strandberg's view. In Section 7.3, I will argue that dispositional desires can also manifest themselves in various ways, such as by producing emotions (or so-called reactive attitudes).

the depressed mother would feel guilty exactly because she lacks motivation to help her son even though she believes that it is the right thing for her to do. In this way, the depressed mother's occurrent desire manifests itself in a different way, not by producing motivation but rather by producing the emotion of guilt in her. This explanation, however, does not apply to dispositional moral judgments. As I have explained, if an agent fails to be moved by her dispositional moral judgments, she need not feel guilty as a result as she may remain unaware of failing to act in accordance to her dispositional judgment.

Generally, in Section 6.3, I have thus continued to investigate what the most plausible form of internalism is and more specifically whether we should be conditional or unconditional internalists. In Section 6.3.1, I first reintroduced unconditional internalism which was already first introduced in Section 2.5.1. I then went through some of the strategies adopted by the internalists in response to the externalist counterexamples such as the depressed people. I explained how many internalists have defended different forms of conditional internalism as a response to those objections. Instead of following this strategy, in Section 6.3.2, I tried to deal with the externalist counterexamples within the framework of unconditional internalism. I argued that there are good reasons to believe that an agent's moral judgments always entail that she also has corresponding dispositional desires to be motivated and so we should accept a form of unconditional internalism formulated in terms of dispositional desires.

In Section 6.3.3, I tried to respond to an objection to the previous kind of unconditional internalism with dispositional desires. I suggested that even depressed people will have dispositional desires to be motivated that correspond to their moral judgments, whereas it is more plausible to think that agents who lack such dispositional desires have not even made the

relevant moral judgments in the first place. In section 6.3.4, I then considered Strandberg's objection which requires us to provide a further clarification of what the difference between dispositional moral judgments and occurrent moral judgments is given that both kinds of moral judgments entail the very same kind of motivation. In response to this objection, I argued that either we can understand that differences between the two kinds of moral judgments either in terms of how transparent they are or we can think that occurrent moral judgments can also manifest themselves by producing various emotions, such as guilt, whereas merely dispositional moral judgments cannot produce such emotions.

6.4 Conclusion

To sum up, this whole chapter—Chapter 6—partially achieves my goal of discovering the most plausible form of internalism. My investigation here started by evaluating non-constitutional internalism in section 6.2. I suggested that *de dicto* and *de re* views are about different subject-matters and because of this I mainly focused on the implications of the relevant *de dicto* and *de re* views. As a consequence, I then discussed the four views that can be constructed on the basis of combining *de re* internalism and externalism and *de dicto* internalism and externalism in different ways. It turned out that any combination that includes *de re* externalism is implausible given the arguments already discussed in Chapters 3-5. Yet, I also explained why I can remain neutral between the other two possible combinations of (i) *de dicto* internalism and *de re* internalism and (ii) *de dicto* externalism and *de re* internalism, given that my main interest in this thesis is to defend a form of *de re* internalism. If that view is compatible with both *de dicto* internalism and *de dicto* externalism, then both of those views are acceptable to me as long as they are compatible with *de re* internalism in the way that I suggested.

I then continued to investigate another form of internalism, unconditional internalism, in Section 6.3. I first considered the traditional forms of unconditional internalism and the strategies which the defenders of such views have for tackling the externalist objections to internalism. I then argued that we could defend unconditional internalism by formulating the view in a slightly different way. As a consequence, I outlined a form of unconditional internalism in terms of dispositional desires as a version of unconditional internalism that can avoid the externalist objections. In support of the new form of unconditional internalism, I provided an argument suggesting that even depressed people and at least some amoralists have dispositional desires for acting accordingly when they have made moral judgments.

Furthermore, I also replied to a challenge concerning how to distinguish occurrent moral judgments from dispositional moral judgments, which in this situation cannot be done by relying on what kind of motivation those moral judgments are connected to given that at best both kinds of moral judgments only entail dispositional desires. In response, I suggested that we can draw the distinction in two ways. Firstly, it can be argued that occurrent moral judgments are more transparent than dispositional moral judgments. And, secondly, occurrent moral judgments can also manifest themselves by producing, not only motivation, but also other moral emotions such as guilt.

Chapter 7: Strong vs. Weak Internalism and Direct vs. Deferred Internalism

7.1 Introduction

In this chapter, I will continue the exploration of different forms of internalism, which began in the previous chapter. The first half of this chapter thus investigates whether we should accept a strong or a weak form of internalism. In section 7.2, I will briefly remind the readers of the difference between strong and weak internalism (see Section 2.4). Following this, in section 7.3, I will try to describe three key differences between desires and motivation. Then in Section 7.4, I will use these resources to formulate a spectrum on which the strength of desires can be evaluated. Finally, in section 7.5, I will be able to investigate how strong dispositional desires you are required to have in order to count as someone who has made a genuine moral judgment. My conclusion will be that, for a moral judgment to count as genuine, the corresponding dispositional desire should be able to manifest itself in ways other than by merely producing motivation, for example, by producing the so-called reactive attitudes.

In the rest of this chapter, I will explore which one of the remaining, contrasting forms of internalism—direct internalism or deferred internalism—is more plausible. In section 7.6, I will review the main motivations which philosophers have had for arguing for deferred internalism and the crucial argument supporting that view. In section 7.7, I will argue against both individualist deferred internalism and communal deferred forms of internalism. The discussion will enable me to conclude that direct internalism with dispositional desires is a more plausible view than any form of deferred internalism. This conclusion will then mark the end of my exploration of which version of internalism is the most plausible view to accept. I will finally draw all the relevant conclusions in Chapters 8.

7.2 Revisiting Strong and Weak Internalism

As explained in Sections 2.4.1 and 2.4.2, traditionally, strong and weak forms of internalism have been formulated in the following way:

> *Strong internalism with motivation*: Necessarily, if an agent judges that it is right to φ in circumstances C, then she has overriding motivation to φ in C.

> *Weak internalism with motivation*: Necessarily, if an agent judges that it is right to φ in circumstances C, she has at least some motivation to φ in circumstances C.

There is a simple reason for why the previous strong form of internalism has always been unpopular (see Section 2.4.2), which can also be elucidated with an example. Imagine an official who oversees which building project will get supported by public funding. He initially intended to review all applications following a procedure described in a set of rules, which he believes is the right thing to do. However, there are lobbyists hired by the relevant stakeholders and they are willing to offer huge financial rewards for the official if only he will give their applications a priority, which would mean violating the previous rules. Such offers are very tempting and, as a result, it often happens that the official ends up accepting the bribes even if he thinks that such behavior is obviously wrong.

In this case, the corrupt official's motivation to act according to his judgment about what is right is intuitively overridden by his desire for money and wealth and all the nice things he can get with them. It would not be hard to find similar cases where ordinary people's motivations

are intuitively overridden. However, strong internalism with motivation would rule out this possibility. It entails that you have not made a genuine moral judgment unless you have overriding motivation to act accordingly. Given how unintuitive the consequences of this view are in the previous kind of cases, not many philosophers have ever endorsed the view.[49]

Given that strong internalism with motivation thus turns out to be highly implausible, we might then wonder if, as internalists, we should therefore accept weak internalism with motivation instead. However, even the weaker forms of (unconditional) internalism with motivation are problematic because of the counterexamples such as the depressed mother which have been discussed in the previous chapters (Section e.g. 2.5.2). In that case, we think that the mother's moral judgments have not changed, even if she has no motivation whatsoever to act accordingly. This is the case because, as I suggested in Section 6.3.3, her dispositional desire remains intact even if it is no longer producing motivation. So, neither version of internalism formulated above seems very plausible to me.

Given the forms of internalism with dispositional desires that I discussed in Chapter 6, we can, however, also draw a similar distinction between strong and weak forms of (unconditional) internalism formulated in terms of dispositional desires. These two alternatives can now be formulated in the following way:

[49] The previous form of strong internalism thus makes weakness of will impossible. It suggests that no agent could ever act against her best judgment. This kind of a view dates back to at least Plato. In *Protagoras,* Plato's character Socrates suggests that weakness of will is impossible. Given an agent has enough knowledge to make the best judgment, she will not be willing to act contrary to the judgment, or so Plato argued (358d). This view was then criticized by Aristotle in *Nicomachean Ethics* (1146b5-1147b20). Even if there has been debate about whether we can be weak-willed, generally speaking, it is accepted that we sometimes can give into temptation and so weakness of will must be acceptable.

Strong internalism with dispositional desires: Necessarily, if an agent judges that it is right to φ in circumstances C, she has a strong dispositional desire to φ in circumstances C.

Weak internalism with dispositional desires: Necessarily, if an agent judges that it is right to φ in circumstances C, she has a weak dispositional desire to φ in circumstances C.

In the rest of this first part of the current chapter, I will investigate which one of these two views is more plausible. In section 7.3, I will first clarify further what dispositional desires are and, more importantly, how they differ from motivation. Based on results of that discussions, then, in section 7.4, I will consider how the strength of dispositional desires should be understood, and more specifically, what the difference between strong and weak dispositional desires is. Finally, in Section 7.5, I will compare strong and weak form of internalism with dispositional desires and I will also consider which one of the two views is more plausible.

7.3 The Difference between Desires and Motivation

In Section 6.3.2, I already drew a distinction between dispositional desires and motivation. Although both of these mental states have the world-to-mind direction of fit, the latter can be roughly understood as the typical consequences of the former. For example, if Mary has a dispositional desire to drink her favorite latte, this mental state will not be able to prompt Mary to drink the coffee by itself. Mary's dispositional desire, however, will be able to produce a motivational state in her and that state of motivation can then prompt her to drink the coffee.

In addition to the previous sketch of the rough difference between dispositional desires and motivation, it is possible to clarify the difference between the two states further also in other ways. Firstly, dispositional desires and motivation are usually thought to have different kind of content. Secondly, unlike motivation, desires for different outcomes are usually not thought to necessarily involve or include beliefs, especially not means-end beliefs. Thirdly, dispositional desires are sometimes thought to be multi-track dispositions that can manifest themselves in a number of different ways, whereas motivation cannot do so. Let me then explain these three differences in turn, starting from the first one.

Let me slightly revise the previous example to illustrate the first difference between dispositional desires and motivation. Let us imagine that Mary loves coffee, and after having a worked hard all morning, she really wants to have some on a break. There are then at least two different things we might say about Mary's mental states:

Statement 1: Mary desires that she has a cup of coffee.

Statement 2: Mary is motivated to buy a cup of coffee.[50]

These two statements nicely exemplify how dispositional desires and motivation have different kind of contents. This difference is reflected by the different grammatical form of the previous two statements. In Statement 1, the verb 'desire' is followed by a that-clause—'that Mary desires to have a cup of coffee'. This that-clause picks out a proposition that characterizes an outcome which Mary desires to obtain. In contrast, in the Statement 2, the predicate '…is motivated to…' is followed by a description of an action, the action of buying a cup of coffee.

[50] This basic idea of how desires differ from motivation was also discussed by Marks (1986, 140).

It is notable that we could not replace the action in Statement 2 with the outcome in Statement 1. It would not make sense to say that Mary is motivated *that she has a cup of coffee* as no matter how much Mary is motivated, she cannot get the coffee thereby directly but rather she must do some actions in order to obtain it. This comparison indicates that the content of a desire is a proposition representing a certain outcome whereas the content of motivation is an action.[51]

The second difference between dispositional desires and motivation is that any motivation to do an action arguably already includes a means-end belief, whereas a desire for an outcome does not and need not include such a belief necessarily.[52] To clarify this difference, it is helpful to consider Mary's case again. It seems that Mary's desire alone will not be able to lead to the consequence that she has a cup of coffee, unless some additional means-end belief, for example, that she can buy a cup of coffee from a nearby coffee shop combines with her desires to produce action. In contrast, suppose that Mary has some motivation to buy a cup of coffee. Then, if her motivation to do this is not blocked by external forces or overridden by stronger motivation, her motivation will lead to the act of buying a cup of coffee. Mary's motivation does not seem to require an additional means-end belief because such means-end beliefs have been included in her motivation itself already. However, if a means-end belief, for example, the belief that Mary can buy her favorite latte from a coffee shop is absent, then she could not be claimed to have the motivation to buy a cup of coffee in the first place. This is why it is generally thought that motivation involves beliefs, whereas desires do not.

[51] It is true that we can say that Mary desires to buy a cup of coffee. You might think that this means that desires too can have actions as their contents. However, it has been suggested that this claim should be better understood as referring to Mary's motivation, and that is an elliptical way of saying that Mary desires a certain outcome in which she buys a coffee.

[52] In his response to an externalist objection, Smith (1996b) also holds a similar view when he endorses the idea that the combination of desires and means-end beliefs will constitute motivation.

The third difference between desires and motivation was already mentioned in the previous chapter, but it still needs some supplementary elaboration. In Section 6.3.4, I explained how desires can be thought of as multi-track dispositions, which means that desires can manifest themselves in various ways, for example, by causing different emotions in an agent. For instance, in Mary's case, when Mary talks to her colleagues during a break, she might complain that she should have brought a cup of coffee with her this morning, as she eagerly desires to have one. Likewise, because of her present desire, Mary might regret not having a cup full of coffee with her. By comparison, it does not seem to be very plausible to think that motivation to do an act would have similar multi-track features. The motivation she has for buying a cup of coffee itself does not make her regret that she does not have a cup of coffee already. Thus, compared with how desires can manifest themselves in many different ways, motivation to do a certain action can manifest itself only through the relevant action the agent is motivated to do.

7.4 The Strength of Dispositional Desires

The last section explored some of the distinctive features of the dispositional desires. There is, however, also another issue that needs to be discussed before we can answer whether strong or weak unconditional internalism with dispositional desires is more plausible. In order to be able to answer that question, we would have to know how strong and weak dispositional desires should be understood in the present context. In the case of motivation, the distinction between weak and strong forms of internalism can be drawn easily enough: strong views require that the agent has overriding motivation (that is not overridden by the agent's other motivation), whereas the weak views require that she only has some motivation to act. However, in the case of dispositional desires, it is not even clear what having an overriding dispositional desire would mean. This is why we need a new account of the strength of dispositional desires. The three

differences between desires and motivation discussed in Section 7.3 are about to give us a perspective for considering the ways we can evaluate the strength of dispositional desires. In this section, I will suggest that such an account can be created on the basis of the idea that the strength of dispositional desires can manifest itself in three different ways.

The first dimension of the strength of dispositional desires could be claimed to be how strong motivation is usually produced by the given dispositional desire. To illustrate this idea, let us consider the following case about an agent, call him Tim. Let us assume that Tim has a dispositional desire to become a good person. Tim has acquired this dispositional desire through the way in which his family brought him up and the inspiration provided by many of the good people around him. Being a good person is not, however, easy for Tim because it requires overcoming many temptations. For example, sometimes Tim is very tempted not to buy a train ticket because it does not seem like the ticket will be checked. He thus often feels the temptation not to buy a ticket even if he knows that free-riding is incompatible with the idea of being a good person.

Although in this case there is a conflict between different desires within Tim's set of desires, let us assume that his dispositional desire to be a good person produces strong motivation that overrides the motivations produced by his other dispositional desires. In the previous example for instance, the motivation produced by Tim's desire to be good outweighs the motivation produced by his desire not to buy a ticket. Because of this, it is plausible to say that Tim's dispositional desire is a strong one—it tends to produce powerful motivation that can outweigh other motivations and thus push him to act accordingly.

The second dimension of the strength of a dispositional desire can be understood in terms of in how many different kinds of situations the desire tends to produce motivation. If we compare two dispositional desires that tend to produce equally strong motivations, then, other things being equal, the one that can generate that motivation more often and in more different kinds of situations should be thought to be a stronger dispositional desire than the other one. Take Tim again as an example. Assume that, in addition to his dispositional desire to be a good person, Tim also has many other dispositional desires, many of which conflict with that desire. Perhaps he desires occasionally not only not to buy a train ticket, but also to tell white lies to his friends, to cheat on his wife or to steal office stationery. One possibility is that Tim's dispositional desire to be good produces motivation only when he is tempted not to buy a ticket but not when he is tempted to do things like, for example, telling a white lie. In this case, it is natural to think that Tim's dispositional desire to be good is a weak one for this very reason. In contrast, if Tim's desire to be good produces motivation in all of the previous contexts and many other ones too, we could claim that his dispositional desire is a strong one. In fact, the more there are different kinds of situations where a dispositional desire can produce motivation, the stronger that desire can be claimed to be.

There is also a third dimension on which the strength of dispositional desire varies, namely, in how many different ways a dispositional desire tends to manifest itself. Earlier, I already explained how many dispositional desires are multi-track dispositions that produce not only motivation to act but also emotions such as regret and guilt in the agent. This means that weaker desires tend to manifest themselves only by producing motivation, whereas stronger desires can be thought to be able to manifest themselves also in other ways. Consider the following situation. Let us assume that Tim has always judged that it is wrong to cheat on his wife. Once,

however, Tim does cheat on his wife due to an impulse, but he does not feel happy at all afterwards. Instead, Tim experiences a strong feeling of remorse for what he has done. A feeling such as this could be argued to indicate that Tim has a relatively strong dispositional desire to act in accordance with his moral belief, even if the dispositional desire is too weak to manifest itself by producing always overriding motivation.[53]

I have then introduced three different dimensions on which the strength of dispositional desires can vary. The existence of these three dimensions of strength suggests that the strength of dispositional desires itself is more like a spectrum, which means that there is not a cut-off point that could be used to determine whether a given desire is a strong or a weak one. Rather, the strength of dispositional desires seems to come in degrees so that it is better to just consider how strong or weak a given dispositional desire is or whether the desire is stronger or weaker than some other desire. I will then rely, in the next section, on this enriched understanding of the strength of desires as I explore the strong and weak forms of weak internalism with dispositional desires.

7.5 Strong and Weak Internalism with Dispositional Desires

From the discussion of the last section, we can conclude that the strength of a dispositional desire depends on the following three things: how strong motivation tends to be produced by the desire, in how many situations the desire tends to produce motivation and in how many different other ways the desire manifests itself. The strength of dispositional desires understood in this way enables us to formulate strong and weak versions of internalism with dispositional

[53] Obviously, there is more than one way for dispositional desires to manifest themselves. In additional to entailing emotions, weak dispositional desires can also manifest themselves through prompting an agent to act in the right way again or to remedy the consequence of a wrong behavior.

desires in a new way. It also helps us to consider which one of these two views is more plausible. As the strength of dispositional desires can now be understood as a spectrum, we can start the exploration of weak and strong forms of unconditional internalism with dispositional desires from the weakest possible version of the view.

An extremely weak unconditional internalism with dispositional desires could be stated in the following way:

> *Weak internalism with dispositional desires:* Necessarily, if an agent judges that it is right to φ in circumstances C, then she has a weak dispositional desire to φ in circumstances C.

We can notice two things about this formulation. Firstly, unless an agent has a dispositional desire to act accordingly, she does not count as someone who has made a genuine moral judgment. Secondly and more importantly, the view places no constraints at all on the strength of the relevant dispositional desire. As long as the dispositional desire is capable of producing even very weak motivation in just one situation and nothing else, it is enough to count as a kind of dispositional desire that would be required for making a genuine moral judgment.

However, even if the previous view is extremely weak and undemanding, the externalists would still want to present counterexamples to it—for example, a case of an amoralist like Patrick—to claim that an agent who makes a moral judgment need not have even the weakest possible dispositional desire to act accordingly. As a response to potential objections of this kind, it would be helpful to remind the reader of the view that I already defended in Section 6.3.3. In

that section, I argued that we are inclined to ascribe a moral judgment to an agent only if she has a dispositional desire to act accordingly, a desire that can manifest itself at least in certain circumstances where this can happen very easily. With regard to Patrick's case, even if Patrick's very weak dispositional desire and the motivation it produces are observed only when he has access to the magic button that can be used to save innocent people without any effort at all, the fact that Patrick has at least sufficient motivation to press the button still supports the idea that Patrick has a very weak dispositional desire. And, as I argued above, if he lacked even that dispositional desire, then we would be unwilling to ascribe a genuine moral judgment to him.

Let us then return to the *weak unconditional internalism with dispositional desires* introduced in the previous paragraphs. As already mentioned, it is noticeable that this view is extremely undemanding. This is because even the weakest possible dispositional desire is sufficient on this view—nothing more is required for making a genuine moral judgment. For example, assume that the following is true about the case of Patrick and his magic button:

1) The dispositional desire can only produce very weak motivation so that it would be very easily overridden by other competing motivations.
2) The dispositional desire only produces such motivation in one situation, e.g., when Patrick has the magic button.
3) The dispositional desire does not manifest itself in any other way.

According to the previous view, even if Patrick's relevant dispositional desire to act in accordance with his moral judgment is as weak as 1) to 3), Patrick will still count as someone who has made a genuine moral judgment. However, the question is: assuming that internalism

is true, should we rather accept a stronger form of internalism with dispositional desires? On such views, Patrick in the previous case would not count as someone who has made a moral judgment because these views set further constraints on how strong of a dispositional desire is required for making a genuine moral judgment.

As I indicated in Section 7.2, strong internalism with dispositional desires can be formulated as follows:

> *Strong internalism with dispositional desires*: Necessarily, if an agent judges that it is right to φ in circumstances C, she has a sufficiently strong dispositional desire to φ in circumstances C.

This stronger form of internalism suggests that anyone who has made a genuine moral judgment must also have a sufficiently strong corresponding dispositional desire to act accordingly. Different versions of this view can then set the threshold of how strong dispositional desires count as sufficiently strong in different ways. To put this view in another way, we would not deem a given alleged moral judgments to be a genuine one unless the corresponding dispositional desires would exceed the relevant threshold—i.e., the desire would have to have more strength than Patrick's dispositional desire as it was just described by relying on the three dimensions of strength explained above in Section 7.4.

There are then three ways in which one could argue for a stronger form of unconditional internalism with dispositional desires in the present framework.

1) It could be argued that a genuine moral judgment requires that a dispositional desire is able to produce, at least in one situation, motivation of a certain strength—motivation that can outweigh at least many other weak motivations.
2) One could also argue that a genuine moral judgment requires that the corresponding dispositional desire is able to provide motivation at least in a certain range of cases and not just in one.
3) Finally, it could further be argued that a genuine moral judgment requires that the corresponding dispositional desire is able to manifest itself not only by producing motivation but also in some other ways.

I will examine each of the three ways in order to investigate whether we should accept a stronger form of internalism with dispositional desires.

7.5.1 Moral Dispositional Desires and Strength of Motivation

First, let us see whether, when you make a genuine moral judgment, you are required to have a dispositional desire that can produce motivation of a certain strength. Up to this point, in the described depression and magic button cases (Section 6.3.3). I have assumed that the depressed mother has only weak motivation to help her uncle and Patrick has only weak motivation to save the innocent people. This is because I have assumed that these weak motivations will be enough to motivate the mother and Patrick to press the magic button in the described cases. However, as I only took a single motivation into account in the described cases, it is hard to evaluate how weak or strong the motivation generated by the relevant dispositional desires would be in these cases. Because of this, it is necessary to consider cases in which the depressed

mother and Patrick have also some other competing motivations to do different things in the situations in which they have the access to the magic button.

Let us assume that, for example, the depressed mother has two conflicting motivations when she has the opportunity to press the magic button. In addition to being motivated to help, the depressed mother also has some motivation to get some hazelnut ice-cream, which she has desired to have for a while. The magic button, as it is designed, can help the depressed mother to satisfy either one of her motivations, but it cannot bring it about that both are satisfied at once. This means that the depressed mother has to make a choice between helping her uncle and getting the ice-cream.

Although both of the depressed mother's motivations could be very weak, one of them will still presumably outweigh the other. Let us further assume that the depressed mother's motivation to help her uncle is even weaker than her motivation to get the ice-cream and so she ends up pressing the button in order to get ice-cream. Strong and weak forms of internalism with dispositional desires will then draw different conclusions from this case.

Many stronger forms of internalism with dispositional desires will suggest that the depressed mother has not made a genuine moral judgment in the previous case, because her corresponding dispositional desire is so weak that it is not able to produce sufficiently strong motivation to outweigh even her most trivial competing motivations. In contrast, the weaker forms of internalism with dispositional desires will accept that the depressed mother still counts as someone who has made a genuine moral judgment even if her dispositional desire can only produce very weak motivation that is outweighed by her other weak motivations.

I am inclined to believe that weak internalism with dispositional desires is the more plausible view in this situation. Consider the depressed mother's situation more generally. In her case, it is true that if she did not suffer from depression, she would presumably be pretty strongly motivated to help her uncle. However, when we take into account that the depressed mother suffers from depression, it is less convincing to argue that her dispositional desire should be able to produce strong motivation also in these circumstances. In this situation, we would think that the depressed mother's dispositional desire's ability to produce anything more than merely weak motivation to help her uncle is easily affected by her psychological condition and so the motivation produced by that desire can be outweighed by other motivations, for example, by her motivation to get some ice-cream. Even if the depressed mother's motivation to get ice-cream outweighs in this case her motivation to help her uncle, this consequence is still compatible with other cases where the mother does not suffer from any mental disturbances so that her dispositional desires will be able to produce strong motivations that outweigh the others. Due to this reason, we should not be skeptical about her commitment to morality because her dispositional moral desires can only produce very weak motivation in her depressed state.

There might still be another concern that weak internalism with dispositional desires will be less plausible in Patrick's case. We can suppose that Patrick is in the same situation as the depressed mother was just above. We can assume that Patrick has weak dispositional desire (that is still capable of producing weak motivation) to help his sick uncle, whereas at the same time he has also a weak dispositional desire to get some ice-cream (which is also producing weak motivation to do so). If Patrick eventually is more motivated to get some ice-cream and so his motivation to help his uncle turns out to be even weaker than his weak motivation to get

ice-cream, the defenders of the stronger forms of internalism with dispositional desires would claim that he has not made a genuine moral judgment in this situation.

They would further argue that there are further important differences between the cases. In the depressed mother's case, if the mother were not depressed, the dispositional desire corresponding to her moral judgment would produce strong motivation. In contrast, Patrick does not suffer from any such disturbances. So, in his case the dispositional desires corresponding to his alleged moral judgment can only ever produce very weak motivation.

To see how serious this concern is, we can return to the situation in which Patrick is more motivated to get ice-cream than he is to help his uncle. In this case, we are aware that Patrick is and has always been almost an amoralist. He is someone who is far less concerned about moral issues than ordinary people even if he cares a little bit given that he does have a weak dispositional desire to do what he thinks is right. This fact should be taken into consideration before we conclude that Patrick does not make genuine moral judgments in the first place. Instead, it could be argued that, because Patrick has weak dispositional desires that both 1) correspond to his moral judgments and 2) can produce only very weak motivation, he is making genuine moral judgments even if the problem is that, despite those judgments, he is not sufficiently sensitive to the demands of morality of which he is aware through his judgments. If Patrick cared more about what is right and wrong according to his judgments, he would probably be more motivated to help his sick uncle. This view that Patrick can make genuine moral judgments even if he is not sufficiently moved by them would explain, for example, why we intuitively would hold Patrick responsible for his choice of getting ice-cream and why we also blame him for doing so. By contrast, if we thought that he was not even capable of making

genuine moral judgments, it would be more difficult to think that he deserves blame for his actions.[54]

7.5.2 Moral Dispositional Desires and the Range of Cases

The second way to defend a stronger form of internalism with dispositional desires is to argue that, when you make a genuine moral judgment, you need to have a dispositional desire that can produce motivation in you in a wide range of different kinds of cases. Up to this point, I have only focused on the magic button cases in my defense of the weaker forms of internalism with dispositional desires. However, it should be noted that all the magic button cases discussed above share two essential features. As we have just seen, firstly, in these cases we tend to assume that the agents involved have no competing motivations that could compete with their motivations to act in accordance to their moral judgments. Secondly, it is also stipulated in these cases that the agents can do what they take to be right almost effortlessly. The weakest form of internalism with dispositional desires would then require that, when you make a genuine moral judgment, you need to have a dispositional desire that can produce motivation at least always when the previous two conditions are satisfied. However, the relevant dispositional desires, would, on this view, not need to be able to produce motivation in any other cases.

At this point, some defenders of the stronger forms of internalism who are inspired by the conditional internalists (Section 2.5.3) would probably suggest that the relevant dispositional desires required by our moral judgments must be able to produce motivation, not only when the previous two conditions are satisfied, but also more generally in all situations in which agents

[54] Susan Wolf (1993, 121) discusses a similar case that pertains to my point here. She admits that, if we can imagine an agent who is incapacitated, the 'mental deficiency' readily exempts her from being morally responsible.

who make those judgments function well psychologically and are practically rational. Many of the arguments and intuitions supporting the traditional form of internalism could also be thought to support this addition. In the rest of this sub-section, I will, however, show that the argument presented in Section 2.5.3 would not be able to support the stronger form of internalism according to which the dispositional desires required by moral judgments must be able to produce motivation in all situations in which the agent is practically rational and psychologically normal.

Let me first describe how, for example, Smith's argument from Section 2.5.3 for internalists could be employed here. You might recall that Smith argued that our moral judgments are actually about what we have reasons to do. He then claimed that judgments about reasons are beliefs about what we take our fully rational versions agents to desire us to do. In this situation, Smith also claimed that, in so far as you are rational, you need to have motivation to act accordingly to your moral judgments. If you lacked such motivation, this would not cohere with what you think your own better version wants you to do. This same argument could also be used to argue that dispositional desires required by genuine moral judgments would have to be able to produce motivation always when an agent is practically rational.

Let us then consider a case. Imagine that Sam makes the following judgment: I should be kind to other people. Note that Sam's judgment is perfectly general, he takes the requirement to be kind to apply to everyone and not just towards some smaller group of people. We can further assume that Sam also has the corresponding dispositional desire to be kind to others. This is indicated by the fact that he is highly motivated to be kind to other people in a vast number of different kinds of cases (and not only to his family or friends but also to strangers). However,

there is one context in which he has no motivation whatsoever to be kind to others, namely, at work where he is responsible for the human resources at a large IT company. I am assuming that in this context, too, Sam has the same dispositional desire, but it is just that this desire is unable to produce motivations in this specific context.

Now, the view under consideration claims that the dispositional desire required by moral judgments must produce the corresponding motivation always when the agent is practically rational. According to this view, we would have to claim that Sam is irrational if he lacks motivation to act to be a kind person at work. But why should we accept this view? To answer this question, let us consider Smith's account of practical rationality.

I have already explained in Section 2.5.3 what full rationality consists of on Smith's view. According to him, in order for an agent to count as fully rational, she has to satisfy four requirements: she should have no false beliefs, she should have all the relevant true beliefs, she should have a systematically justifiable set of desires, and she should not suffer from any physical or psychological disturbances (Smith 1994, 156-161; 1996a, 160). It could be argued that Sam above fails to satisfy the requirement of 'having a systematically justifiable set of desires'. This is because Sam is at least sometimes motivated to be cold and ruthless to his colleagues. The crucial point then is that Sam's dispositional desire to be cold and ruthless towards his colleagues clearly conflicts with dispositional desire to be kind to others that corresponds to his moral judgment. This incoherence can then be used to argue that even Sam must be irrational in the previous case.

Yet, against the previous conclusion, it can be argued that, overall for Sam not having the motivation to be kind to his colleagues can in this case be more coherent and hence more rational than having the motivation that would match his moral judgment that it is right to be kind to other people.[55] For example, let us assume that in the previous case, Sam has a strong desire to be very successful in the company. And, he also believes that the best way to achieve this goal is to follow the rules of the company by setting aside any personal feelings that would normally be considered to be constitutive of kindness. It is true that Sam's lack of motivation to be kind to others at work conflicts with his dispositional desire to be kind to everyone and his corresponding moral judgment. In this sense, Sam could be argued to be at least locally incoherent and irrational. However, when we consider his psychological make-up more generally, it becomes clear that Sam is more coherent if he lacks the motivation to be kind to others at his work, given how badly such motivation would fit his other central cares and concerns. For this reason, we should think that it can be globally rational for an agent to have a dispositional desire corresponding to her moral judgment even if that desire fails to produce motivation in the agent. As a result, I believe that we should not accept the stronger form of internalism with dispositional desires according to which the relevant dispositional desires must always be able to produce motivation in a range of cases in agents when she is rational.

7.5.3 Moral Dispositional Desires and Reactive Attitudes

In the last two sub-sections, I have examined two ways in which someone could try to argue for stronger forms of internalism with dispositional desires, both of which have proved to be implausible. To complete our investigation, let us finally consider the last way in which

[55] For a similar argument, see Nomy Arpaly (2004, 61).

someone could try to argue for stronger forms of internalism with dispositional desires. This exploration aims to find out whether, when you make a genuine moral judgment, you are required to have a corresponding dispositional desire that can manifest itself not just by producing motivation but also in some other ways. In order to answer that question, I will start from considering what distinguishes moral normative judgments from non-moral ones, given that both kinds of judgments are expressed with the same words and they are also both connected to motivation. I will then suggest that one good way to draw the distinction is to take seriously the idea that moral judgments are inherently related to various reactive attitudes, whereas non-moral normative judgments are not related to those attitudes in the same way. This means that it is much more plausible to think that genuine moral judgments must correspond to dispositional desires that can manifest themselves also by producing reactive attitudes. As a result, I will argue that this way of defending a stronger form of internalism with dispositional desires is actually much more plausible than the previous two ways.

The argument I will present begins from the similarities between moral judgments and other kinds of normative judgments. The first similarity is that both kinds of judgments are expressed with the same words. Consider the vocabulary we frequently use, for instance, 'good', 'right', 'ought to', 'should' etc. We can use these words to express moral judgments, for example, the thought that it is morally good to be polite to strangers; the thought that it is morally right to save an innocent person's life; or the thought that we should keep the promises we have made. Sometimes we also use these same words to express, for example, prudential judgments which too are normative judgments even if they are not moral judgments. So, for example, we might say that the (prudentially) right thing to do is to put extra layers of clothing given that the temperature will drop significantly today or that I should use an umbrella as it is raining heavily

outside. Considering that the previous words can be used to express both moral and non-moral normative judgments, we cannot distinguish these judgments from each other merely on the basis of vocabulary used.

Furthermore, prudential non-moral normative judgments also resemble moral judgments in that they too seem to be reliably connected to motivation. We would normally expect that my judgment that I should use an umbrella when it is raining will lead me to have corresponding motivation to use one. This is in the very same way as, we would expect that if an agent judges that it is morally right to save another person's life, she will have some motivation to act accordingly. In both cases, genuine judgments seem to require that the agents who make the judgments have corresponding dispositional desires that will produce motivation in them. Thus, it also seems difficult to see how we could distinguish moral judgments from prudential non-moral normative judgments merely on the basis of whether they produce motivation.

A question then naturally arises: how could we tell the difference between moral judgments and non-moral judgments if this cannot be done by relying on the idea that moral judgments are related to motivation in a certain internal way? One plausible response to the previous question suggests that moral judgments have a certain important distinctive feature, which non-moral judgments lack. People who accept this view think that the dispositional desires that are required by genuine moral judgments must be able to manifest themselves also by producing moral emotions, whereas dispositional desires that correspond to the non-moral judgments need not be able to manifest themselves in these ways.[56] These emotions, or the so-called 'reactive

[56] Blackburn (1998, 61-68), Eriksson (2014), Gibbard (2003, 152-158) and Toppinen (2015, 152-156) all argue explicitly for a necessary connection between moral judgments and multi-track dispositional desires that manifest themselves via producing this type of reactive attitudes. Other philosophers who

attitudes', as Strawson puts it are 'essentially reactions to the quality of others' wills towards us, as manifested in their behaviour: to their good or ill will or indifference or lack of concern' (Strawson 2008, 15). According to Strawson (2008, 16), there are three main kinds of reactive attitudes:

(1) Personal reactive attitudes, such as resentment, are based on the consequence that someone fails to treat you as she is legitimately expected from.
(2) Vicarious reactive attitudes, such as approval and disapproval, are based on the consequence that someone failing/succeeding to treat a third-party in a way in which she/he can legitimately expect from that person;
(3) Self-reactive attitudes, such as guilt, are based on the fact that someone failing to behave in a way in which she can legitimately expect from herself.

Let me illustrate each of these three attitudes. First, take resentment, which Strawson used as an example of a typical personal reactive attitude. Imagine that Mary, for instance, needs a bit of help and her close friend Alex could help her in that way very easily. However, instead of actually giving any help or even an explanation for not helping, Alex just coldly and indifferently refuses to help Mary. In this situation, the resentment Mary feels towards Alex could be thought of as her emotional reaction to Alex's indifference. It is an emotion that is based on the fact that Alex fails to treat Mary in a way that she can legitimately expect from him.

---

have connected moral judgments to these emotions include Brandt (1979, 163-176), Copp (2001, 25-26), Gibbard (1990, 47), and Hooker (2000, 72-75).

Secondly, as reactive attitudes are fundamentally reactions to others' wills, an agent can acquire such an attitude even if she is not involved in the moral affair directly. Consider the previous example again. Although Charles does not know either Mary or Alex, when hearing about them he might still well experience 'the vicarious analogue of resentment' (Strawson 2008, 15). Charles can attain this attitude because he can imagine what his emotional reaction would be if he were in the very same kind of a situation in which Mary is. If that situation, Charles himself would also be resentful towards Alex in the same way as Mary is in the actual case. But as Charles is not actually involved in Mary and Alex's case, Charles cannot respond to Alex's indifference and coldness to Mary with resentment. Rather, from a third-person's point of view, Charles can have an attitude of disapproval towards Alex's indifference to Mary, which is an attitude based on the fact that Alex is failing to treat Mary in a way in which Mary can legitimately expect from Alex.

The third kind of reactive attitude Strawson referred to is what we normally call guilt. It is common for us to feel guilty or be remorseful after we realize that we could have done the morally right thing, but we failed even to try. Suppose that, in the previous case, Alex gradually realizes that Mary has every reason to resent him because he has deeply disappointed her. In that situation, it is easy to imagine that Alex would strongly feel that he should have done his best to help Mary and he would also hope that he could change what he did and makes things even better between himself and Mary, even if he is also aware that this will not happen. In this situation, Alex's attitude of feeling guilty reacts to himself, but it is also an associated attitude because it is based on his realization that he has failed to behave in a way in which Mary can legitimately expect from him.

At this point, we can also notice that there are certain important connections between the various reactive attitudes. Simon Blackburn has thus suggested that these attitudes form 'a network of emotions' (Blackburn 1998, 9). The notable thing is that even if there are different kinds of reactive attitudes, those attitudes are not solitary, but instead they rely on each other and they can also arise as a consequence of the other reactive attitudes. As illustrated by the previous examples, Charles's disapproving attitude towards Alex is based on Alex's response to Mary and Mary's resentment of Alex. The attitude of disapproval can be thought to be a consequence of Mary's resentment. Likewise, we can also suppose that Alex's guilt and remorse (if there is any) are attitudes connected to Mary's resentment and Charles's disapproval, the attitudes which make him realize that he should not have treated Mary's request with coldness and indifference.

It could be then suggested that the difference between moral and other prudential non-moral normative judgments, such as prudential ones, lies in what kind of dispositional desires are internally related to them. The claim would be that only moral judgments require dispositional desires that can manifest themselves by producing the previous kind of reactive attitudes, whereas the dispositional desires related to other normative judgments need not be able to manifest themselves in these ways. If this would be the case, then we can tell moral judgments apart from non-moral judgments through identifying whether the dispositional desires corresponding to a given judgment must be able to manifest itself not only by producing motivation but also by producing the previous kinds of reactive attitudes.

Consider an example of a prudential judgment. Imagine that on the basis of her doctor's advice, Mary judges that she ought to go to bed early, at least before 12 pm, for the sake of her health.

In this case, even if Mary fails to go to bed early because she keeps watching television, we would not expect Mary to feel necessarily guilty as a consequence. Furthermore, we also would not expect anyone else to have third-personal reactive attitudes towards Mary in this case. We would not expect anyone to blame her for failing to get to bed early, or to praise her for having done so successfully. This is, obviously, different than what we expect to be the case in the previous case.

At this point, we can also add that it is intuitively implausible that the dispositional desires that are required in the case of moral judgments would not need to be able to manifest themselves by being able to produce the previous types of reactive attitudes. Consider again a slightly different version of Patrick's case (see the original one in Section 6.3.3). In this new version of the case, Patrick again believes that he has made a genuine moral judgment because he thinks that he has applied the relevant moral terms in the same way as others. In this situation, let us assume that Patrick also has some motivation to act in accordance to what he judges to be 'right'. Let us, however, also assume that in this case Patrick never has any of the previous reactive attitudes. He would not blame others for failing to press the magic button in his situation even if he thinks that this would be the right thing to do. Likewise, he would not feel guilty himself for failing to do so himself and he is not pleased when he saves the people by pressing the button.

In this situation, I believe that we would not ascribe a genuine moral judgment to Patrick even if there is a connection between Patrick's moral judgment and his motivations, and this is exactly because he has no reactive attitudes that would be related to his judgments. As a result of this too, it seems like the most plausible forms of strong internalism with dispositional desires

require that, when you have made a genuine moral judgment, you must have a dispositional desire to act accordingly that can manifest itself, not merely through motivation, but also by producing the previous kinds of reactive attitudes.

In this Section 7.5, I began from the weakest forms of internalism and I then explored three different ways in which the stronger forms of internalism with dispositional desires could be argued for. I have argued that the first two ways of strengthening the relevant forms of internalism do not seem to work. The first method—the arguments to the conclusion that a genuine moral judgment requires that the corresponding dispositional desire must be able to produce motivation of a certain strength—proved to be implausible. This is because I have argued that if the dispositional desire corresponding to a moral judgment is able to produce only very weak motivation, this is not so much a reason to doubt whether or not the agent has made a moral judgment but rather a reason to question her sensitivity to moral demands, of which she is aware through her judgments. The attempt to defend the second way—the idea that a genuine moral judgment requires that the corresponding dispositional desire is able to produce motivation at least in a certain range of cases—also turned out to be problematic. This is because as I have argued any attempt to show that the dispositional desire that would be able to produce motivation in a wider range of cases, for example, always when the agent is rational, are bound to fail.

It, however, seems that the third way is more plausible. It can be more plausibly argued that a genuine moral judgment requires that the dispositional desire corresponding to it is also able to manifest itself by producing certain kinds of moral reactive attitudes and not just motivation. There were two main kinds of support for this idea. Firstly, accepting this view helps us to

explain what distinguishes moral judgments from other normative judgments. Secondly, the view also seems to be supported by our intuitions as illustrated by the previous version of Patrick's case.

7.6 Direct Internalism and Deferred Internalism Again

In Section 2.6, I introduced deferred internalism, which has been defended by many internalists as a response to the externalist counterexamples, such as the problems raised by the amoralists and evil agents who fail to be motivated to act in accordance to their judgments.[57] According to all forms of deferred internalism, individual sincere moral judgments do not always need to generate motivation directly. Rather, even if these moral judgments fail to be accompanied by motivation, they can still be regarded as sincere moral judgments as long as they are connected in a certain way to other moral judgments that are connected to motivation in the right way. So even if there are motivationally inert moral judgments, moral judgments generally still must be connected to motivation in an internal way. As a consequence, deferred internalism, when formulated in terms of motivation, can be put in the following way:

> *Deferred internalism with motivation*: Necessarily, if an agent judges that it is right to φ in circumstances C, then she has motivation to
> φ or her moral judgment is connected in a certain way W to some other moral judgments that are accompanied by motivation.

[57] I introduced deferred internalism in Section 2.6 of this thesis. There are also a number of other people who have discussed this topic. See, e.g. Dreier (1990), Blackburn (1998), James Lenman (1999), Jon Tresan (2009b) and Matthew Bedke (2009; 2019).

By contrast, direct internalism is the view according to which every individual moral judgment must be connected to motivation in the right way and so these judgments cannot rely on how some other judgments are connected to motivation in order to count as genuine.[58] As a counterpart to the deferred forms of internalism, we can thus formulate direct internalism with motivation with the following schema:

> *Direct internalism with motivation*: Necessarily, if an agent judges that it is right to φ in circumstances C, then she has corresponding motivation to φ directly.

At this stage, it might also be helpful to briefly recall the two different versions of deferred internalism, which I already discussed in Section 2.6. The first version we can call 'individualist deferred internalism'. This form of internalism states that an agent's genuine moral judgment must either motivate the agent itself or be connected in the right way to her previous moral judgments that motivated her in the relevant way. The second form of deferred internalism can be called 'communal deferred internalism'. According to this view, genuine moral judgments must either motivate the agent themselves or they need to be connected in the right way to other people's moral judgments, where those judgments are accompanied by the required motivation.

Blackburn nicely illustrates the individualist deferred internalism with the example of Satan (Blackburn 1998, 61). We normally think that Satan is an evil figure who intends to behave in the wrong way on purpose, from the bottom of his heart. Even if he makes judgments about what is right, Satan still much rather pursues that which is wrong. Blackburn invites us to

[58] There are, of course, many different versions of direct internalism such as, strong and weak internalism (see Section 2.4), unconditional and conditional internalism (see Section 2.5) or constitutional and non-constitutional internalism (see Section 2.7).

consider what is behind this phenomenon. Satan, as a fallen angel, is supposed to have had the right kind of motivation attached to his moral judgments before he was exiled from the paradise. Furthermore, it is assumed that neither Satan's moral judgments nor his ability to make them changed when that happened. As a figure of darkness, Satan is still aware of what is morally right and wrong. He, however, allegedly now has exactly the opposite motivations than the ones that he had before—he now wants to do what he believes is wrong and he desires not to do what he takes to be right. Blackburn's own attempt to respond to this objection was to argue that, even if Satan's current moral judgments are not connected to motivation directly in the right way, these judgments are still genuine because they are suitably connected to Satan's previous judgments that were motivating in the right way.

The traditional argument for communal form of deferred internalism is a bit more complicated than Blackburn's defence of the individualist deferred internalism. The main argument for the communal form of deferred internalism has always been that amoralists can only exist in a community where moral judgments are generally assumed to motivate. In this way, communal deferred internalism can be immune to counterexamples in which individuals allegedly make moral judgment without having any motivation to act accordingly. This view can be supported by our intuitions of a community called *Amorality* where there are two sets of vocabulary.

The first set of vocabulary employs certain and somewhat strange words to indicate what we consider to be moral behavior. For example, the residents of Amorality call keeping promises and helping strangers as 'gooq' actions and murdering and cheating as 'baq'. However, these terms are not connected in any way whatsoever to what the members of Amorality are motivated to do (nor are they connected to any other attitudinal reactions in them). The residents

of Amorality are never motivated to pursue ‘gooq’ actions or try to avoid ‘baq’ actions. At the same time, those living in the Amorality community also make use of another set of vocabulary, which includes words such as ‘gooqq’ and ‘baqq’ that are often used to evaluate behavior. They use ‘gooqq’ to refer to strange actions such as shaking hands with the left hand, always leaving some food in the bowl even when you do not feel full yet, and so on. They also use ‘baqq’ to refer to behaviors such as telling jokes in front of a group of people or walking a set of stairs one step at a time. The second set of words has an obvious distinctive feature compared to the first set. Whenever someone calls an action ‘gooqq’, she will try to do that action, whereas if a certain behavior is considered as ‘baqq’ most people in the Amorality community will largely avoid that kind of behavior.

Imagine then an ordinary human being who has travelled to visit Amorality. She soon gets familiar with the difference between the two sets of words. She notices that sometimes the locals use ‘gooq’ and ‘baq’ to refer to the same actions as the ones we call ‘good’ and ‘bad’ even if this has no connection to their motivation, whereas other times they use ‘gooqq’ and ‘baqq’ to refer to completely different kinds of actions even if in this case there is a connection to what they are motivated to do. The question then is, which set of words would the visitor translate into our moral language of right and wrong? The first set of language that employs ‘gooq’ and ‘baq’ gets the extension of right and wrong actions right, but the residents of Amorality do not really care about those actions. In contrast, when the locals use the second set of language that employs ‘gooqq’ and ‘baqq’, they seem to care about those actions exactly in the same way as we are concerned about right and wrong actions, even if from our perspective, the residents of Amorality have odd views about which actions they are to do and avoid.

Intuitively, it is more plausible to think that we should translate the second set of vocabulary into our own moral language.[59] Although the first set of vocabulary has the very same extension as our moral terms, the fact that the residents of Amorality do not care about those actions makes it implausible to think that they would use those words to express genuine moral judgments. It could be further suggested that we would not be able to agree or disagree with residents in the Amorality on their views of 'gooq' and 'baq' actions by using our moral concepts. Given that the residents of Amorality do not have any motivational attitude towards either 'gooq' actions or 'baq' actions, it would not make too much sense for us to start arguing with them about which actions are good or bad.

In contrast, the second set of vocabulary does not suffer from the same problems. The residents of Amorality tend to have positive, motivational attitudes towards what they call 'gooqq' behaviors and negative attitudes of avoidance towards 'baqq' behaviors. Furthermore, exactly in the same way as we commend and pursue good behaviors and condemn and avoid bad behaviors, they do so too when it comes to 'gooqq' and 'baqq' behaviors. This suggests that these terms play the same practical role to express moral judgments as our moral terms do in our community. We could thereby have moral disagreements with the residents of Amorality by using the second set of vocabulary. This is why many have found it plausible to think that it would be correct to translate the second set of vocabulary into our own moral language.

This suggests that the residents of a community can make sincere moral judgments and express those judgments by a given set of terminology only when those judgments and terms play the

[59] For the relevant discussions of Amorality, see Bedke (2009, 194-195; 2019, 9-14), Dreier (1990, 13-14), Lenman (1999, 445-446) and Tresan (2009, 185-186).

same practical role as the corresponding judgments and terms play in our own community. Based on this, many communal internalists have concluded that whether moral judgments are taken to be genuine depends on whether they are related to other judgments in a community that are generally accompanied with motivation to act accordingly.

7.7 In Defence of Direct Internalism with Dispositional Desires

Although individualist deferred internalism and communal deferred internalism both have their attractions as we have just seen, it is still worthwhile to consider whether there are good objections to those views. In this section, I will argue that both forms of deferred internalism should be rejected because there are cases where they have implausible consequences. in order to show that this is the case, in this section, I will explain and develop further Matthew Bedke's recent argument against deferred forms of internalism (Bedke, 2019). Following Bedke, I will argue that these forms of deferred internalism will have to make inconsistent *ad hoc* assumptions about exactly when the genuineness of moral judgments depends on other judgments and when it does not.

Let me begin from the way in which Bedke developed the previous case further to suggest that the communal versions of deferred internalism are not very plausible (Bedke 2019, 12). In the previous case, let us further assume that our visitor to Amorality accidentally meets a local resident called Jane. What is special about Jane is that she first tried to use the second set of vocabulary to describe the same actions that we take to be moral behaviors. That is, she first tried to call, for example, the action of keeping promises 'gooqq'. Furthermore, Jane was very much motivated to act accordingly, in the same way as we care about good actions. Likewise,

she also used to call behaviors such as cheating and hurting other people ‘baqq’ and she used to condemn and avoid these bad behaviours in the same way as we do not want to do them.

However, at this point, all the other residents of Amorality were genuinely puzzled. They could not understand what Jane actually meant when she called those actions ‘gooqq’ and ‘baqq’. This is because, even if she had the required motivations, her attributions of ‘gooqqness’ and ‘baqqness’ were very different from anyone else’s. In order to overcome this problem of communication, Jane decided to invent a third set of vocabulary, ‘gooqqq’ and ‘baqqq’. This set of vocabulary has the same extension as the first set (‘gooq’ and ‘baq’), and therefore also the same extension as our moral vocabulary. Yet, these new concepts are also connected to motivation exactly in the same way as the second set of vocabulary (‘gooqq’ and ‘baqq’), and so in the same way as our vocabulary (‘good’ and ‘bad’) are connected to motivation. The problem unfortunately was that, no matter how hard Jane tried, this third set of vocabulary never caught on in the community either. It seems that Jane continued to be the only person who makes judgments to evaluate actions using terms of the third set of vocabulary (‘gooqqq’ and ‘baqqq’).

According to Bedke, it then seems obvious that, when Jane uses the third set vocabulary to make sincere evaluations, she is using them to express genuine moral judgments (Bedke 2019, 12). These judgments have exactly the same extension as our moral judgments and they are connected exactly in the same practical way to Jane’s motivations and reactive attitudes as our moral judgments are to ours. Yet, the problem is that the deferred internalists seem only to be able to agree with this if they make their view asymmetric in a strange way.

So far, the communal deferred internalists have always explained why an agent has ability to make genuine moral judgment in a certain specific way. Generally, according to them, whether an individual's judgment is a genuine moral judgment is determined by whether most people in the agent's community are motivated in the right way by their corresponding judgments. Due to this reason, when talking about the ordinary amoralists in the actual world (about those who are not motivated by their judgments about 'right' and 'wrong'), the communal deferred internalists can accept that these amoralists are making genuine moral judgments. They can make such judgments exactly because most other people's corresponding judgments in our shared community are related to motivation in the right way. On this view thus, the amoralists in our community inherit their ability to make genuine moral judgments from people who are motivated in the right way. This is still the case even if the amoralists' own judgments are motivationally inert.

Yet, there is a crucial problem with the previous way in which communal deferred internalist explain how an actual amoralist can still make genuine moral judgments. Although the previous explanation makes sense of certain cases of amoralists in our community, it contradicts with our intuition that Jane too can make genuine moral judgments when she employs 'gooqqq' and 'baqqq' in her newly invented language. In the case of the actual amoralists, communal deferred internalism implies that whether a moral judgment is genuine depends on the motivational profile of the community to which the agent belongs. If this were true generally, we should not accept that Jane's 'gooqqqness' and 'baqqqness' judgments are genuine moral judgments because absolutely no one else in Jane's community is motivated by their corresponding judgments that employ those concepts. As a matter of fact, no one else in Jane's community

uses those special concepts to make moral judgments. The rest of the residents of Amorality could not care less about the so-called 'gooqqq' and 'baqqq' actions.

However, the previous description of Jane reasonably suggests that we should firmly attribute genuine moral judgments to her when she thinks of actions as 'gooqqq' and 'baqqq' even if most of the other people in the Amorality community are not motivated by those judgments. Obviously, we believe that Jane's judgments are genuine moral judgments because of the motivational profile of Jane's own moral judgments themselves, rather than what judgments, if any, others in her community are motivated by. This explanation is very different from what the communal deferred internalists say about the actual amoralists.

To avoid this problem, the communal deferred internalists could suggest that, even if the amoralists' ability to make moral judgments depends on the motivations of other people in their community, the ability of the lone moralists to make moral judgments does not depend on the motivations of others (or the lack of them) in the same way. Yet, the resulting asymmetric view would just seem too *ad hoc*. The communal deferred internalists would fail to provide a unified explanation of what makes a given moral judgment a genuine one. To remain consistent, communal deferred internalists would have to argue that either both Jane's and the amoralists' capability to make genuine moral judgments depends on the motivational profile of the whole community or that in both Jane's and the amoralists' case their ability to make moral judgments depends only on their own motivational profiles. Unfortunately, neither of these alternatives is available for communal internalists. The first one would implausibly deny that Jane is making genuine moral judgments in the case above, whereas the second alternative would be to give up communal deferred internalism in the first place.

Likewise, individualist deferred internalism also faces a similar challenge regarding Satan's and other amoralists. In Section 6.3.3, I presented the case of the psychopath Anna, who is also an amoralist. In that case, Anna was not motivated to do what she thought she judged to be right. As explained in Footnote 36, psychopaths are often thought to remain unmoved by their moral judgments, because it is usually assumed that they are less concerned about the welfare of others. It is also often thought that psychopaths suffer from emotional deficits—they lack the ability to feel empathy, remorse, and even guilt. Let us imagine that, thanks to the new advances in medical technology, we become suddenly able to treat the emotional deficits of psychopaths such as Anna with an advanced brain surgery. Let us assume further that, after the surgery that fixes Anna's neural problems, she is immediately able to make moral judgments and also be motivated by these judgments in the same way as other ordinary moral agents.

We can then focus on the following modified version of the case already discussed in Section 6.3.3:

> Immediately after her surgery Anna once again finds a mobile phone that she intends to buy. The sales assistant introduces two phones of the same model to Anna. These phones are identical both in how they function and in their price. The only difference is that, if Anna buys the phone on the left, 20$ will be donated to starving children in Sudan. Anna takes herself to believe that not choosing this phone would be morally wrong, exactly in the same way as all ordinary people believe as well. Although Anna used to be a psychopath before, her neural problem has now been fixed. Because of this, Anna can make moral judgments and also be motivated by those judgments in the normal way.

> As a natural consequence of her moral judgment, Anna thus desires to buy the phone one the left. She even tells the sales assistant that she really cares about the welfare of the poor starving children in Sudan.

The description of Anna's behavior and motivations in this case clearly illustrates that, now immediately after her surgery, Anna can now be motivated by her moral judgments. This is why it is intuitively plausible to believe that Anna has made a sincere moral judgment immediately after her mental conversion.

Nevertheless, it does not seem that individualist deferred internalism could explain in the same, consistent way why we should also take Anna's first judgments immediately after his mental conversion to be genuine moral judgments. If we accept individualist deferred internalism generally, we should presumably treat Anna's case exactly in the same way as Satan's case above. When we do so, whether or not Anna's judgments immediately after her operation are genuine moral judgments would depend on whether or not her previous judgments were accompanied by motivation in a right way.

Yet, in the case we are now considering, we know that Anna's previous judgments did not even motivate her at all. As a consequence, a consistent form of individual deferred internalism would be required to claim that, just as Satan's new judgments are genuine moral judgments because Satan's previous judgments were motivational in the right way, Anna's new judgments cannot yet be genuine because her previous judgments were not connected to motivation in the right way. This result clearly contradicts our intuitive belief that Anna has made a genuine moral judgment in the case I just described.

In order to avoid this unintuitive consequence, the individualist deferred internalists again have to adopt an oddly asymmetric view. They would have to claim that whether a judgment that isn't itself related to motivation is a genuine one depends on the agent's previous judgments that were connected to motivation. Yet, according to this proposal, whether an individual judgment that is connected to motivation itself counts as a genuine moral judgment would not depend on how the agent's previous judgments were connected to motivation. Again, this view seems too *ad hoc*. It fails to provide a unified explanation of on what grounds a given judgment counts as a genuine moral judgment (and whether this depends on the previous judgments of the agent).

In order to avoid the previous problem, it again turns out that the individualist deferred internalists have to either claim 1) that both Anna's and Satan's ability to make moral judgments immediately after their conversions depends on how their previous judgments were connected to motivation or 2) that neither's ability to make genuine moral judgments depends on the past in this situation. The problem is that the first option unintuitively denies that Anna is currently able to make genuine moral judgments immediately after her conversion, whereas the problem with the second option is that it makes Satan unable to make genuine moral judgments after his conversion (and furthermore choosing the alternative requires giving up deferred internalism). To claim that Satan is able to make genuine moral judgments even after his conversion, however, was actually one of the key motivations to adopt individualist deferred internalism in the first place.

In this section, I have thus argued that both individualist and communal versions of deferred internalism fail. The communal deferred internalism cannot explain consistently why an agent can make genuine moral judgments even when the rest of her community is not motivated by the corresponding judgments. Likewise, the individualist deferred internalism cannot explain consistently why an agent can make a genuine moral judgment immediately after a conversion even if her previous moral judgments were not connected to motivation in the right way. These arguments suggest that we should not accept any deferred form of internalism at least when these views are formulated in terms of motivation.

The previous discussion of deferred and direct forms of internalism was formulated in terms of motivation even if in this thesis I am not defending any form of internalism with motivation (see Chapter 6 above). But rather what I am only defending is a version of internalism with dispositional desires. We can thus ask also whether a deferred or a direct form of internalism with dispositional desires would be more plausible. Deferred and direct forms of internalism with dispositional desires can be stated as follows:

> *Deferred internalism with dispositional desires*: Necessarily, if an agent judges that it is right to φ in circumstances C, then she has dispositional desire to φ or her moral judgment is connected in a certain way W to some other moral judgments that are accompanied by the relevant kinds of dispositional desires.

> *Direct internalism with dispositional desires:* Necessarily, if an agent judges that it is right to φ in circumstances C, then she has a corresponding dispositional desire to φ directly.

Here, I will only simply state that we should accept the direct form of internalism with dispositional desires rather than the deferred one. This is because the same objections I made to deferred forms of internalism with motivation also apply to deferred forms of internalism with dispositional desires. Deferred forms of internalism with dispositional desires too fail to explain whether an agent's ability to make genuine moral judgments depends on other people's moral judgments that are accompanied by dispositional desires. Deferred internalism with dispositional desires cannot explain consistently whether an agent's ability to make genuine moral judgments depends on her previous moral judgments that are accompanied by dispositional desires.

In contrast, direct forms of internalism with dispositional desires will not face the same problem. According these views, an individual judgment must be accompanied by a dispositional desire but not necessarily by motivation. Thus, a given judgment, such as Jane's judgment in the previous case, can count as a genuine one as long as it is accompanied by the relevant dispositional desire, even if others in the agent's community are not motivated by their otherwise similar judgments. This means that, on my view, whether an amoralist is able to make genuine moral judgments only depends on whether she has the relevant dispositional desires (for my argument to this conclusion, see Section 6.3.3 above). Due to this reason, whether Satan's and Anna's judgments after their conversions should count as genuine moral judgments depends consistently only on their respective current dispositional desires.[60] This is why my view is not vulnerable to the asymmetrical objections discussed above.

[60] For an internalist account of these types of cases, see Björnsson (2002).

Furthermore, direct internalism with dispositional desires is also compatible with other forms of internalism with dispositional desires I have defended. To accept direct internalism with dispositional desires is consistent with my argument for unconditional internalism with dispositional desires in Chapter 6 and strong internalism with dispositional desires in the first half of Chapter 7. These forms of internalism with dispositional desires actually all suggest that the genuineness of moral judgments only depends on the direct connection between moral judgments and dispositional desires.

7.8 Conclusion

In this Chapter 7, I continued my exploration of which form of internalism is the most plausible, which I already started in the previous chapter. In the first half of this chapter, I discussed how strong form of internalism we should accept. In Section 7.2, in order to remind my readers, I revisited my previous discussion of strong and weak internalism with motivation (see Section 2.4). As I was here more interested in different versions of internalism formulated in terms of dispositional desires, in Section 7.3, I then discussed three essential differences between motivation and desires: 1) they have different content, 2) motivation to do an action already includes a means-end belief (whereas dispositional desires to do actions do not), and 3) dispositional desires can manifest themselves in many different ways, whereas motivation only manifests itself by producing action. In Section 7.4, on the basis of the previous discussion, I then outlined three different ways in which the strength of dispositional desires can vary. In Section 7.5, I argued that a genuine moral judgment requires that the dispositional desire corresponding to it is also able to manifest itself by producing certain kinds of moral reactive attitudes and not just motivation.

The second half of this chapter consisted of Sections 7.6 and 7.7. In these two sections, I focused on whether direct or deferred forms of internalism with dispositional desires are more plausible. In Section 7.6, I examined the motivations which many philosophers have had for arguing for both individualist and communal versions of deferred internalism, as well as the central argument for those views. Then, in Section 7.7, I argued, following Bedke, that the arguments that were supposed to support deferred internalism turn out to be problematic—in slightly modified form the cases on which these arguments are based can be used to ground strong objections to deferred internalism. This is because deferred internalism with motivation will lead to having to adopt inconsistent *ad hoc* assumptions about when the genuineness of moral judgments depends on some other judgments. As these objections to deferred internalism with motivation also apply to deferred internalism with dispositional desires, I suggested that we should accept some version of internalism with dispositional desires.

Chapter 8: Conclusion

In this thesis, I have defended internalism, the view which claims that there is a necessary connection between moral judgments and motivation. More precisely, my final conclusion is that the most plausible form of internalism that we should accept by far is constitutional, unconditional, relatively strong, direct internalism that is formulated in terms of dispositional desires. Even if this form of internalism that I have ended up defending in this thesis is in many ways stronger than some of the other recently introduced internalist views, one advantage it has is that it is still able to accommodate all the famous externalist counterexamples. I have attempted to defend a stronger view of internalism mainly because it seems to me that many of the other sophisticated forms of internalism that have been recently defended have weakened the 'reliable connection' between moral judgments and motivation too much, or so I have argued in this thesis. In the rest of this concluding chapter, I will finally summarize the thesis by explaining how I reached my final conclusion.

In Chapter 2, I created a map of the logical space by explaining what the different forms of internalism there can be formulated are and how those views differ from externalism. I first introduced the basic terms 'moral judgments' and 'motivation' that both refer to certain kinds of mental states. By relying on these two concepts, I was then able to introduce the basic idea of motivational judgment internalism as the view that attempts to explain the reliable connection between moral judgments and motivation in terms of a certain kind of internal, modal connection between those states. The majority of Chapter 2 then went through different forms of internalism that have been widely discussed in contemporary metaethical literature (Sections 2.4-2.7). In these sections, I presented not only arguments for different forms of internalism, but also certain well-known objections based on various famous counterexamples

to them too. This is because trying to accommodate the externalist counterexamples has always been an important motivation for the internalists to put forward new, more sophisticated forms of internalism that are also often weaker too. On the basis of introducing these new views, I was then able to construct a map of logical space of what forms of internalism there can be. With the help of this map, we can not only locate existing forms of internalism and see how they differ from one another but also see new forms of internalism there could be that we could investigate further. Finally, after introducing the different forms of internalism carefully, in Section 2.9 I explained externalism as the main alternative to different versions of internalism.

In Chapters 3-5, I then argued that we should reject externalism. Given that externalism and internalism are two exclusive options—we have to accept one or the other, this provided my main argument to the conclusion that we should accept at least some form of internalism. The main argument I provided in these chapters against externalism and thereby for internalism was an attempt to defend and develop Smith's fetishism argument by showing that all the main externalist responses to this argument fail.

In Chapter 3, I first outlined the main idea of the fetishism argument. The fetishism argument starts from the observation that changes in our moral judgments normally cause changes also in what we are motivated to do. Although this observation is generally accepted by both internalists and externalists, Smith (1994) argued that only the internalists can provide a plausible explanation of the observed reliable connection between moral judgments and motivation. He claimed that the externalists would have to rely on something other than the moral judgments themselves (for example, an additional desire to do whatever is right) to explain that reliable connection. Smith then argued that the externalist explanations of that kind,

however, will be problematic. This is because ordinary moral agents would in the externalist framework end up caring more about an abstract property of moral rightness than ordinary concrete moral considerations that we usually think moral agents should be moved by. According to Smith, externalism thus turns moral agents into moral fetishists.

In Chapter 4, I discussed and argued against the externalists' objections to the fetishism argument that attempt to defend the previous externalist view of moral motivation. Many externalists have claimed that it is not problematic if moral agents are motivated by the additional desire to do whatever is right as long as they also have other more specific desires to do the things that are right. Moreover, many externalists have also suggested that a general desire to do whatever is right would also be useful and even required to motivate us in cases in which there is moral uncertainty, or we are strongly tempted to do something immoral. In response to these externalist views, I argued that it would be fetishistic to be motivated by the general desire to do whatever is right even if we had the specific desires to do the right things at the same time. I also argued that all the other externalist attempts to defend the explanation of the reliable connection between moral judgments and motivation that rely on the relevant *de dicto* desire fail for different reasons.

In Chapter 5, I discussed and argued against the externalists' attempts to explain the reliable connection between moral judgments and motivation in other ways that do not rely on the general desire to do whatever is right. In Section 5.2.2, I argued that Lillehammer's practicality option principle cannot explain the way in which both sides in a moral disagreement are supposed to be equally motivated by their respective moral judgments. Then, in Section 5.3.2, I argued that Cuneo's virtue-based alternative explanation cannot explain situations in which

an agent who is not fully virtuous can still be expected to be motivated to do what she judges to be right. Following Dreier, in Section 5.4.2, I argued that Copp's morally suggestible person model is also implausible because it unintuitively makes the prospect of changing one's moral views something to be afraid of. Lastly, in Section 5.5.2 I discussed Dreier's second-order desire model. I explained that having the relevant second-order desire to desire to do what one judges to be right should be deemed to be a requirement of rationality. I then argued that, because of this, the second-order desire model actually collapses into a form of internalism, and so it is not available for the externalists.

In Chapters 6-7, after having argued against externalism in the previous chapters, I then investigated different forms of internalism to find out which form of internalism would be the most plausible one. As a result of my exploration, I concluded that the most plausible form of internalism is constitutional, unconditional, relatively strong, direct internalism that is formulated in terms of dispositional desires.

In Chapter 6, I began with evaluating non-constitutional internalism. I first explained how views about different subject-matters are. As *de re* views are about certain specific mental states, constitutional forms of internalism and externalism have been the main focus of this thesis. As I also explained in the beginning of Chapter 6, the fetishism objection is a decisive objection to *de re* externalism. It shows that there must be some mental states such that they are internally connected to motivation. In contrast, non-constitutional forms of internalism and externalism are the *de dicto* views that are about whether the term moral judgment applies to states that are accompanied by motivation, irrespective what the nature of those judgments is. I then explained why I would remain neutral between *de dicto* internalism and externalism.

Then in the rest of Chapter 6, I defended unconditional internalism formulated in terms of dispositional desires (whereas usually both unconditional and conditional forms of internalism are formulated in terms of motivation). In Section 6.3.3, I explained how unconditional internalism with dispositional desires can deal with the traditional externalist counterexamples to the previous unconditional forms of internalism. My defence of this view is based on a range of thought experiments involving magic buttons. I argued that our widely shared intuitions about this kind of cases show that we do not think that an agent has made a genuine moral judgment unless she has a corresponding dispositional desire that can manifest itself by producing motivation at least in some cases.

In Chapter 7, I first explained how strong form of internalism we should accept. In Section 7.3, I discussed three fundamental differences between motivation and desires, which are 1) they have different content, 2) motivation to do an action already includes a means-end belief whereas dispositional desires do not include such beliefs, and 3) motivation can only manifests itself by producing motivation whereas dispositional desires can manifest themselves in many other ways too. These differences between motivation and desires further indicate that there are three ways in which the strength of dispositional desires can vary. They also tell us how one could argue for the stronger forms of internalism with dispositional desires. After examining each method, I concluded in Section 7.5, that a genuine moral judgment requires that the dispositional desire corresponding to it should also be able to manifest itself not only by producing motivation but rather also by producing certain kinds of moral reactive attitudes. This consequence gives us sufficient reason to accept internalism with dispositional desires which is slightly stronger in this one specific way.

In the rest of Chapter 7, I finally defended direct internalism with dispositional desires. I argued, following Bedke, that the traditional arguments which were supposed to support deferred internalism are problematic. The problem is that all the known forms of deferred internalism with motivation will have to adopt inconsistent *ad hoc* assumptions about when the genuineness of moral judgments depends on some other judgments and when they do not. I suggested that this objection to deferred forms of internalism applies to both communal and individualist versions of deferred internalism (and also whether these views are formulated in terms of motivation or dispositional desires does not make a difference either). As different forms of deferred internalism with dispositional desires are thus implausible, I believe that we have sufficient good reasons to accept a direct form of internalism formulated in terms of dispositional desires instead.

References

Aboodi, R. (2015) The Wrong Time to Aim at What's Right: When Is *De Dicto* Moral Motivation Less Virtuous? *Proceedings of the Aristotelian Society*, 115.3:307–314.

Aboodi, R. (2017) One Thought Too Few: Where De Dicto Moral Motivation is Necessary. *Ethical Theory and Moral Practice*, *20*(2): 223-237.

Aiken, H. (1944) Emotive Meanings and Ethical Terms. *The Journal of Philosophy*, 41 (17): 456-470.

Anscombe, G.E.M. (2000) *Intention*. 2nd ed. Cambridge, MA: Harvard University Press.

Aristotle (2000) *Nicomachean Ethics*. Crisp, R. (ed.). Cambridge: Cambridge University Press.

Arpaly, N. (2002). *Unprincipled virtue: An inquiry into moral agency*. Oxford: Oxford University Press.

Bedke, M. (2009) Moral Judgment Purposivism: Saving Internalism from Amoralism. *An International Journal for Philosophy in the Analytic Tradition*, 144 (2): 189–209. doi:10.1007/s11098-008-9205-5.

Bedke, M. (2019) Practical Oomph: A Case for Subjectivism. *The Philosophical Quarterly*, 1–21. doi:10.1093/pq/pqz024.

Bjömsson, G. (2002) How Emotivism Survives Immoralists, Irrationality, and Depression. *Southern Journal of Philosophy*, 40 (3): 327–344. doi:10.1111/j.2041-6962.2002.tb01905.x.

Björnsson, G., Björklund, F., Strandberg, C., et al. (2015) *Motivational Internalism*. Oxford: Oxford University Press.

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

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

Blair, R.J.R. (1995) A Cognitive Developmental Approach to Morality: Investigating the Psychopath. *Cognition*, 57 (1): 1–29. doi:10.1016/0010-0277(95)00676-P.

Brandt, R. (1979) *A Theory of the Good and the Right*. Oxford: Clarendon Press.

Brink, D. (1986) Externalist Moral Realism. *The Southern Journal of Philosophy*, 24 (5): 23-41.

Brink, D. (1989) *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Brink, D. (1997) Moral Motivation. *Ethics*, 108 (1): 4–32. doi:10.1086/233786.

Bromwich, D. (2016) Motivational Internalism and the Challenge of Amoralism. *European Journal of Philosophy*, 24 (2): 452–471. doi:10.1111/ejop.12053.

Carbonell, V. (2013) *De dicto* desires and morality as fetish. *An International Journal for Philosophy in the Analytic Tradition*, 163 (2): 459–477. doi:10.1007/s11098-011-9825-z.

Cholbi, M. (2006a) Belief Attribution and the Falsification of Motive Internalism. *Philosophical Psychology*, 19 (5): 607–616. doi:10.1080/09515080600901939.

Cholbi, M. (2006b) Moral Belief Attribution: A Reply to Roskies. *Philosophical Psychology*, 19 (5): 629–638. doi:10.1080/09515080600901954.

Cholbi, M. (2011) Depression, Listlessness, And Moral Motivation. *Ratio*, 24 (1): 28–45. doi:10.1111/j.1467-9329.2010.0048.x.

Cooper, J. M. and Hutchinson, D. S. (Eds.). (1997). *Plato: Complete Works*. Indianapolis: Hackett Publishing.

Copp, D. (1995) Moral Obligation and Moral Motivation. *Canadian Journal of Philosophy*, 25 (1): 187–219. doi:10.1080/00455091.1995.10717438.

Copp, D. (1997) Belief, Reason, and Motivation: Michael Smith's 'The Moral Problem.' *Ethics*, 108 (1): 33–54. doi:10.1086/233787.

Copp, D. (2001) *Morality, Normativity, and Society*. Oxford: Oxford University Press.

Cuneo, T. (1999) An Externalist Solution to the 'Moral Problem.' *Philosophy and Phenomenological Research*, 59 (2): 359–380. doi:10.2307/2653676.

Dancy, J. (2000). "Should we pass the buck?" In O'Hear, A. (ed.), *Philosophy: The good, the true, and the beautiful*. Vol.47. Cambridge: Cambridge University Press.

Darwall, S. (1983) *Impartial Reason*. Ithaca N.Y.: Cornell University Press.

Dorr, C. (2002) Non-cognitivism and Wishful Thinking. *Noûs*, 36 (1): 97–103. doi:10.1111/1468-0068.00362.

Dreier, J. (1990) Internalism and Speaker Relativism. *Ethics*, 101 (1): 6–26. doi:10.1086/293257.

Dreier, J. (2000) Dispositions and Fetishes: Externalist Models of Moral Motivation. *Philosophy and Phenomenological Research*, 61 (3): 619–638. doi:10.2307/2653615.

Eggers, D. (2015) 'Unconditional Motivational Internalism and Hume's Lesson.' In Björnsson, G., Björklund, F., Strandberg, C., et al. (eds.) *Motivational internalism*. Oxford University Press. pp. 85–107. doi:10.1093/acprof:oso/9780199367955.003.0005.

Eriksson, J. (2006) *Moved by Morality: An Essay on the Practicality of Moral Thought and Talk*. Department of Philosophy. Uppsala University.

Eriksson, J. (2014) Elaborating Expressivism: Moral Judgments, Desires and Motivation. *Ethical Theory and Moral Practice*, 17 (2): 253–267. doi:10.1007/s10677-013-9434-3.

Gibbard, A. (1990) *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Clarendon Press.

Gibbard, A. (2003) *Thinking How to Live. Cambridge*, MA: Harvard University Press.

Goldman, A. (1970) *A Theory of Human Action.* Princeton, NJ: Princeton University Press.

Hare, R.M. (1952) *The Language of Morals*. Oxford: Clarendon Press.

Hare, R.M. (1963) *Freedom and Reason*. Oxford: Oxford University Press.

Hooker, B. (2000) *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*. Oxford: Clarendon Press.

Hume, D. (2007) *A Treatise of Human Nature: A Critical Edition, Volume 1*. Norton, D.F. and Norton, M.J. (eds.). Oxford: Oxford University Press.

Kauppinen, A. (2007) The Rise and Fall of Experimental Philosophy. *Philosophical Explorations*, 10 (2): 95–118. doi:10.1080/13869790701305871.

Kennett, J. and Fine, C. (2008) 'Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths'' In Sinnott-Armstrong, W. (ed.) *Moral Psychology. Volume 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press. 173–190.

Kennett, J. and Fine, C. (2008) 'Could there be an empirical test for internalism?' In Sinnott-Armstrong, W. (ed.) *Moral Psychology. Volume 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press. 217–225.

Korsgaard, C. (1996) 'Reflective endorsement.' In O'Neill, O. (ed.) *The Sources of Normativity*. Cambridge: Cambridge University Press. 49–89.

Korsgaard, C. (1986) Skepticism about Practical Reason. *The Journal of Philosophy*, 83 (1): 5–25. doi:10.2307/2026464.

Kripke, S. (1980) *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lenman, J. (1999) The Externalist and the Amoralist. *Philosophical quarterly of Israel*, 27 (3–4): 441–457. doi:10.1007/BF02383189.

Lillehammer, H. (1997) Smith on Moral Fetishism. *Analysis*, 57 (3): 187–195. doi:10.1093/analys/57.3.187.

Marks, J. (1986) *The Ways of Desire: New Essays in Philosophical Psychology on the Concept of Wanting*. Chicago, Ill.: Precedent.

McDowell, J. and McFetridge, I.G. (1978) Are Moral Requirements Hypothetical Imperatives? *Proceedings of the Aristotelian Society*, Supplementary Volumes, 52: 13–42.

McDowell, J. (1979) Virtue and Reason. *The Monist*, 62 (3): 331–350.

Mele, A.R. (1996) Internalist Moral Cognitivism and Listlessness. *Ethics*, 106 (4): 727–753. doi:10.1086/233670.

Mele, A.R. (2003) *Motivation and Agency*. Oxford: Oxford University Press.

Miller, A. (1996) An Objection to Smith's Argument for Internalism. *Analysis*, 56 (3): 169–174. doi:10.1093/analys/56.3.169.

Miller, A. (2013) *An Introduction to Contemporary Metaethics*. 2nd ed. Cambridge: Polity.

Nichols, S. (2002) How Psychopaths Threaten Moral Rationalism: Is it Irrational to Be Amoral? *The Monist*, 85 (2): 285–303. doi:10.5840/monist200285210.

Nichols, S. (2004) *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.

Nussbaum, M. (1999) *Sex&Social Justice*. Oxford: Oxford University Press.

Prinz, J. (2006) The emotional basis of moral judgments. *Philosophical Explorations*, 9 (1): 29–43. doi:10.1080/13869790500492466.

Putnam, H. (1973) Meaning and Reference. *The Journal of Philosophy*, 70 (19): 699–711. doi:10.2307/2025079.

Roskies, A. (2003) Are Ethical Judgments Intrinsically Motivational? Lessons from 'Acquired Sociopathy.' *Philosophical Psychology*, 16 (1): 51–66. doi:10.1080/0951508032000067743.

Roskies, A. (2006) Patients with Ventromedial Frontal Damage Have Moral Beliefs. *Philosophical Psychology*, 19 (5): 617–627. doi:10.1080/09515080600901947.

Roskies, A. (2007) 'Internalism and the evidence from pathology' In Sinnott-Armstrong, W. (ed.) *Moral Psychology. Volume 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press. 191-206.

Sayre-McCord, G. (1997) The Metaethical Problem. *Ethics*, 108 (1): 55–83. doi:10.1086/233788.

Schmidtz, D. (1994) Choosing Ends. *Ethics*, 104 (2): 226–251. doi:10.1086/293599.

Shafer-Landau, R. (1998) Moral Judgement and Moral Motivation. *Philosophical Quarterly*, 48 (192): 353–358. doi:10.1111/1467-9213.00105.

Shafer-Landau, R. (2003) *Moral Realism: A Defence*. Oxford: Clarendon.

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

Smith, M. (1995) Internal Reasons. *Philosophy and Phenomenological Research*, 55 (1): 109–131. doi:10.2307/2108311.

Smith, M. (1996a) Normative Reasons and Full Rationality: Reply to Swanton. *Analysis*, 56 (3): 160–168. doi:10.1111/j.0003-2638.1996.00160.x.

Smith, M. (1996b) The Argument for Internalism: Reply to Miller. *Analysis*, 56 (3): 175–184. doi:10.1093/analys/56.3.175.

Smith, M. (1997) In Defense of 'The Moral Problem': A Reply to Brink, Copp, and Sayre-McCord. *Ethics*, 108 (1): 84–119. doi:10.1086/233789.

Smith, M. and Harcourt, E. (2004) Instrumental Desires Instrumental Rationality. *Proceedings of the Aristotelian Society*, Supplementary Volumes, 78: 93–129.

Smith, M. (2008) 'The Truth about Internalism.' In *Moral Psychology, Volume 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press. 207–215.

Stevenson, C.L. (1937) The Emotive Meaning of Ethical Terms. *Mind*, 46 (181): 14–31.

Stocker, M. (1979) Desiring the Bad: An Essay in Moral Psychology. *The Journal of Philosophy*, 76 (12): 738–753. doi:10.2307/2025856.

Strandberg, C. (2012) Expressivism and Dispositional Desires. *American Philosophical Quarterly*, 49 (1): 81-91.

Stratton-Lake, P. (2000) *Kant, Duty and Moral Worth*. London: Routledge.

Stratton-Lake, P. (2002). "Introduction." In Stratton-Lake, P. (ed.), *Ethical Intuitionism: Re-evaluations*. Oxford: Clarendon Press.

Strawson, P.F. (2008) *Freedom and Resentment and Other Essays*. Oxford: Routledge.

Suikkanen, J. (2005). Reasons and Value—in Defence of the Buck-passing Account. *Ethical Theory and Moral Practice* 7(5), 513-535.

Suikkanen, J. (2014) *This Is Ethics: An Introduction*. Chichester, England: Wiley-Blackwell.

Svavarsdóttir, S. (1999) Moral Cognitivism and Motivation. *The Philosophical Review*, 108 (2): 161–219. doi:10.2307/2998300.

Svavarsdóttir, S. (2006) How Do Moral Judgments Motivate? In *Contemporary Debates in Moral Theory*. Oxford: Blackwell. 163–181.

Tiberius, V. (2015) *Moral Psychology: A Contemporary Introduction*. Oxford: Routledge.

Tolhurst, W. (1995) Moral Experience and the Internalist Argument against Moral Realism. *American Philosophical Quarterly*, 32 (2): 187–194.

Toppinen, T. (2004) Moral Fetishism Revisited. *Proceedings of the Aristotelian Society*, 104 (1): 307–315. doi:10.1111/j.0066-7373.2004.00095.x.

Toppinen, T. (2015) 'Pure Expressivism and Motivational Internalism.' In Björnsson, G., Björklund, F., Strandberg, C., et al. (eds.) *Motivational Internalism*. New York: Oxford University Press. 150–166.

Tresan, J. (2006) De Dicto Internalist Cognitivism. *Noûs*, 40 (1): 143–165. doi:10.1111/j.0029-4624.2006.00604.x.

Tresan, J. (2009a) Metaethical Internalism: Another Neglected Distinction. *An International Philosophical Review*, 13 (1): 51–72. doi:10.1007/s10892-008-9042-y.

Tresan, J. (2009b) The Challenge of Communal Internalism. *The Journal of Value Inquiry*, 43 (2): 179–199. doi:10.1007/s10790-008-9141-9.

van Roojen, M. (2015) *Metaethics: A Contemporary Introduction*. Oxford: Routledge.

United Nations Children's Fund (2016) *Female Genital Mutilation/Cutting: A Global Concern* [Online]. [Viewed 24 November 2019], Available from: https://www.unicef.org/media/files/FGMC_2016_brochure_final_UNICEF_SPREAD.pdf

Wallace, R.J. (2006) 'Moral Motivation.' In *Contemporary Debates in Moral Theory*. Oxford: Blackwell. 182–196.

Wedgwood, R. (2004) The Metaethicists' Mistake. Philosophical Perspectives, 18, 405–426.

Wiggins, D. (1991) Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs. *Proceedings of the Aristotelian Society*, 91: 61–85.

Williams, B. (1981) *Moral Luck: Philosophical Papers, 1973-1980*. Cambridge: Cambridge University Press.

Wolf, S. (1993). *Freedom within Reason*. Oxford: Oxford University Press.

Zangwill, N. (2003) Externalist Moral Motivation. *American Philosophical Quarterly*, 40(2), 143-154.

Zangwill, N. (2008) The Indifference Argument. *An International Journal for Philosophy in the Analytic Tradition*, 138 (1): 91–124. doi:10.1007/s11098-006-9000-0.