Motives, reasons, and morality: the Humean theory of motivation, reasons internalism, and the problem of objective morality

Elliot Cross

University at Albany, State University of New York, ecross@albany.edu
MOTIVES, REASONS, AND MORALITY:
THE HUMEAN THEORY OF MOTIVATION, REASONS INTERNALISM,
AND THE PROBLEM OF OBJECTIVE MORALITY

by

Elliot Cross

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Arts and Sciences
Department of Philosophy

2014
In this dissertation I argue that the Humean theory of motivation, and reasons internalism ought to be accepted and are compatible with moral objectivity. The Humean theory of motivation is the view that only desires, and not beliefs, are capable of motivating. Reasons internalism is the view that only those considerations capable of motivating can be reasons. The combination of these two views entails that we only have a reason for action when we have a relevant desire. Accordingly, if we hold both of these positions then we cannot have reason to be moral unless we desire to be moral. This seems to conflict with the idea that morality is objective and applies to everyone.

Firstly, I argue that desires are dispositions to feel attraction to certain things. Next, I argue for Humeanism. To have a desire is to be in a state that we seek to make the world fit, as opposed to beliefs, which we attempt to make fit the world. Because of this, desires and not beliefs motivate. I then proceed to argue against judgment internalism. This position holds that judgments about moral reasons are necessarily motivating. I argue that such judgments are merely usually motivating, because most people have a desire to be moral, which causes them to be motivated by what they take to be their moral reasons, but it is possible to make a moral judgment and not be motivated by it. I proceed to defend reasons internalism, and argue against the claim that legitimate reasons need not be capable of motivating. I argue that no consideration can be a reason for action if we cannot act on it, which means that it must be capable of motivating us.

I then reject some potential approaches to making objective morality, Humeanism, and reasons internalism compatible, before bringing together Humeanism, internalism, and objective morality, and demonstrating that they can be understood as
mutually consistent, if we reject the claim that morality is necessarily reason giving. I argue that we may have no reason to be moral, and yet may still be immoral even then.
# Table of Contents

Abstract........................................................................................................... ii

**Introduction** ................................................................................................. 1

**Chapter 1: Definitions** ................................................................................. 4

1. Desire ........................................................................................................... 4
2. The Object of a Desire .................................................................................... 5
3. The Feeling of Desire .................................................................................... 6
4. Persisting Desires .......................................................................................... 8
5. The Origin of Desires ................................................................................... 14
6. Whims .......................................................................................................... 17
7. Beliefs .......................................................................................................... 20
8. Motivation .................................................................................................... 24
9. Reasons ........................................................................................................ 30
10. Conclusion .................................................................................................. 32

**Chapter 2: The Humean Theory of Motivation** ............................................. 33

1. Introduction .................................................................................................. 33
2. Desires, Beliefs, and Direction of Fit ............................................................. 35
3. Direction of Fit: Beliefs ................................................................................ 36
4. Sobel and Copp's Objection ......................................................................... 40
5. Direction of Fit: Desires .............................................................................. 42
6. The Realist View of Desirability .................................................................. 45
7. Direction of Fit Applied to the HTM .............................................................. 50
8. Changing Our Beliefs .................................................................................. 54
9. Mele’s Objection ......................................................................................... 57
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 3: Judgment Internalism</td>
<td>64</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>64</td>
</tr>
<tr>
<td>2. Why Judgment Internalism?</td>
<td>65</td>
</tr>
<tr>
<td>3. Limitations on Any State Having Both Directions of Fit</td>
<td>67</td>
</tr>
<tr>
<td>4. Direction of Fit of Judgments</td>
<td>70</td>
</tr>
<tr>
<td>5. That People Usually Are Motivated by Their Judgments</td>
<td>74</td>
</tr>
<tr>
<td>6. Amoralism</td>
<td>76</td>
</tr>
<tr>
<td>7. Failing to Judge Correctly</td>
<td>82</td>
</tr>
<tr>
<td>8. Reasons</td>
<td>88</td>
</tr>
<tr>
<td>9. Motivation Again</td>
<td>89</td>
</tr>
<tr>
<td>10. Conclusion</td>
<td>90</td>
</tr>
<tr>
<td>Chapter 4: Reasons Internalism</td>
<td>92</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>92</td>
</tr>
<tr>
<td>2. Reasons</td>
<td>92</td>
</tr>
<tr>
<td>3. Backgrounding Conditions</td>
<td>94</td>
</tr>
<tr>
<td>4. Williams' Argument for Internalism</td>
<td>96</td>
</tr>
<tr>
<td>5. Sound Deliberation from the Subjective Motivational Set</td>
<td>101</td>
</tr>
<tr>
<td>6. Unwanted Desires and Whims</td>
<td>107</td>
</tr>
<tr>
<td>7. Reasons Externalism</td>
<td>108</td>
</tr>
<tr>
<td>8. McDowell's Objection</td>
<td>110</td>
</tr>
<tr>
<td>9. Reasons Prior to Desires</td>
<td>111</td>
</tr>
<tr>
<td>10. Korsgaard's Objection</td>
<td>115</td>
</tr>
<tr>
<td>11. Nagel's Objection</td>
<td>116</td>
</tr>
<tr>
<td>12. Conclusion</td>
<td>118</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>-----</td>
</tr>
</tbody>
</table>

**Chapter 5: The Problem**  
1. Introduction                                      | 119 |
2. Similar Desires, Similar Reasons                 | 124 |
3. Constitutivism: Velleman                          | 130 |
4. Social Basis of Shared Desires                    | 133 |
5. Hobbes                                           | 136 |
6. Schroeder                                        | 143 |
7. Constitutivism: Korsgaard                         | 146 |
8. Conclusion                                        | 150 |

**Chapter 6: Morality and Reasons**  
1. Introduction                                      | 152 |
2. "Ought" Being Concerned with Reasons              | 153 |
3. The Inescapability of Morality                    | 158 |
4. No Reasons for Desires                            | 161 |
5. Practicality                                      | 163 |
6. Morality and Rationality                          | 165 |
7. What We Can Say About a Non-reasons Based Morality| 166 |
8. Further Objections                                | 170 |
9. Why Be Moral                                      | 174 |
10. Conclusion                                       | 178 |

**Bibliography**                                     | 180 |
Introduction

In this dissertation my aim is to show how it is possible to hold three particular, popular positions – the Humean Theory of Motivation, reasons internalism, and moral objectivity – simultaneously. This dissertation, in solving this problem, can be viewed as having three subgoals. Firstly, to argue for the Humean Theory of Motivation. Secondly, to argue for reasons internalism. Thirdly, to argue that morality need not necessarily provide reasons to everyone.

Moral objectivity is a very popular position and perhaps more intuitive than moral subjectivity. The Humean Theory of Motivation is still a widely held opinion, holding that desires and not beliefs motivate. Reasons internalism, as defined by Williams, holds that only considerations which are capable of motivating at least after sound deliberation can be reasons, and there is much intuitive appeal to this claim as well, because what we care about seems to be what decides what reasons we have. Individually, then, all three of these theses are appealing, however the three theses are in tension with each other, and it appears that we may be forced to give up one of them. I seek to avoid having to surrender one of these theories, and instead explain how they can be made compatible with each other.

The problem with holding all of these theories together is if reasons internalism and the Humean Theory of Motivation are correct, then it appears that while reasons can be obtained from those considerations which have a connection to our desires, no beliefs about a consideration will be sufficient to turn that consideration into reasons. For if desires motivate and beliefs do not, and if reasons must be capable of motivating, then reasons can be based on desires but not on our beliefs. While beliefs can direct our motivations towards certain ends, our ends themselves are not dependent on our beliefs but on our desires. We might then conclude that should somebody not desire a
moral end, then she would have no reason to acquire that moral end, and should she not desire to do the moral thing generally, then she might have no reason to ever be moral. Should she believe that a certain action is a moral action, it need not follow from that that she has any reason to act in that way. Should somebody, however, have no reason to act in a particular way, then we may well conclude that it cannot be said that she acts wrongly for not doing what she has, after all, no reason to do. This seems incompatible with moral objectivity, which holds essentially that should two people in the same circumstances act identically, then either both act rightly or both act wrongly. It holds that the fact that two people have two different sets of desires should not have the result that acting in the same ways in the same circumstances may be immoral for one of them and not immoral for the other.

While there is not space in this dissertation to argue for the existence of objective morality, I do argue for the Humean Theory of Motivation and reasons internalism, in order to dismiss the solution to this problem of rejecting one of these theses. My suggested solution instead is to deny that morality necessarily provides reasons. This is not a popular position, although there have been those who have previously held it, like Foot,¹ but nevertheless it is the correct one. Morality not being reason-giving is not the problem that it might initially appear, as most of what we want to be true about an objective morality can still be established even granting this point. It is perfectly possible to judge that somebody has acted wrongly, even if he had no reason to act otherwise, just as we can judge that somebody acted cruelly or illegally, even if she had no reason not to so act, for example. Being reason-giving is not, then, necessary to be objective. In other words, morality can still apply to everyone, even if it doesn’t give them reasons.

¹ "Morality as a System of Hypothetical Imperatives", in 2002.
In the opening chapter, I will discuss some initial concepts, such as what desires and beliefs are, as well as a few other important entities for my project, as foundation for the arguments to come.
Chapter 1: Definitions

Much of my position that is to follow will depend on my concept of desires, and so I will first explain what I take them to be. I don't want to examine in depth the issues surrounding the exact nature of beliefs, since that is a very large and contentious issue in itself and unnecessary for this project, but some discussion of the nature of beliefs is necessary insofar as I must mark off the features of beliefs that distinguish them from desires and make them unsuitable to be the basis for motivation, although the actual arguments for that latter position won't be discussed until the following chapters. Finally, a brief discussion of the basic idea of a motivation and a reason will also follow.

1. Desire

There are three features of desires which seem to me to be obvious and fundamental, which are as follows: desires have an object; desires involve a feeling towards this object; and desires are exist in us even when not in the forefront of our minds. The latter two features appear to be in tension with each other, as we wish to say that a desire requires a feeling and yet continues to exist even when no feeling is present. My concept of desires is based in part on an attempt to resolve this tension, but I do not think it is an artificial move to try and combine these two features into one account as I think these are the features that we naturally think desires possess, and I am attempting to produce a concept of desire which basically matches our intuitions about what desires are.

The account I will give of desires is that they are dispositions to be attracted towards, or repulsed from, certain objects. If we have a desire for an object, then we feel attracted to that object when it is in mind, but the desire persists even when the object is not in mind. The three basic elements of desire are therefore: an object; a feeling of attraction towards (or repulsion from) that object; and a disposition to have
this feeling in certain situations. I will discuss these three elements in turn, explaining why I think each one is a necessary component of desire.

2. The Object of a Desire

Desires have an object, which is the thing that we desire. An object of desire is obviously part of the concept of a desire, as any desire is a desire for something. This "for" should be construed broadly, in the sense that a desire might be "for" not getting something, i.e. an aversion to that object. In any case, it's incoherent to think of a desire without an object – that would be a "wanting to get" with nothing that we want to get. A "wanting to get" that has no object, insofar as that is even possible, should be described as mere restlessness rather than a proper desire.

The object of a desire, in the sense of the thing that we desire, can be virtually anything. We will generally have a representation of that object in mind to occasion the feeling of our desire, but this representation is not the object of desire itself, as clearly it would generally be incorrect to say that we feel attracted to the representation of what we desire – we are attracted to the thing itself. I want a nice house, for example, not the representation of a nice house. The object, then, is something that we will have a representation of which features in the mental occurrence of a desire, but that representation is not itself the object of a desire. The use of the term "object" should not be taken to be limiting the focus of a desire to a concrete entity; in fact, generally the object will be something like some state of affairs we are attracted to or repulsed by. Of course we can, and do, have desires which are for some state of affairs that does not obtain, in which case we are led to the slightly strange conclusion that the object of a desire need not necessarily exist for us to have a desire for it. Such desires are so common and familiar that I don't find this conclusion terribly concerning; it seems that as long as we can think of it, we can have a desire for it. The objects of desires could
be physical objects, like a jewel or another person; something non-physical, like recognition or achievement; or something fairly nebulous, like that another person be happy. It can be very narrow, like the desire for one specific painting, or very, very broad and wide-ranging, like the desire for world peace. I won’t draw any boundaries on what we can desire, as the possible objects of desires are innumerable, and I couldn’t possibly even categorize them all. Therefore, I will try to frame my ideas in such a way that they are compatible with whatever the object of a desire may be.

A desire has been “satisfied” when we have acquired the object of that desire, although that is not necessarily to say that the desire is thus completed and goes away once it is satisfied. Some desires will be completed and dismissed in that sense, like a desire to climb Mount Everest once, while others will remain or return after a time, like the desire to read a book or the desire to keep one’s current job. Desires are, though, the kinds of things which are theoretically satisfiable, although they might in actuality be unsatisfiable; by this I mean that all desires could theoretically be satisfied by acquiring their object, but some objects may in actuality be impossible for us to acquire, like world peace.

3. The Feeling of Desire

A desire includes a feeling about the object of the desire, either a feeling of attraction towards or a feeling of repulsion away from that object, which we have when the object is in mind. Although these are two phenomenologically different kinds of desire, one with a feeling of attraction (“I want this”), the other with a feeling of repulsion (“I don’t want this”), for simplicity’s sake I will generally talk about desires of attraction, since these kind of desires are what we typically think of when we talk about desires. The necessity of having this feeling whenever the object is in mind is ceteris paribus, as I will explain in the next section. It’s difficult to define this feeling without lapsing into
circularity (e.g. "the feeling of desire is the feeling of wanting something"), so I will simply say that the feeling of attraction is primitive. I assume that the reader is familiar with this feeling of wanting something when presented with something we find desirable and won’t find such a move too troubling. It is not necessary for me to analyze this feeling any further, and I suspect it is impossible to do so anyway.

By "when the object is in mind" I mean when it is in front of us in the world or when we are thinking about it, such as when we remember it or imagine it. In other words, the object is in mind either when we recall the mental representation we have of the object, or when we directly perceive the object. When we see a book we want, we will feel attracted towards it and want to get it, and if we remember that book later we will feel much the same way, although probably less strongly because the presence of the book to our mind is less strong, being only a memory rather than direct perception. The strength of the feeling will also vary between different times, for whatever reason: sometimes a desire to walk will be stronger than a desire to drive; at other times the desire to walk will be weaker than the desire to drive. This difference in strength of attraction might depend on the presence or absence of other desires that have some kind of relation to the object of that particular desire, thus lending their strength of feeling to the first desire, e.g. a desire to get some fresh air will strengthen a desire to walk. The strength of feeling also varies between desires for different objects in general, as some objects we simply want more than others. The feeling of wanting to have a family would be stronger in most people than a feeling that they want to plant a tree. This difference in feeling can be one factor that causes particular desires to vary in strength, but it is probably not the whole explanation of what causes one desire to be stronger than another in the sense of one desire overriding other desires. Consider that sometimes desires which are only faintly felt can win out over other more strongly felt desires, for example, a feeling of repulsion towards a horrible tasting medicine is
probably more strongly felt than the feeling of attraction towards a more nebulous concept of general healthiness, and yet usually we will say that we more strongly desire to be healthy than to avoid the medicine. This is not something I need to analyze here, and it may be impossible to do so in a systematic way since what causes one desire to win out over another likely depends on the various features of individuals' characters and sets of desires and so on, and so is different for each person.

I don't intend to talk about the various strengths of desires any further. It is not necessary to my project to discuss how people actually come to order their desires, or how they in fact reason and come to decisions about what to do; I am instead concerned with the relationship between desire, motivation, and reasons, and about what constitutes a reason in general. I do not need to discuss the particulars of the decision making process and how people form particular intentions for particular actions in order to do this. The gap between motivation and action can be a large one, and explaining it is not a task I wish to take on here.

4. Persisting Desires

Desires, as I've said, include a feeling of attraction we feel towards certain objects. That cannot, however, be all there is to say about them, because defining desires merely in phenomenological terms leads to the conclusion that we lose our desires whenever we are not experiencing the feeling associated with them. If desires are just a feeling of attraction towards some object and nothing more, then the desire only exists so long as the feeling does, but while there is no feeling that we constantly experience, there are many desires that we would want to say we constantly have (if and when we have them). If desires only consisted of some feeling, then it would be impossible to say that we continue to have any desire when we are not feeling anything towards the object of the desire, which would be almost all the time since we cannot
have more than a few objects in mind at any given moment, but generally have many desires. This would lead to the conclusion that any desires we had, we would have only very intermittently and rarely, when we happened to think about what it was that we desired. This is not a conclusion to be welcomed. This would mean, for example, that when we are unconscious we no longer have a desire to live, or even that whenever we are simply not thinking specifically about our continuing to live (i.e. almost all the time), we have no desire not to be killed, and this is obviously incorrect. We do not think that a person crossing the street thinking about a meeting she's going to has no desire not to get run over and killed, or that she has no desires of any kind outside of those relating to her meeting. The common thought is that we have our desires whether or not we are actively thinking of them all the time. Nonetheless it is only when we think about the object of a desire that we feel an attraction towards it. Whenever the object isn't in mind, we don't have the feeling of attraction towards it, which is hardly surprising as it would be very strange if we were to have a feeling of attraction towards an object which wasn't present in our minds to have a feeling about. What I want to establish is that this lack of feeling does not mean that we no longer have that desire, but that we have the desire even when not having that feeling – in other words, we still desire an object even when not thinking about it. If desire fundamentally includes this feeling of attraction towards an object, but must be present even when this feeling is not, the solution must be that desires are fundamentally dispositions, specifically, a disposition to have this feeling of attraction (or repulsion) whenever the object of the desire is being thought about or directly perceived.

It is necessary then to say something about what I understand by the term “disposition”. Fundamentally I accept Martin's “Finkish” account of dispositions, which is widely, although not universally, accepted, and which holds that dispositions are real.

---

2 It's not unusual to make this point, for example, Fehige in Millgram 2001, p 50.
properties that cannot be understood merely counterfactually or as reducible to other, non-dispositional, properties. Martin gives an example of a live wire to explain what he has in mind when calling dispositions real, non-counterfactual properties. Typically the description of the wire being live will be a dispositional one, specifically, that the live wire will shock anyone who touches it, so it has a disposition to shock if touched, regardless of whether anyone actually does touch it, and so is live. Martin rejects concepts of dispositions which would explain this property of the wire counterfactually, that is, that having the disposition of being live means that if it was in a situation in which somebody touched it, then the person would be shocked. Counterfactual explanations are appealing as explanations of dispositions because we want an account of dispositional properties that does not rely on the disposition actually being activated, for example, we want an account of the wire being live that does not depend on somebody actually touching it, as whether or not somebody touches the wire has nothing to do with it being live, and therefore it's natural to think in terms of what would happen if it were to be touched and use this as the basis for an account of dispositions. However, Martin feels that this does not go far enough to capture the nature of dispositions as real properties – or as "bare dispositions," because it fails to account for a situation where the wire will not shock anyone even if touched, and yet is still live. Martin imagines a device attached to the wire, which he dubs an "electrofink", which immediately cuts the power if somebody touches the wire, which means that if someone touches the wire she will not be shocked. On the counterfactual account the wire attached to an electrofink is not a live wire, since there will be no possible world in which the wire, so modified, would shock someone, so it is not true that if somebody touches the wire, she will be shocked. Martin argues that we think the wire can still be justified.

---

3 Lewis 1997, p 143.
4 As explained in Martin 1994, p 2-3.
5 McKitrick 2003, p 350.
live regardless of whether or not anyone would ever be shocked by it, and therefore the disposition exists even if the counterfactual account fails. To put it another way, we surely still think that there is a real difference between the live wire attached to the electrofink, and a wire that is actually dead. The counterfactual account of dispositions cannot capture this difference, as in both scenarios it is true that somebody touching the wire would not be shocked, and so they are identical for the counterfactual account.

Martin thinks that our intuitions are that being live is a real property of the wire that exists beyond a mere account of what could happen in certain circumstances.

I draw similar conclusions to Martin's about when it can be said that we have a desire. Even if the object of a desire never came to our mind, and so we never had the feeling of attraction towards it, in fact even if it were impossible for us to have a feeling of attraction towards it, we could still very well have a desire about it. In order to ascribe the possession of a desire to us it is not necessary that we ever actually think of the object of that desire and feel attracted to it. It is possible, for example, that a particularly unreflective individual might go through his entire life never once thinking about his own mortality, or there might even be somebody who can't think about her mortality, thus never manifesting their desire to live, but this does not necessarily mean that such people have no desire to live. It's not too far-fetched to think of a person who never thinks of his own mortality: he could live a safe life where nothing ever seriously threatens him, and then dies suddenly from a heart attack, never having occasion to think of his own mortality, and yet nonetheless he could have a desire not to die. Somebody who can't think about her own mortality could be somebody who lacks the capacity to ever think about her own mortality, such as somebody suffering from very serious mental illness or brain damage, which leaves her with no concept of who she is or the passage of time or the like, but is still capable of engaging in some activities and enjoying some things (i.e. not somebody who is completely lacking in all higher brain functions).
Such a person might be incapable of ever thinking about the end of her life, and there would be no counterfactual story where she does so while remaining as she is, yet nonetheless it seems inaccurate and unintuitive to describe her as lacking any kind of desire to go on living. If we killed such a person it would still seem to be accurate to say that it could not be considered a matter of indifference to that person that she had been killed, or that she did not desire to live anyway. Or imagine that somebody has some kind of mental feature similar to the electrofink, say a smoker who has been hypnotized to feel a great feeling of aversion towards cigarettes whenever she thinks about smoking one. I think it would still be correct even in that circumstance to say that the smoker has a desire to smoke a cigarette the feeling of which is being blocked by the mental conditioning, because there is a difference in the desires of the non-smoker who has never desired to smoke and those of the conditioned smoker. So it seems correct to say that somebody may have a desire for something even if she was prevented from ever feeling any attraction towards it. These dispositions are a real property that we may possess even if the feeling that we have a disposition to experience never actually occurs.

I say this to clarify my thinking about desires, but I do not wish to make my position rest on the presence of such unfelt desires, and indeed I will try to avoid talking any further about them. This is because I do not wish to simply assume that people have all manner of desires that they never feel, as that would be too convenient for whatever motivational theory one might wish to establish. It would be implausible for me to claim that everybody has certain desires even if they don't feel that they do, claiming that their feeling these desires is being prevented. It is hard to see how we would ever know that we have or don't have certain desires if we never experience the

---

6 A person who was wouldn't have any desires. Desires are mental states, and so somebody with no mental states has no desires.
feeling of attraction, however, I think it is at least possible that we could have desires that we never become aware of, where the feeling of attraction never arises.

There is also some debate on whether dispositions are to be understood as reducible to physical properties, which some have argued convincingly against, claiming that in fact dispositions often cannot be reduced to some base physical properties.\(^7\) The primary reason for believing in irreducibility is that certain fundamental, which is to say non-reducible, physical properties like the charge of a particle can only be understood in terms of dispositions: the negative charge of an electron can only be described in terms of how the particle behaves, and the effects that it produces, and that is to provide nothing more than a dispositional account of the charge.\(^8\) Since the negative charge of an electron is a fundamental property, and the dispositional explanation is the most fundamental account of that property, we have a disposition that cannot be reduced into further physical properties – in fact we might conclude that dispositions are the most fundamental physical accounts that can be given. It does not seem necessary (nor possible) for me to settle this question here, as this is a question that requires a good deal of work in metaphysics, but for whatever impact it may have on my arguments I will state that I accept the argument that dispositions cannot always be reduced to more fundamental physical properties.

Dispositions, then, are real properties, not necessarily reducible to other properties, expressed as tendencies for certain things to happen in certain circumstances, and they still exist even if those circumstances never occur, or are prevented from occurring, and the disposition is thus never expressed. I hope not to discuss the true nature of dispositions any more than this to avoid being side-tracked into metaphysical issues. Hopefully the dispositional nature I ascribe to desires is

\(^7\) E.g. McKitrick, 2003.  
\(^8\) Ibid, p 355-356.
general enough to be acceptable to most, even to those who hold a different conception of dispositions in general.

One final point here: above I noted that the necessity of feeling a certain way towards the object of a desire when the object is being thought about was *ceteris paribus*. As with the electrofink it’s possible that some other factor like depression could prevent us from feeling our usual feeling of attraction to something which is the object of one of our desires. Just as the electrofink blocks the live wire from actually shocking anyone while the disposition of being a live wire still remains, so too depression could block someone from actually feeling attracted to the object of her desire without that removing the desire itself from her. Such cases are borderline cases, but it is plausible to say that if we normally feel attracted to an object when it is in mind, then even if right now the object doesn’t produce the required feeling, we still have a desire for it provided that there is some circumstantial factor that is preventing us from feeling the way we usually do.

5. The Origin of Desires

I have discussed what desires are, and I will in later chapters discuss what role they play, but there is also the question of how they come to be, of why we form the desires we do. The origin of desires is not something that my account will directly rest on, so I don’t wish to discuss this question in great depth, however some arguments later will touch on this issue so I must say something about it. I will say that generally we gain desires through experiencing what certain things are like, in combination with having our own individual characters which lead us to feel attracted to certain things.
and not other things.\textsuperscript{9} It’s obvious that somebody with few experiences will have few desires, because there are few objects that such a person is capable of bringing to mind and thus potentially being attracted to or repulsed by.

While typically a desire will be gained by experiencing some object and discovering if we like it or not, this is not the only way desires can be gained. Some of our desires may be produced by other desires rather than by experience with new objects, for example, desires can be produced as means to some end. Typically, when we have a desire for some object, coming to believe that another object is the means to acquiring the former is, \textit{ceteris paribus}, going to produce an attraction for that object as well. Of course, this desire will be for that object considered as a means to our end and not for the object considered as itself, in other words, the object which is a means will not be desired for its own sake, but solely as a means to our end. These desires for means are derivative of the original desire for the end and their existence depends on that desire. For example, if I want to go to Timbuktu, I will need money to do that, and so I will desire money as the means to being able to go to Timbuktu, but only insofar as I need the money to go there. If I come to no longer want to visit Timbuktu, or if I win a free holiday there, I will also no longer desire to get the money to do so, obviously. A common means necessary for achieving any desired ends is continuing to live. Assuming my desire to visit Timbuktu is of the typical sort where I want to see the sights and so on, clearly in desiring this I desire that I be alive at the time I go there. There are many, many other desires which also have as means our being alive, for example, acquiring a certain job will require me to be alive to enjoy it. Most people don’t generally make this explicit in their thoughts, and take living for granted unless there is doubt over the question of whether they will be alive at that point. So although the

\textsuperscript{9} Milgram 1997, chapter 3 and throughout, argues for this method of desire formation; although I agree with this general conclusion, I do not agree with all the particulars of his argument.
feeling part of the desire to live may never have been experienced, nevertheless it seems that virtually anyone who has any desires centered on themselves must have a desire to live, at least until the completion of those desires. Since it is not necessary to know that we have a desire in order to have that desire, we can assume that most people have some kind of desire to live if they have desires that require it, which most people do, even if they don’t think in these terms. Of course the desire to live as a means is not the only desire concerning our continuing to live, as most people have a desire to live just for its own sake as an end.

What about cases where we know that something is a means to our end and circumstances are such that we could take these means, and yet we refuse to do so? Does this mean that we don’t have such desires? No, rather the desire to take some means can conflict with some other desire, which may outweigh the first desire in our practical reasoning. It is of course possible that we have a desire of repulsion towards those means considered just as themselves, not in the context of being means to some end, and this desire outweighs our desire to take these means for the sake of our end. I mean that if the means to some end we want, are something themselves that we strongly don’t want, then we may not end up acting upon those means because the desire not to do so is stronger. For example, if acquiring some scientific data we desire requires conducting painful tests on animals, and we desire not to cause suffering to animals, then we may refuse to take these means to our end. Therefore, it is not true that we always take the means to our ends, or indeed that we always feel attracted to

---

10 This argument is similar to Williams’ argument on the evil of death in "The Makropulos Case", in 1976. I qualify that virtually everybody will have the desire to live in order to leave room for unusual cases I may not be thinking of, and also for the possibility that somebody could have some desire which is predicated on their dying. Certain desires about self-sacrifice, or the cessation of suffering obviously come to mind, but somebody could also have a desire like "I want to be remembered after I die", which of course also requires their death at some point. 11 Assuming, of course, they are not suffering from some kind of irrationality or other unusual mental state.
the means, if a feeling of repulsion towards them is more strongly felt than any potential attraction.

Let me sum up my definition of desires then. Desires are dispositions to have a feeling of attraction (or repulsion) towards some object whenever that object is in mind, either by being recalled or by being directly perceived. The object is what that feeling of attraction is aimed at, and to say that desires are dispositions is to say that the desire persists between, before, and after particular instantiations of the feeling of attraction or repulsion.

6. Whims

The positions I will argue for, namely the Humean Theory of Motivation, and reasons internalism, will mean that desires will become the basis for reasons. This view is often held to be vulnerable to an objection which has been called the problem of "too many reasons". That is, the position of desire based reasons allows certain mental entities to give rise to reasons which we don't think in our commonsense thinking are reason-giving, thus, we are left with a theory in which we appear to have more reasons than we intuitively think we do, and so something must be wrong with the theory. Specifically, whims and passing fancies and such are not usually considered as providing reasons to do something (that I have the sudden urge to throw myself from the top of a tall building when I find myself on top of such a building is not thought to give me an actual reason to jump), yet it might seem that making desires the source of our reasons commits us to making these reason-giving too as they appear to essentially be kinds of desire. Why whims should not be considered reason-giving is, however, not generally discussed; it usually appears to be taken as obvious. My reply to this

---

12 Schroeder 2007, p 84.
objection (which I am not hugely concerned by) will come in chapter 4, but I should say something about what whims are first.

Whims clearly are things which are similar to desires, in that they include a feeling of attraction to some object. The whim to jump from someplace high is a feeling of wanting to jump, so a feeling of attraction to the thought of jumping, or something like that. However, desires are dispositions to have such feelings whenever the object is in mind, meaning not just when the object is physically present before us, but also when we think about the object of the desire. Whims, on the other hand, are not dispositional, and therefore they are not identical to desires. Thus any theory that desires are reason-giving does not make whims reason-giving by definition, since whims are something different from desires; although the conclusion that whims are reason-giving may still follow depending on why it is that desires are reason-giving. Whims fail to be dispositional because we do not necessarily feel the required attraction whenever the object is in mind, even ceteris paribus. I have in the past felt an urge to jump from a tall place, yet now when I think about being on top of a building, I do not feel any attraction towards jumping, and in fact I more consistently feel no attraction, indeed an aversion, towards doing so. We do not always experience the same whims even when the objects of previous whims are directly in front of us – I do not feel an urge to jump every time that I am atop someplace high. Furthermore, we do not say that we have whims that we are not currently feeling, as we do with desires; if a whim is not being felt it doesn’t exist, simple as that. Someone who is asleep currently has no whims, but does have desires. So the dispositional nature of consistently feeling the same way towards
certain objects, and of being present even if the feeling is not, is missing, and thus
whims cannot be quite the same thing as desires, even if closely related.¹³

This distinction between desires and whims is intuitive; after all when we think
about whims we usually distinguish them from desires as different kinds of mental state.
On what grounds do we usually see a difference between desires and whims? Partly it
is the very fact that we think they are not reason-giving, whereas we are more likely to
think that our desires do give us reasons, but this cannot be the entirety of the reason
since certainly some of those who completely reject desire based reasons still wish to
keep a distinction between whims and desires. I think that in large part what seems
different about these two entities is precisely that desires are dispositional and whims
are not, that is, that whims are not something we consistently experience as desires
are. We refuse to accept whims as important precisely because they are "passing
fancies", things that we only feel sometimes, usually for a limited duration,¹⁴ that aren't
necessarily aimed at the sort of objects we'd feel comfortable saying "this is the kind of
thing I want". That I had a whim to do something, we might say, does not mean I really
wanted to do it, because it's just something that came over me inexplicably, and is not a
proper desire of mine. I think this distinction follows quite naturally from our usual
conceptions of whims and from what I have said about desires.

¹³ Sometimes we’d be inclined to call such things "urges" rather than whims, for example, if I
have the urge to throw a glass in anger. Whims typically have a connation of not being serious
or harmful, unlike what we call urges, but I take it that urges in this sense are just the same
thing as what I’m calling whims: some mental entities which are similar to desires, but not
dispositional. For convenience’s sake, I will use the term “whims” to refer to both.
¹⁴ Sometimes cases like these are whims or urges, but other times they may be desires which
are only felt in rare situations. For example, I may desire to punch someone who is irritating
me, despite never normally being violent, after being sufficiently provoked. This could be a
mere urge, or it could be a desire which generally is just not felt because the circumstances in
which the disposition asserts itself are very rare (e.g. we have a disposition to feel attracted to
the idea of punching someone who sufficiently provokes us). There is nothing else special
about these desires, however, and what I’ve said so far about desires will apply to them.
7. Beliefs

A proper, full discussion of the controversial topic of beliefs is outside the scope of this dissertation, but I must say something to distinguish them from desires, and explain what I take them to be to enough of an extent to explain why I shall argue in the next two chapters that they cannot give rise to motivation on their own. I will attempt to take a basically neutral account of beliefs which would be accepted by a large number of philosophers, rather than trying to create and argue for my own original account of beliefs. There are two basic elements which appear to be constitutive of beliefs: that they can be true or false and aim at being true, and that they consist in our having a stance towards a proposition.

What's obviously fundamental about beliefs (and central to my arguments about them) is that they, as Williams puts it, "aim at truth." The purpose of beliefs is to capture what is true, and their success in doing so is what we primarily judge them upon, and whatever state they are will be one that we will, as far as possible, attempt to make track what we take to be true. Mental states which are not sensitive to the truth in this way, which we don't attempt to make track truth, simply are not beliefs. Other considerations – for example, their utility to us, or how they are formed, and so on – are secondary for evaluating beliefs compared to their success at capturing the truth. This is the most important distinguishing factor between beliefs and desires, as assignments of truth or falsity are not appropriate for desires (as I've described them), as they only consist of a disposition of attraction or repulsion to something, as opposed to beliefs where truth values are central.

---

15 Williams 1976, p 136.
16 I will argue for this claim in section 3 of chapter 2, however I think it is also intuitive.
17 Smith 1995, p 7. Being attracted to something simply does not have a truth value, although obviously the fact that I am or am not attracted to something can be either true or false.
However, this sensitivity to truth does not yet give us an account of what kind of mental state beliefs are. Sometimes beliefs have been analyzed in terms of dispositions to act, or in some other behaviorist or intentional manner, where to believe something simply means to have a propensity to make certain assertions or act in a certain way; in which case there would be basically nothing to say about belief as a mental state. Such accounts, while undoubtedly capturing part of what is important about the role of beliefs, seem unsatisfactory as a complete explanation of what beliefs are, as they have, for example, difficulty explaining the difference between someone who has a belief that she never acts on and somebody who doesn’t have that belief (and thus obviously would never act on it). Since there does appear to be a real difference between these two people, a difference in what is going on in their minds and not merely a difference in dispositions, I will not rely on a merely behaviorist type of account. It is true that how we judge what beliefs a person has depends at least in part on how she acts, e.g. that somebody is crossing a bridge probably allows us to conclude that she believes the bridge is safe to cross, but then it seems to me that this is true of any mental state at all, that we judge their existence in other people by looking at their actions, for there is no other way for an internal mental state to be evinced but in some kind of external action or expression. I won’t discuss this feature of beliefs, then, since as it is common to all mental states it will not pick out any feature that is constitutive of being a belief in particular.

The typical contemporary view of belief is that to believe something is to have a stance on the truth value of some proposition, that is, to take an attitude towards that

---

18 Ayer attributes a position like this to Wittgenstein, for example. Ayer 1985, p 58. Dennett has a more modern and subtle version of this view, 1987.
19 Williams 1976, p 144.
proposition concerning its truth or falsity. For example, to believe that Paris is in France is to take the proposition that “Paris is in France” to be true. Just as a desire must be for something, so too a belief must be about something, otherwise we would have a belief without believing anything: in other words, a belief must have an object, and this object is usually taken to be a proposition. A proposition is whatever is it that is expressed by a sentence which has a truth value. The stances which it is possible to take on a proposition, and thus the kinds of beliefs that we can have about them, are not limited to merely binary truth/falsity, but can and do also include varying degrees of certainty, such as believing that something is probably true, or that it is almost certainly false, or even that we don’t know either way, and so on. However, what’s vital is that the stances that we can take on these propositions are all concerned with whether or not the proposition accurately expresses the nature of the world, that is, whether or not it’s true or false, and are not the kind of feeling of attraction or repulsion that desires involve. They are factual, not affective. The stances we can take in beliefs vary only in the degree of certainty of the truth value that we are assigning to the proposition, not in the kind of value we are assigning to the proposition, since they all concern truth values. It might be objected that it is also possible to take a stance towards a proposition like “hoping X is true” or “wishing X is false”, which involves something more than just pure truth value determination. However, such states are not constituted only by beliefs, but are best explained as being combinations of belief and desire: for

20 E.g. Fodor has a version of this view, 1981. See also: http://plato.stanford.edu/entries/belief/
21 Schwitzgebel, http://plato.stanford.edu/entries/belief/ It’s not necessary for me to give a proper account of propositions here, as my arguments will rely on the act of believing rather than the object of belief. The Russelian view is that propositions include objects and their properties, while the Fregian view is that they are the senses of sentences. Either one is acceptable for my arguments.
example, "wishing that p is just desiring that p and believing that p is very unlikely."²² Our actual stance towards the proposition remains a purely truth value stance; the belief that goes into the state of hoping that X is true, for example, has a purely true/false stance, probably that it is uncertain whether the proposition is true or false, and is combined with a desire that proposition be true.

It is sometimes claimed that having a belief involves having in our minds some kind of representational content about the state of the world:²³ Russell says, for example, that "belief consists in an idea or image combined with a yes-feeling."²⁴ Fodor thinks that to have a belief is technically to think something about such a representation, and the representation in turn "expresses the proposition."²⁵ I’ll leave this issue aside here, as it is not important for my purposes, and is rather controversial. It has also often been argued that beliefs must be dispositional,²⁶ usually for much the same reason as desires (e.g. a sleeping person still has beliefs). This may or may not be correct, however it is clear from what has previously been said that they are a quite different kind of dispositions than desires. Nothing in my arguments will depend on whether or not beliefs are a kind of disposition.

In short, then, to have a belief is to take some stance about some proposition, which is to say, to hold a proposition to be true or false with some degree of certainty.

²² Smith 1995, p 117. Smith provides this as a possible explanation, without fully committing himself to it.
²⁴ Russell 1948, p 164.
²⁵ Fodor 1981, p 201. He feels that since a proposition is not a mental entity, it cannot be what is directly present in our minds for us to have a belief about, and therefore the medium of representation is a necessary part of belief.
²⁶ E.g. Russell 1948, p 162.
8. Motivation

I should say something here about what we take motivation in general to be, as groundwork for my following argument that desires can motivate while beliefs cannot. Motivation is some mental state that comes before intentional action, is relevant to that action being done on purpose, and acts as an explanation for our doing it. That motivation has to do with intentional action is fairly obvious, and I will offer some arguments below for this claim. Motivation must have to do with action as the main way we judge that people are motivated to do something is that they do it, and people who claim to be motivated to do something and yet never do it, without there being something which explains why they did not do it (in the sense that they would have acted that way if this factor had not been present), we view as not being motivated. Motivation also contributes to making our doing an action understandable, as we cite our motivation as what immediately led us, what caused us, to act. Nothing which fails to explain why we would act in this sense could be a motivation, therefore for a neutral preliminary account here I'll say that motivation is some mental state which would, in the right circumstances, cause us to act, and in so doing make our action understandable as being done by an agent. I stipulate that motivation causes action in the right circumstances because there can be genuine motivations that we do not act upon. Being motivated to do a particular action does not inevitably cause us to do that action; motivation can be prevented from causing action by numerous other factors, including other stronger motivations, or certain mental or physical states. As an obvious example, no matter how motivated you are to go jogging, if you don't have any legs you're not going jogging. Motivation will also not cause action if we lack relevant beliefs concerning which actions are appropriate and how those actions can be done, but, had the right circumstances been present, the motivation would have led to action. In these cases where a particular action doesn't occur, it is most reasonable to think...
that the mental states which would have explained what we would have done in those
different circumstances are also motivations. I will also argue that desires can properly
fill the role of motivation while beliefs cannot, but that conclusion will be established in
later chapters.

I think that if there is a typical view of what motivation is, it would involve
reference to causation, although it would perhaps not be exhausted by such a
reference. There are metaphysical issues with causation (such as what precisely
causation is) that I cannot get into here however, so if that is unacceptable I would
suggest the following rough account which tries to avoid mention of causation:
motivation is the mental state that would be followed by our doing a particular action in
any situation where we are not prevented from acting in that way, and further,
motivation is what makes this action comprehensible as being done purposefully by an
agent.\(^{27}\) That latter clause makes sure that accidents are not counted as motivated,
because they lack this kind of explanation.

There are many things that could be said to cause or explain an action, such as
our personality in general, our representations of the world, what other people do, and
so on, but these are not usually considered to be motivations. We can start to rule
them out by stipulating that motivation is the immediate cause of action, that is, there is
no further intermediate cause between motivation and action. Between action and our
character, our beliefs, and so on, there is a gap where our doing an action is still not
explained until a motivation is provided, but there is no explanatory gap between
motivation and action. Once we know that somebody is motivated to do something in

\(^{27}\) A motivation will make actions understandable in the sense that we can see that it led to a
person doing such a thing intentionally, but not necessarily in the sense of making the action
seem rational. Somebody could easily have irrational motivations: for example, somebody
wearing a tinfoil hat may offer as a motivation that he wants to stop the aliens from reading
their minds, and this makes sense of his action in the way that I mean, but certainly doesn’t
make it rational.
particular we require no further explanation between the motivation and action about why he actually acted in that way, if he did act in that way (but we would need an explanation if he didn't act in that way). We might well question how someone became motivated to do that action in the first place, which is what I will discuss in the next chapter, or criticize the motivation for various reasons, but we don't question why an action followed from a motivation to do that action.

Only intentional actions are motivated; we do not think that it makes sense to ask for a motivation for accidents and reflexive actions, such as instinctively pulling your hand back from the fire, or turning on the traffic light in ignorance. What precisely intentions are is too big and contentious a topic to discuss here, however it obviously includes (or is) the idea of purposefulness – actions done by accident or mindlessly are not intentional actions, nor are they actions for which it would generally be appropriate to ascribe a motivation. I am not attempting to comprehensively argue for this element of motivation here, I'm simply appealing to our general notion of motivation, which I think includes the idea that only those actions we consciously mean to do are motivated actions. Jumping at a loud noise is not a motivated action.

The fact that motivation must explain our doing of an action introduces a further limitation on what can count as a motivation, specifically, that the mental state we claim as our motivation must be relevant in some way to the action being done by us, because a mental state which is not relevant to our acting cannot help explain the action. This is also implied by the above condition that the action must be an intentional one – an unintentional action is partially distinguished by the fact that the mental state within us that caused the action is not one that helps explain the action as being done by some agent (e.g. a merely instinctive reaction). Agents are capable of reasoning

---

about their actions, and there must be some kind of deliberative link between the motivation and the action such that it is comprehensible to us, given certain facts about the agent, that the motivation would make the agent consciously act the way he or she did.

Certain mental states like moods and emotions cannot sufficiently make sense of actions to be motivations by themselves. If someone smashed a glass because he was furious, while we might accept that in an everyday conversation as an explanation of the action, based on our previous experience of such emotions, a more complete explanation would have to include something like “he wanted to relieve his anger”. If we did not already know that breaking something is sometimes cathartic, it would be difficult to see the connection between anger and the destruction of personal property, which we might otherwise expect to only make him angrier (since people usually dislike having their property broken), and so the emotion would not explain the action to us. Moods and emotions are not directly motivations themselves, then, as a reference to a desire (and usually a related belief) is needed to complete the picture, but they are obviously closely linked to motivation. They can, I think, give rise to new desires or whims, or cause existing desires to manifest themselves, which could in turn be the motivations for action: being in a good mood can often cause us to feel attracted to helping other people, and likewise my anger towards someone can cause me to desire doing something bad to him. These desires could be the motivations for our actions, and so emotions and moods can certainly have an important role in intentional action. To examine these mental states fully is beyond the scope of this project, so in short my position is that they can (and often do) give rise to desires which will themselves be

---

29 They can also give rise to unintentional, and thus unmotivated in my sense, actions, like being rude to someone while in a bad mood without intending to be rude.
motivations to act, for example admiring someone may give rise to a desire to help them, without themselves quite being motivations.

As Dancy notes, it is often difficult to separate motivation from other mental entities when attempting to define it. It's easy to identify motivation with desires or even beliefs, depending on your position about which one causes motivation. Motivation is also particularly difficult to separate from reasons, since both seek to answer the question of why we acted the way we did. One important difference between reasons and motivations is that while both do seem to serve this explanatory role, only reasons can act in a justificatory role in explaining why an agent acted the way she did. A motivation by itself does nothing to justify a course of action, it can only (partially) explain why it happened in those circumstances. A motivation can be immoral or irrational, although generally speaking most people try to be somewhat rational, if not also moral, in their motivations. While not justifying, a motivation is typically present for every intentional action, as lacking a motivation would make it more difficult to understand why an agent did a particular thing instead of another thing, yet for any intentional action there is clearly a meaning as to the question of why the agent acted in such a way. We might notice that those who have a realist view about reasons, the view that being reason-giving is a property of objects in the world and not any kind of mental state, have an easier time distinguishing reasons from motivations, since motivations are surely mental states, and reasons, according to this view, are not. However, even such realists would have to provide some account of motivation, as some motivational state is required to bridge the gap between the recognition of a  

\[30\] Dancy 1994-5.  
\[31\] Irrational motivations weaken the explanatory role for motivations, but I think that is an acceptable conclusion. For example, a severely mentally ill person's motivations will be less comprehensible than a more rational person's, but then we should expect that to be the case.  
\[32\] Hursthouse, 1991, makes a fairly convincing case for a class of intentional actions which may complicate the standard account that intention entails there being a reason and motive present.
reason and an actual action, because a mental state must be provided to explain our action, and that state acts as the motivating state. I do not know of anyone who is a realist about motivations in the same way that some people are realists about reasons. Motivations are mentally internal states, not real entities in the outside world. So the realist about reasons does not necessarily have less of an explanatory burden here.

There are cases that make the nature of motivation unclear, for example, a case of idle fidgeting while bored seems neither quite motivated nor exactly unintentional and unmotivated, but falls somewhere between the two. We may not be fully aware of what we’re doing while we’re fidgeting, and yet we can provide a kind of explanation, such as being bored, as to why we were doing it when pressed. That seems to make these actions motivated in some sense, but they don’t seem quite like the more self-consciously motivated actions. Obscure cases like these probably arise because the operation of our minds is often mysterious to us and mental states are not always entirely separate from each other; in other words, awkward borderline cases are to be expected. In addition, as Goldman has pointed out, we almost never consciously practically deliberate in the way accounts of practical reason often assume we do. We don’t usually bother to figure out a precise intention or motivation for what we’re doing, particularly if an action is unimportant in scope or the occasion to act is immediate (we are much more likely to do so when thinking about long-term plans for our future, for example). Because we often don’t deliberate, it may not always be clear with some actions whether an action is unintentional – and thus the motive is unknown or nonexistent in the usual sense, as is arguably the case with fidgeting – or whether it is an action where the motive could possibly be identified by the agent upon further reflection. This vagueness is not ideal, but I can only make the old appeal that we cannot look for more precision in an enquiry than is possible.

---

Motivation then, in conclusion, is the mental state which would be, in the right circumstances, immediately prior to an intentional action and relevant to the action, and thus explains why we acted.

9. Reasons

Here I will only make some very basic points about what reasons are as most of my discussion of reasons requires argument and must wait until later chapters. Typically, reasons are viewed as considerations that weigh in favor of doing an action. Saying they "weigh in favor" isn't a particularly enlightening thing to say, but we're dealing with fundamental concepts and it's hard to get any more precise in discussing what reasons are — in fact it has been claimed that the concept of reasons is primitive.\(^\text{34}\) I won't go quite that far, as I think there are some useful things that we can say about reasons, for example, whether reasons are considerations or some kind of mental state, such as desires. I will hold that reasons are considerations, as opposed to saying that they are desires themselves, since there are many convincing objections to that latter claim.\(^\text{35}\)

Reasons are sometimes split into two different types, one being the explanatory (or motivating) reasons, and the other being the justificatory (or normative) reasons. Both kinds of reasons are answers we might give for why we did an action, but answer two different senses of the question "why?". Explanatory reasons are those which explain why we actually did a particular action as opposed to doing any other action; these would be the considerations that, in some sense, lead to our motivation to do it and are the considerations that we actually acted upon. For example, that my car is unsafe can be a motivating reason for me to sell it. Justificatory reasons are

\(^{34}\) For example, Scanlon, 1998, p 17.  
\(^{35}\) Goldman, 2005, discusses some common ones. p 508-509.
considerations that endorse a particular action, in the sense of justifying it by reference to some standard, such as rationality or morality, and so explain why it might be correct to say we ought to do that action. That my car is unsafe can be a reason why I ought to get rid of it.

Dancy argues, plausibly I think, that there really aren't two types of reasons here but rather two different questions that could be asked about our reasons: one question about which consideration(s) would explain our acting, and one about which would justify it. Certain considerations can be cited as answers to why we acted the way we did for either or both of these questions, in other words, the same consideration can be an explanatory reason and/or a justificatory reason, depending on how we take it and how we act upon it. The unsafe car above is an example of the same consideration being cited in these two different ways. Alternatively, there could be two different considerations for the same action which produce a different explanatory and justificatory reason for it, for example, I may help a stranger in need because it pleases me (explanatory) and/or because it is my duty (justificatory). I will view these two different kinds of reason as Dancy suggests, accepting that considerations can act in either role depending on the situation. Considerations that we actually act upon are our explanatory reasons, and considerations that make it, for whatever reason, appropriate to act in a certain way would be justificatory reasons. Obviously any justificatory reason we actually act on will also be an explanatory reason, although we may have justificatory reasons that we don't act upon and thus are not additionally explanatory reasons. Any explanatory reason may or may not also be a justificatory reason. My point is that there is no strict line we can draw to say that in general “these types of considerations only count for explanatory reason and these only for justificatory”.

---

Reasons are considerations that explain why we acted in one of two senses, either in the explanatory sense or in the justificatory sense. There is obviously much more to be said about reasons, and I will return to them in chapter 4.

10. Conclusion

To sum up this chapter, desires are dispositions to feel attracted to or repulsed by certain objects. Whims (or urges) are similar to desires in that they incorporate a feeling towards an object, but they are not dispositional. Beliefs are stances towards propositions, and lack this feeling of attraction altogether, being solely concerned with truth values. Motivations are mental states which provide a certain explanation for our doing an action. Reasons are considerations that explain why we acted in either an explanatory or a justificatory sense.

These are what I take to be my starting positions, and in the chapters to come I will argue for further conclusions about these entities and how they relate to each other. In the next chapter I will argue that desires can be the source of our motivations, and that beliefs cannot be.
Chapter 2: The Humean Theory of Motivation

1. Introduction

In this chapter and the next I will argue that desires motivate, while beliefs do not. Desire based motivations, I will establish, are just those desires which can serve the explanatory role that motivations have towards our actions. I will argue for the Humean Theory of Motivation (HTM), the theory that desires are motivations, and further that beliefs are not motivations.

Humeans argue that whenever motivation is present, that motivation must have been produced by some kind of a desire. I will not hold such a strong position, but will rather simply argue that desires motivate and beliefs do not, and this is what I will mean when I talk about the HTM. In other words, I will leave open the possibility that some third kind of mental state could motivate; in fact, I believe that only desires can motivate, but it’s not possible to run through every possible mental state and argue that none of them could motivate within this dissertation. So I will leave that possibility aside as, in any case, the primary debate around moral motivation is usually whether or not beliefs can motivate, or whether desire must also be present. I will deny that beliefs can produce motivation, as Nagel for example holds,37 or at the very least produce motivation on their own without a desire also being present. Beliefs do have an important role for action on the typical HTM view, that of directing our action, and I also take this position on the roles of desires and beliefs towards action. This view basically holds that desires determine our ends, and beliefs direct us on the means we could take to achieve those ends, so desires are motivation producing, and beliefs are motivation directing. I will argue that beliefs cannot produce motivation simply because they are the wrong kind of thing to motivate, an argument which will be primarily based on the

37 Nagel 1979.
on the different directions of fit of these two mental states, as I will explain, and is drawn largely from Smith.\(^{38}\) The direction of fit distinction I will discuss in detail below, but it is essentially the view that when we have a desire we naturally try to make the world fit the desire, and when we have a belief we try to make the belief fit the world. Therefore, desires and not beliefs are suitable for being the direct causes of action, since action involves changing something in the world and only desires call for change in the world. In this chapter I will establish that desires have the appropriate direction of fit while beliefs have the other direction, and explain how that makes desires suitable for motivation. The next chapter will argue against the position that certain judgments might have both directions of fit, and thus those judgments can motivate us instead of desires.

A brief note here about the name of the Humean Theory of Motivation: while it is inspired by Hume’s writings on motivation, whether Hume himself actually held this position as it has come to be understood, and what he took his position on motivation to actually entail, is open to serious debate.\(^{39}\) This issue of interpretation is not one that I take a stance on, as I am only concerned with whether or not the position is itself true and not whether it is an accurate reflection of Hume’s thinking. The position referred to by this name is at this point much more than a claim about what Hume thought; in fact it is often not intended to be an actual claim about what Hume himself thought at all. The truth of the HTM has sometimes seemed so obvious to many who support it that full arguments for it are often not given,\(^{40}\) although in the face of increasing opposition this has become more necessary than ever.

---

\(^{39}\) E.g. Cohon 2008.
\(^{40}\) For example: "There is a famous sentence: 'Reason is and ought only to be, the slave of the passions.' This sentence...expresses a view to which I, like every man who attempts to be
2. Desires, Beliefs, and Direction of Fit

The nature of desires and beliefs is fundamental to establishing which of them can motivate, so a brief recap of what I’ve already said about them is appropriate here. In the last chapter, I argued that desires were dispositions to have a feeling of attraction or repulsion to an object, and I took the usual position on beliefs to be that they are propositional attitudes, which is to say, we have a belief when we have a position on the truth value of a proposition. Recall too that motivations are the mental states prior, relevant to, and explanatory of an intentional action.

The basic idea of direction of fit comes from Anscombe, famously with her example of the shopping list. Let’s say a man goes shopping with his shopping list in hand and with the intention of buying everything on his shopping list, but on returning home he realizes he forgot to buy something that is on the list. There are two possible responses he could make to rectify this situation and bring it about that he has purchased everything on the list: firstly, he could go and purchase the remaining item; secondly, he could remove that item from the list. Both responses make it so that he has purchased everything on the list, and yet both are clearly very different responses. The first response, to make some change to the world and acquire the missing item, is meant to be analogous to the desire direction of fit, while the second response, changing his list to fit the situation he is in, is analogous to the belief direction of fit.

A mismatch between a belief and the world will occur when the proposition that we have a belief about does not have the truth value that we take it to have: for example, if a proposition does accurately reflect the world and we take it to be false, then there is a gap between how our belief has captured the world and how the world

---

reasonable, fully subscribe...Desires, emotions, passions...are the only possible causes of action." Russell 1999, p 170.

Anscombe 2000, p 56. The example I give here is slightly reconfigured from her version.
actually is. A mismatch between a desire and the world will occur when the object of the desire is not something which exists in the world: for example, if I desire that it not snow, but in fact it does snow, then how I want the world to be is not how it is. The distinction of direction of fit as I will employ it is essentially the claim that, if there is a mismatch between the world and a desire or a belief, having a desire is to be in a state which \((ceteris paribus)\) we attempt to fit the world to, which I will call having the desire direction of fit, whereas having a belief is to be in a state which \((ceteris paribus)\) we attempt to fit to the world, which I will call having the belief direction of fit. This fitting would be brought about by, in the case of belief, changing our position on the truth value of a proposition to match what we now take to be its actual truth value, so changing our belief; in the case of desire it would be brought about by attempting to change the world to bring about the state of affairs that we desire, which of course can only be done through action, so maintaining our desire. My task now is to establish that beliefs and desires do indeed have these respective directions of fit, and rule out the possibility of their having the opposing direction of fit.

### 3. Direction of Fit: Beliefs

Beliefs are states such that, when we are aware that there is a mismatch between the world and the belief, we attempt to modify the belief to better match up with the world. The main reason for this is that, as stated last chapter, it is constitutive of beliefs that they are attempts to accurately reflect the way the world is: “If a man recognizes that what he has been believing is false, he thereby abandons the belief he had.”\(^{42}\) Of course, beliefs do not always succeed in accurately capturing the state of the world, and many may never do so, but we attempt to have accurate beliefs so far as we are capable, because that is what beliefs are for. If a belief is inaccurate and we

\(^{42}\)Williams 1976, p 137.
continue to hold it, we generally must not be aware that it is inaccurate. If we become aware that a belief is failing to accurately reflect the state of the world, we feel a need to change it, assuming we are basically rational. For example, if we become aware that a proposition we were holding to be true is actually false, then we feel that it is necessary to change our stance towards it, and thus our belief. It can be that this need is somehow overridden, or that we try to purposely ignore it as best we can, but this is only to say that it is defeasible, not that it is not necessary. The belief direction of fit is required simply due to the nature of what beliefs are: states with which we attempt to capture the correct truth value of some proposition, as I stated in the previous chapter.

Railton argues that this feature is true of beliefs by definition, that this sensitivity is a constitutive part of what makes a state a belief, and suggests that mental states which are not sensitive to matching up with the world in this way simply are not beliefs. It is true that if somebody had a state she claimed was a belief, but accepted evidence that it was wrong and still did not change this mental state to accommodate this evidence, we would conclude that whatever was going on in her head was not a proper belief. We might conclude that this person is trying to have a genuine belief but is very irrational (or simply very stupid), but often we are most likely really drawing one of two conclusions, neither of which is truly ascribing a belief to such a person. Say we meet a man who apparently believes that aliens are living in his backyard, even though various pieces of evidence have been offered showing that this is not the case, and in fact he has stated that he accepts this evidence and allows that it’s true that aliens are

43 There are many issues to be raised here about how aware we can be that a belief is false - it seems that sometimes people hold beliefs in some kind of bad faith, some kind of awareness that the belief is false or at least unlikely, or even simply deliberately avoid evidence that might cause them to change their beliefs. These are beyond the scope of my work here. I take it that it is uncontroversial that generally beliefs are and should be what we genuinely accept as true. 44 For example, if you are very distracted and don’t have time to really process what you just learnt. 45 Railton 2003, p 297.
not living in his backyard – but he still believes that they are. Firstly, and most likely, we might conclude that this person only thinks that he understands the evidence against his belief, but in fact he really does not, due to whatever failures are going on in his theoretical reasoning. He’s simply not managing to truly comprehend how the evidence offered shows that his favored proposition is false, and so he is still having a belief, albeit a mistaken belief, because he’s not actually accepting evidence to the contrary, but only appears to be doing so. Alternatively, perhaps he does seem to understand the evidence and yet still holds his belief. We might conclude that he has some kind of mental compulsion to take the proposition that aliens are living in his backyard to be true, and so whatever evidence to the contrary is offered he will invariably come back to taking this proposition to be true. Such a compulsion could be thought of as not a real belief, or it might be considered a belief but only as the symptom of mental illness: i.e. as a sign of non-normal mental functioning. That a belief which is not sensitive to evidence would be considered a symptom of mental illness only demonstrates that beliefs are supposed to be sensitive to evidence – we think that the fact that somebody has one which is not indicates that something has gone wrong in their thought processes, and so we clearly think that beliefs are mental states which ought to be sensitive to evidence in order for our mental states to be functioning correctly. In any case, we only can see that beliefs are mental states that ought to be modified after the acceptance of evidence that it is not accurate, or in other words, are states with the belief direction of fit.

Williams suggests another argument that supports giving beliefs this direction of fit, this one based on the utility of correct beliefs. His argument is simply that if beliefs are not sensitive to the truth then they are of no use to us. If we accept that something we believe is false, for example, and yet try to believe it anyway, we are

46 Williams 1976, p 151.
forced to abandon our notions of evidence as that which counts towards or against what we should believe, since obviously we’ve just rejected them in regards to this belief. That means that we will be believing at will rather than following what the evidence indicates, and this leaves us with no way to determine what is true and what false, and thus no way to guide ourselves through the world. In short, this will leave us at risk of not being able to do or achieve anything, because how we know what to do, and how to do it, depends on our beliefs. For example, I can only keep living because I have various beliefs about how and where to get water to drink; if I simply decided to believe that water will find its way into my body somehow, I would have a difficult time surviving.

Beliefs’ sensitivity to evidence means that they fill an important practical role for us because having a reasonably accurate conception of what the world is like is necessary to move through the world at all, and to achieve the ends we want to achieve. If our beliefs don’t track the truth at all, then it will be impossible for them to be useful to us, and therefore, this is another reason why we attempt to change them when they don’t accurately portray the world. We could of course make-do if we kept only a few of our beliefs unresponsive to evidence, but we must make it a general rule of our belief forming that our beliefs track the truth: we can’t know which beliefs will be helpful to us ahead of time, and knowing which beliefs are useful to us will itself require a large network of true beliefs (and we can’t know ahead of time which beliefs we’ll need to be able to judge other beliefs). Clearly we’ll do better the more of our beliefs that are accurate, and so we should aim to always make our beliefs capture the truth as far as we’re able, even if we inevitably fall short of this goal some of the time.

Since these arguments show that beliefs must have the belief direction of fit, clearly they also, in addition to ruling out beliefs having no direction of fit, rule out beliefs only having the desire direction of fit and not the belief direction of fit. Railton’s argument applies here – if our concept of beliefs includes their being sensitive to
evidence, then they must generally change to fit the evidence, and not change the
evidence to fit them. It might seem that Williams’ argument does not actually rule out
the possibility of their only having the desire direction of fit, in the following way: if we
use our beliefs as a guide on how to change the world, then they will be action-guiding
even while inaccurate, and will come to be generally practically useful once we have
changed the world to fit them. This response will not actually help, however, as the
problem is that without proper truth-sensitive beliefs in the first place we will not be able
to change the world in order to match these “beliefs”, for obviously achieving our goals
with regularity will require knowing enough about the world to take the means
necessary to do that, and that we cannot do without proper beliefs. It is clear, then, that
beliefs have what is viewed as the typical “belief direction of fit”, which is to say, we
attempt to make the mental state accurately reflect the world, as opposed to the desire
direction of fit, which requires not changing the mental state to better match the world.

4. Sobel and Copp’s Objection

Sobel and Copp provide an objection against Smith’s direction of fit distinction,
by arguing that there are cases where it seems that beliefs do not change to fit the
world even when confronted with perceptions that the belief is wrong. They argue that
it is fairly common for us not to change a belief when faced with certain perceptions that
we don’t take to be accurate, for example, we don’t believe that a stick becomes bent
as it is submerged in water, even though it appears to be bent, because we don’t take
the perceptions of it being bent to be real evidence that counts against our belief. This
means that we only change our beliefs when confronted with perceptions that conflict
with that belief when we believe those perceptions to be accurate. If we must believe
the perceptions to be accurate before we allow them to count against our beliefs, then it

47 Sobel and Copp 2001, p 47.
looks like at least some beliefs can be concerned only with fitting other beliefs and not the world, as I claim when I talk about the belief direction of fit.\textsuperscript{48} Therefore the direction of fit distinction doesn’t explain the relationship between beliefs and evidence after all, and doesn’t capture the sensitivity of beliefs towards evidence, but is rather giving an account of beliefs working on other beliefs.

It’s difficult to give a full reply to Sobel and Copp without getting into the question of how we move from perceptions to beliefs, which I can’t do here, but I will say that their type of examples don’t seem to establish what they want them to establish. It is true that I don’t take the perception of the stick being bent as evidence, but that is because having previously examined such sticks I already have standing beliefs that such sticks are not really bent, which is to say, I have come to believe that these perceptions are misleading. There was a time when I was not aware of this optical illusion, and so the first time I examined such instances I must have formed my belief in reference to the perceptions I was having at the time, such as putting my hand in the water and feeling the unbent stick. We come to believe that certain perceptions are inaccurate by contrasting our perceptions and forming judgments about them and thus beliefs about them. Now when I judge that the stick is unbent I refer to such beliefs, but originally I had to judge from my perceptions in the usual way that any belief is formed, and the way I decide which perceptions to give weight to and which to ignore must be that I take some of them to accurately reflect the way the world is and others not, and the former is what I’m looking for. So it seems that the belief direction of fit standard must be present in order for me to sift through all my various perceptions to form beliefs. To judge which perceptions to discard and which to believe are accurate is to use the belief direction of fit, so these kinds of examples help establish the belief direction of fit as being present in our beliefs.

\textsuperscript{48} Ibid, p 46.
Sobel and Copp also appeal to examples of people who hold beliefs stubbornly even in the face of contradictory evidence, such as a stubborn assurance that God exists no matter the evidence, as examples of beliefs without the belief direction of fit.\footnote{Ibid, p 47.} Certainly some people fail to deliberate correctly and so their beliefs aren't as sensitive to evidence as they should be, but I don't think this is really an objection to the direction of fit distinction. I think we're likely to conclude that such people either aren't really accepting the evidence against their beliefs, or their beliefs aren't proper beliefs after all, just as I discussed above. The former is the more commonplace conclusion to draw about such people – we think they are ignoring evidence, not reflecting on it fully, and if they did so, then their beliefs would change to match the evidence. If that's not the case, then I think it is reasonable to conclude that these people don't really believe what they claim to believe after all, because, as previously argued, it's constitutive of beliefs that they be sensitive to truth. Stubbornly holding on to a proposition is not an example of a belief which does not have the belief direction of fit, but of having a mental state which fails to be a full belief.

5. Direction of Fit: Desires

Desires are states such that, when we are aware that there is a mismatch between the world and the desire, we attempt to change the world to better fit the desire, ceteris paribus. There are two components to this direction of fit, the first of which is that desires do not change to fit the world, and the second is that we attempt to change the world to more closely fit desires, if that is possible. The latter will be argued for as part of the argument for the HTM in section 7; it is the former that will concern me in this and the next section.
If something is the object of my desire, if I am attracted to some state of affairs, and this state of affairs does not exist, this need have no impact on my desire, meaning that I can still be and usually am still attracted to this state of affairs, as opposed to ceasing to be attracted to it. This difference between beliefs and desires must surely have something to do with the fact that while beliefs are concerned with truth values, desires are not, but are concerned with whether we are attracted to or repulsed by some object. Whereas to have a belief is to take some proposition to be true or false to a varying degree, to have a desire is to take no such position on the truth value of an object, but only to have a disposition to be attracted to it, which itself has nothing to do with any possible truth value of an object. We might of course be attracted to a possible truth value and take that as our object of desire. For example, we could specifically desire that a proposition be true, but that feeling of attraction still does not depend on the actual truth value of the proposition since the fact that the proposition was not actually true would not necessitate us ceasing to desire that it be true. I could desire that it be true that there is an afterlife, while still believing that there is not – clearly many people do exactly that. Desires are dispositions to be attracted to an object, and while it can be true or false that we have such a feeling and disposition to have the feeling, and it can be true or false that the object of our desire actually exists, clearly the feeling of attraction itself and the disposition to have it are not the kind of things that themselves admit of truth values. However we define a truth value and whatever we attribute as being necessary to having one, it should be obvious that feelings do not meet these requirements (a feeling of attraction includes no referent, no representation of the world, is not coherent, etc.), and thus dispositions to have feelings will not meet them either, since having a disposition only adds conditions on when we have this feeling. If I state that I am attracted to something, a reply like “that’s false” can only mean that it’s false that I feel that way – to claim that a feeling itself is true or
false is clearly nonsensical. There are other possible grounds of criticism against a desire (and feelings in general), for example, that a desire is imprudent or immoral, but not that it is false.\textsuperscript{50}

Because desires are not concerned with truth values, there is no immediate problem if the object of a desire is not something that obtains in the world, unlike with beliefs where there is clearly a problem if a proposition believed to be true is not a true proposition, because to believe is to attempt to capture accurately what’s true. The object of the desire is something that we have a feeling of attraction for: the object is picked out because we have this feeling towards it, not because we take it in some sense to be an accurate reflection of the state of the world as is the case with beliefs. This means that a contradiction between the world and the object of the desire does not mean that our desire is an unsuccessful desire, and so does not put us under pressure to change or abandon the desire simply because of this mismatch.

While it’s clear that desires do not have a constitutive problem with their objects not matching up with the world, it’s less clear that we might not have prudential reasons for wanting to make desires fit the world. In other words, as it was in our interest to make our beliefs fit the world as per Williams’ argument, it might also be in our interest to make our desires fit the world, so that we might be able to better satisfy our set of desires, but abandoning those which are unrealistic, for example. That might be the case, however it doesn’t really affect the truth of the HTM (since both instances appear to depend on desires motivating us to make some change in ourselves), so I will leave that possibility aside here.

\textsuperscript{50} On some accounts of desire it might be possible to assign a truth value to a desire, if desires include judgments, for example, but not, as I’ve explained, on my account.
6. The Realist View of Desirability

Desires have the belief direction of fit if we are realists about desirability; a recent example of such a realist about desirability is Parfit, in his *On what Matters*.\(^51\) Parfit’s realist view about desirability is fairly collapsed into his realism about reasons, but his position is that certain objects in the world really are such that we have reason to desire them, and as such certain things in the world really are desirable (and give rise to reasons to act), and that it is only appropriate to desire those things which actually are desirable. He also holds this for reasons and goodness. If desirability is a real quality, then it seems like we must conclude that desires should fit the world in the sense that they must be formed in response to the correct objects, namely those which truly are desirable. This is analogous to belief formation, and indeed Parfit takes it to be so; as we must be sensitive to facts about the world in order to determine what is really true and form correct beliefs, so too must we be sensitive to facts about the world in order to form correct desires.

Parfit further argues that the reasons we have to desire anything will be “irreducibly normative truths”,\(^52\) and notes that those who believe that reasons depend on psychological facts, as I do since I believe that what we desire depends on our character and other psychological facts about ourselves, are generally skeptical about the existence of such truths. I am indeed skeptical about the existence of such truths, as I’m not sure what they are supposed to be or how we’re supposed to know that they really do exist. I will now discuss those two concerns, although only briefly since the existence of such truths is a very large metaphysical question that I cannot fully cover.

---

\(^51\) Parfit 2011, Vol. 1, p 50 and *passim*.
\(^52\) Ibid, p 109. Parfit also claims this is true for reasons to believe and reasons to act. He contrasts his position to naturalism and calls it non-naturalism.
here. I hope nonetheless to provide the outline of an argument against Parfit’s claims here.

I have trouble envisaging what irreducible, non-natural, normative truths would be, beyond a declaration by fiat that something just is right or just is wrong, with no further explanation required or possible. The question of what these truths are supposed to be is tied up with the question of how we know they exist, for surely the reason we think natural properties really do exist is because of our various experiences with the natural world, and the successes of science, and so on. It’s important to determine how we supposedly know that such normative truths exist then, and this in turn does not have a very satisfying answer since it appears that how we are supposed to know that these truths exist is that they are self-evident or via our intuition (which amounts to much the same thing), specifically, that our intuitions inform us that we have reason to want certain things and not other things. Parfit’s agony argument, for example, is the claim that certain theories which reject his claims about these truths are committed to saying that we might have no reason to avoid being in agony next Tuesday if we don’t desire to avoid such pain; according to Parfit this is very implausible, and this implausibility is what he rests his rejection of such claims on. Perhaps he is right that it is extremely implausible to think that we might have no such reasons (I am not actually so sure), but my concern is that his argument relies at its base on the appeal to implausibility and therefore on our intuitions about what is desirable. The question of whether normative facts like desirability actually exist is, it seems to me, particularly poorly suited for appeals to intuition.

We see many examples of a disconnect between our intuitions and the truth when it comes to natural facts, for example, in the famous slit experiment. This

---

53 Ibid, p 73.
experiment takes a plate with two narrow slits in it, places some photosensitive material a distance behind the plate, and fires a light source at this plate through the slits. The light travels through both slits, is somewhat scattered due to the narrow openings, and some of it rejoins on the other side. Because light can be thought of as being comprised of waves, when two waves pass through the two different slits, it sometimes ends up that peak of one wave and the trough of another wave meet on the other side, causing the two waves to cancel each other out. In other places, a peak meets a peak, or a trough meets a trough, strengthening the resulting light, and so on the photosensitive material a pattern of light and dark banding emerges in line with the strengthened and cancelled out waves. This experiment has also been carried out by firing one wave of light at a time at the plate; it seems obvious that in this case what we’d expect to result is a fairly even distribution of light on the photosensitive material, since there will be no interference from other light waves, as there is only one light wave passing through at a time. In fact, what actually happens is that the exact same pattern of banding emerges as before, and one of the conclusions scientists have drawn from this is that the single wave of light, which can also equally be considered a single particle, is in some sense passing through both slits at once. That a single particle could be in two places at once and interfere with itself is surely an extremely unintuitive claim, and yet that is precisely what seems to happen.\footnote{There are numerous other examples that could be used, of course, such as the Earth is not really stationary, that heavier things don’t fall faster than lighter things, and so on.} If our intuitions about real natural facts can fail so badly to capture the truth, there’s little reason to think that our intuitions about supposed real non-natural facts would fare any better.

The obvious objection to a case like this is that natural facts are not like normative facts, and that simply because our intuitions don’t work very well in determining scientific truths does not mean they will not successfully pick out normative
truths. However, I don’t feel that there is much evidence that our intuitions are very good at this latter either, particular not when we consider how specific they need to be.

We see a large disparity of normative intuitions between different times and places, and not at all the kind of similarities that I would expect to see if our intuitions tended to reliably pick out the same kinds of things as being desirable. Consider, for example, that for a thousand years many people in China thought it was desirable for parents to break almost every bone in their young daughter’s foot and bind it into a compressed form, crippling her for life, in order to meet their society’s standards of beauty. To me it is extremely self-evident that this is not a desirable thing for parents to do, but clearly they did not agree. The problem is that if desirability is a real quality that we’re supposed to be able to pick out via standard moral judgments, as is implied by Parfit’s arguments, then this implies that we should expect to be able to reach agreement about many quite specific things as being desirable, and this is not generally what we find happening. In contrast, a theory which requires agreement on a smaller number of general features of human nature, like Utilitarianism, for example, which can get started if we only agree that happiness is good, is on much stronger ground when appealing to intuitions. It’s more plausible to think that such agreement as there is in our intuitions is due to basic facts about human nature, such that we all (or virtually all) find happiness desirable, and not due to some non-natural property inherent in external objects.

I am not positively establishing here that these non-natural normative facts do not exist; I am, as I stated, skeptical of their existence, not convinced that they don’t exist. I don’t know how to establish their existence one way or the other, particularly not in an appropriately brief fashion. It is not a crippling objection to their existence that different people have different views on what is right and wrong, for of course different people have different views on what is true and what is false, and yet the vast majority of people still believe that there really are natural facts and that some of those different
beliefs are simply wrong. So the fact that different people have wildly different normative intuitions does not necessarily allow us to conclude with certainty that there is no truth of the matter as to what is desirable. Differences in opinion over what natural facts there are can be explained by appealing to failings on the part of the person, who may fail to properly recognize natural truths by being irrational or by being led astray by other mistaken views, and so on, and the same might be said for differences of opinion in regard to normative truths. The crucial difference is that we have developed further standards of judgment that we can appeal to in order to resolve differences of opinions about natural facts, and this is what allows us to conclude that some intuitions are simply wrong, whereas Parfit does not have further standards in regards to non-natural facts. As the example of the slit experiment shows, natural truths often do not line up with what we find plausible, and how we know which natural facts there are has little to do with our intuition, but rather on the predictive success of our theories and the fact that we can derive practical results from them. We do not need to take the oddities of quantum mechanics on faith, for example, as we can see that many of the things stated by the discipline must be true because much of the technology we use in our lives depend on facts determined by people working in that area.\textsuperscript{55} We can’t resolve differences in intuitions in regards to natural facts by appealing to more intuitions because clearly that can’t help without some explanation as to why some intuitions are better than others, which is to say, without reference to some further standard of judgment, and so likewise we cannot merely rely on intuitions with normative facts either. Since Parfit’s argument depends on intuitions about what is self-evident and what is not, notably to establish the claim that desirability is a real property, I do not find it a convincing argument that such properties exist and therefore, in turn, that desires should be based on them and thus have the belief direction of fit.

\textsuperscript{55} Transistors and lasers, for example, are based on the theories of quantum mechanics.
7. Direction of Fit Applied to the HTM

The point in establishing that desires have the desire direction of fit and that beliefs have the belief direction of fit is that it is because of these features that desires, and not beliefs, are suitable to provide motivation. I will now argue for the HTM, establishing that the desire direction of fit is necessary for a mental state to be motivating, and therefore that desires can be motivating, and beliefs are not. The argument for the HTM will not truly be complete until the next chapter, when the claim that some judgments have the desire direction of fit will be addressed, thus ruling out beliefs from motivating.\(^{56}\)

Smith argues that desires must be solely responsible for motivation because they have the same direction of fit as the state of having a goal, in fact he thinks that "having a goal just is desiring",\(^ {57}\) and that motivation is the pursuit of a goal.\(^ {58}\) This view makes sense for Smith because he defines desires in terms of their functional role in grounding various dispositions to act and so on;\(^ {59}\) however clearly my concept of desiring, involving as it does a feeling of attraction or repulsion, means that I cannot very simply conclude that having a goal just is desiring, for I do not think it is very plausible to claim that having a goal necessarily involves having such a feeling as I have attributed to desires. For example, it’s possible that I might have a goal to get to work on time and yet have no particular feelings towards that goal whatsoever (my feeling of attraction being towards getting money, etc.). Or we might desire something which we don’t take as a goal: that the weather be pleasant tomorrow, for example, is something that we could desire and yet not take as a goal (since there’s nothing we can do about the weather). I don’t intend to discuss what a goal is here, although I do think

\(^{56}\) As Coleman 2008, p 128, notes.
\(^{58}\) Ibid, p 44.
\(^{59}\) Ibid, p 52.
that goals are provided by our desires, which won’t be surprising given that I am about
to argue that desires are necessary for motivation, and it’s pretty obvious that having a
goal is closely related to having a motivation. It seems likely that we cannot have a
goal without being motivated, although we may be motivated without having a goal,
depending on how broadly or narrowly we define a goal. Even though I cannot accept
all the details of Smith’s argument, I do agree with the general thrust of his argument,
and so this is how I will argue for the truth of the HTM, specifically that desires and not
beliefs are sources of motivation because being motivated requires a state with the
desire direction of fit, which I have previously argued desires have and beliefs do not.

As previously discussed, to have a desire is to have a mental state to which we
naturally attempt to make the world fit, that is, we try to match the world to the object of
our desire, ceteris paribus,\(^{60}\) whereas beliefs are states that we tend to change to better
match the world. The point of essentially all intentional action is to produce some
change in the world, which is our intent, and acting will mean having some kind of
object that we are trying to bring about. For example, if I study with the intention of
passing my exams I have the state of affairs where I have passed these exams as my
object. Alternatively the object we are trying to bring about may be the action itself – I
may play a game because I like it, for no further reason than that, so my object in
playing the game is only to play the game (and have fun in so doing), but my playing
the game is still something that I have intentionally brought about in the world. Notice
how this means that intentional action has the same direction of fit as desires, in that we
usually try to bring about the object of our desire (assuming that is possible), making
the world fit the desire, just as we try to bring about the intentional object of our action
and fit the world to our intention. We try to bring about what we desire, because having

\(^{60}\) By which I intend to leave room for cases when we have no idea how to do this, or if we are
incapable to trying to do this, and so on.
a desire is having the kind of state that makes us try to change the external world to
better match our desire. This is because of the feeling of attraction that desires include,
because such a feeling is just the kind of thing that expresses itself as our trying to get
what we are attracted to, unless some further factor overrides – when we judge
whether somebody is attracted to something we consider if she ever tries to acquire it
or bring it about in appropriate circumstances, and if she doesn’t, we conclude that she
isn’t attracted to that object after all. Of course, being attracted is also a
phenomenological matter and not merely a question of behavior, so our judgments
about somebody else may be wrong, but clearly there is some immediate connection
between being attracted to something and trying to get it that we expect to be and find
to be necessary to feelings of attraction. So being attracted to something results in our
trying to bring about the thing we are attracted to when that is possible, and when
nothing else overrides that attraction.

In contrast beliefs are not like that, in that if the proposition we believe to be
ture turns out to be false, we modify the belief rather than the world. We certainly must
modify our beliefs to line up with the world because otherwise they will not track what is
ture and false, and that is the purpose of beliefs, as previously discussed. Therefore,
there seem to be only two possible situations that we can find ourselves in with our
beliefs and their matching the world: either the belief we have is correct, in which case
there is no need for us to do anything as our belief is a successful one and there is no
gap between world and belief, or our belief is incorrect, in which case we modify the
belief. There are only these two cases because, fundamentally, the only purpose of
beliefs is to accurately capture some truth value, and they do that either correctly or

---

61 Anscombe 2000, p 68.
62 Right now this argument is vulnerable to objections about beliefs of the type that judgment
internalists would propose, which themselves call for action. I will turn to their position in the
next chapter.
incorrectly. We can have varying degrees of certainty in what we believe, but that too can be evaluated, as either we have the right amount of certainty given the evidence available, or we don’t. In neither one of these cases is there a move to bring about some change in the world, because to do so would be to undermine the sensitivity to the state of the world that beliefs must have to be beliefs – beliefs are supposed to accurately reflect the world, and to do that they must change to fit it. In that case, how could beliefs give rise to an intention or some other mental state which had the aim of modifying the world in some way? The essence of beliefs is to correctly capture the nature of the world, not modify it. Say beliefs did call for us to modify the world somehow, the modification would have to be of one of two possible things: it could call for modification of the aspect of the world captured by the proposition which is believed, in which case the belief is self-defeating since it would make itself no longer accurate (either no longer actually accurate, or no longer apparently accurate if in fact the belief is incorrect to start off with), as a proposition believed to be true, for example, would only be true of the old, unmodified world. The other alternative is that the modification would be of an aspect of the world not captured by the proposition which is believed, in which case, it’s extremely difficult to see how the belief could say anything at all about those states of affairs, for the belief references only the propositions it takes a stance on, and so really has nothing to do with anything outside of those propositions. It seems clear that a belief, or any other mental state, could not lead us to do anything about a state of affairs which it has no connection to.

Motivation is the mental state which is the direct explanatory mental cause of action. States with just the belief direction of fit could not explain action, as those are not states that it makes sense to act on, as discussed. In contrast, we make sense of somebody feeling attracted to something by thinking that she will try to get what she is attracted to, that she will try to bring about the object of her desire. Motivated actions
themselves are attempts to bring about some object. Obviously, then, desires can make sense of actions as being done by an agent because why somebody acted is answered by noting that she is attracted to what the action achieves, and so desires can act as motivations for our actions. It’s worth noting that it still remains to be explained why it is that desires do actually motivate. I take this to be a contingent fact of our psychological makeup, that feelings of attraction typically lead to action, when action is possible or appropriate. Whenever we are attracted to something, we will generally try to acquire it; this is simply what it means for us to be attracted to something. It need not necessarily be this way – it is possible that some other creature might react to feelings of attraction in a completely different way, say by falling asleep – but this is how we, as a matter of fact, are constituted.

8. Changing Our Beliefs

There is an obvious objection to this claim that beliefs cannot motivate, while still accepting that they only have the one direction of fit, and that is the case of our changing our beliefs in order to more accurately reflect the world. For example, when we come to learn that a proposition we held to be true is in fact false, we will (most of the time) change our position on that proposition accordingly. It might seem like this is a case of belief causing some action, specifically the action of changing our belief, and therefore, presumably, some motivation must have occurred for that action to happen. If that is the case, then I need to establish either that it is a desire which is doing the motivating, or that this is not a motivated action after all. My response is that it is a bit of both, depending on whether it is a case of our changing our belief more or less involuntarily when faced with evidence, or when we change our belief after deliberation.

63 I am not, therefore, using "explanation" in a purely causal sense here. Beliefs, or indeed physical brain states or the natural laws of the world, can explain certain events in a causal sense, but not in the sense of making it clear why somebody did something, or that it was done intentionally.
A belief might change because of some other belief we have come to hold, when we see that this new belief forces us to revise a previously held belief, or it might change due to our perceptions of the external world changing. In the latter case, where beliefs are changed by something in the outside world, there are two features to notice when determining what motivates such a change. Firstly, the belief which is being changed here seems to be in the wrong position to be what is motivating us, since the belief is the state that is being changed, not the state that is causing the change. The belief is what we are working on, and therefore it seems like there must be something further that precipitated this work, because if it was just the belief self-causing its own change why would this change be happening now and not at any other time, since the belief would have been present before this action occurred? Presumably our changing our belief occurs now because something else arose and caused it to happen now. Therefore, in such cases that belief is definitely not acting as motivation for that change, and whatever is cannot, since we are considering the second kind of cases here, be another belief, but must be something else.

So secondly we have to consider what could possibly act as a motivation in these cases. I offered as a definition of motivation a mental state that makes sense of an action as being done by some agent, however, these beliefs are being changed because of the way the world is. The state of the world is not a mental state, and therefore no state of the world can be a motivation, and so if changes in beliefs are directly caused by the external world, then this would not be a motivated action at all. However, changes in beliefs are not quite directly caused by the external world, as that would mean that the world would be reflected directly in our beliefs, and that clearly is not entirely the case, as if nothing else, we must at least have had experience of states of affairs in order to have beliefs relating to them. So at the very least whatever mental states make up perception must be between the belief and the world – might these
states, then, count as motivation? I am hesitant to view them as motivations as I’m not sure that such states would make sense of our changes in belief as being done by an agent or as being intentional. It does not seem like, when I am confronted with a pen on a desk and have a perception of such, I have any kind of intention or goal in forming the belief that right now there is a pen on this desk, I simply do come to believe that without any motivational input from me. It seems like very young children manage to form such beliefs, even if they lack other mental abilities that we might think are necessary for being an agent (impulse control, for example). Directly forming and changing beliefs in response to perceptions of the world is not a motivated action, but a reaction to our perception, and in any case even if it were motivated, certainly beliefs would not be doing the motivating. The situation can be no doubt different with more complicated beliefs which aren’t directly formed from just perceptions, which can require more active reasoning on my part.

In the case of changes in beliefs which are due to prior beliefs, there are two different ways this happens: either changes are brought about reflexively by our automatically altering our beliefs to line up with new beliefs which are caused by changes in our perception (say second-order beliefs related to the first-order changed beliefs), or we deliberate over what to believe. The former is obviously not motivated action, as above. The latter is a different matter, and does seem to be motivated. Williams’ utility argument for the truth sensitivity of beliefs which I discussed earlier suggests a reply for the HTM for these kinds of cases, namely that people have a desire to have accurate beliefs and that this desire is what motivates our changing of beliefs in such circumstances. Properly discussing how we deliberate over our theoretical reasons is too large a topic to discuss here, but certainly a desire to have accurate beliefs must play an important part when we weigh the evidence and try to consider what we ought to believe. We are trying to produce the best, most truthful
beliefs we can have, because doing so will help us pursue our ends, or simply because we have a desire to know the true nature of things as much as possible. It’s hard to imagine that somebody who didn’t have a desire to have accurate beliefs would spend much time deliberating over what to believe, and our experience of such people is that they indeed don’t appear spend much time deliberating. People who don’t care too much whether their beliefs are accurate more or less form beliefs in either the reflexive way or perhaps by relying on authority or so on, but they tend not to formally deliberate and thus not be motivated to do much with their beliefs. The desire to form accurate beliefs is what motivates us to change our beliefs when we are deliberating over whether or not to change them.

9. Mele’s Objection

Mele argues that Smith’s argument for the HTM based on the direction of fit of desires fails because we are not motivated by our desire for the world to be a particular way in certain cases: e.g. when we desire something to be different about the past. There’s nothing we can do about the past, and we cannot attempt to make the world fit whatever object we desire about the past, so presumably we aren’t motivated to do anything about the past no matter what our desires concerning it might be. In reply we should note that firstly, the desire is still not revised because it doesn’t match the world, so it still has the direction of fit such that we do not attempt to fit it to the world, e.g. my desire that some historical figure would have been kinder is not revised upon learning that he was not kind at all. Secondly, there is an obvious difference between not doing something because we can’t do it and not doing something because we aren’t motivated to do it. It’s true that we can’t change what has happened in the past, and so we don’t attempt to do so as we wouldn’t even know how to attempt to do so, but this

---

64 1995, p 394.
only means that the usual response to such a desire is blocked, not that we are having a different response to the desire. If, somehow, we gained a method of affecting the past, then people who desire to change something about it might well do so, provided some other concern didn’t override that motivation, just like with any other motivation. The old question of whether or not you would kill a young Hitler if you had a time machine, for example, indicates that it is possible for us to have motivations towards changing the past, were it possible. This is a similar case to someone who is paralyzed desiring to walk: the fact that she cannot do anything to work towards that action does not mean she is not motivated to do it. That circumstances prevent us from attempting to fulfill a desire does not mean our reaction towards it has changed, that we are no longer motivated by it, only that we have no avenue in which to pursue that motivation.

Most importantly, however, I would also respond that the HTM need not claim that all desires are motivating. The account I gave of motivation earlier was essentially a functional one: whichever desires can explain action, are motivations. Desires to change the past may well not be involved in any action, and therefore are not motivations (but they could become motivations should we gain access to a time machine, for example). The HTM typically claims that all motivations are desires, but the reverse does not necessarily follow. It is no objection to my version of Smith’s argument, then, that not all desires motivate.

10. Nagel’s Objection

Nagel’s influential argument against the HTM in *The Possibility of Altruism* essentially attempts to show that even if desires must be present in all cases of motivation, they need not ultimately be doing the motivating - in other words, it can provide a way for someone to accept a version of Smith’s argument and yet still deny
the HTM. I will argue that his objection fails, and so Nagel fails to refute the HTM. Nagel's position is that desires can be "motivated" by beliefs, that is, arise purely as a reaction to a belief, therefore it is the belief that is the truly motivating state as it gives rise to both desire and action. The desire must be present when motivation is present, but the desire does not actually produce the motivation, it is simply a necessary link between belief and action. In some ways his position is similar to a reversed HTM, in that where the HTM holds that while beliefs are necessary for action, desires are the state that motivates, he argues that desires are necessary for action, but beliefs are the state that actually motivates. I will attempt to show that he is mistaken in his central case of future desires, where he holds that beliefs are motivating us, and argue that it is in fact desires that are motivating us in such cases. Other elements of his position will be taken up in chapter 4.

Nagel's strongest argument against the claim that only desires motivate is that we can have motivations relating to future desires, which he argues is impossible for the HTM to account for. For example, if I know that the food in my pantry will soon run out, we would expect that I would be motivated to go and procure more food. Doing so will undoubtedly lead to the future satisfaction of my future desires to sate my hunger, but it is no small detail that these desires are in the future, because it is not clear immediately how future desires can motivate us. No matter what the argument for the HTM, one necessary condition on motivation is that the motivating state must actually be present, must exist, in order for us to be motivated, because obviously a state must be present to cause or explain anything. What Nagel points out is that it seems like my future desires cannot be motivating me now, because right now I don't

---

65 This is how Shafer-Landau uses it in 2003, p 137-8.
66 1978, chapter V. All references in this section are from The Possibility of Altruism, unless stated otherwise.
67 p 27. Actually if we view motivation only as what explains, rationalizes, our action, rather than what causes it, perhaps this premise would not be quite so secure as it appears to be.
have these desires – I will have them in future, but I am acting now when I don't have them, and a desire that I don't have cannot motivate me. So when I go grocery shopping I am not hungry, and I can't have a desire to sate my hunger when I am not hungry; such a desire cannot, therefore, be what motivates me to go grocery shopping right now. The problem for the HTM is that we do indeed go to the grocery store to get food for the future, even when we are not hungry, even when we are sick or have just gorged ourselves and so feel no attraction to the thought of eating at all, and so it cannot be a desire to satisfy our hunger that motivates us in these circumstances. Nagel argues that what motivates me in these cases is my belief that in the future I will have these desires which I can satisfy by acting now, so in the above case, the state that motivates me is my belief that in the future I will be hungry. I am being motivated by a belief, not a desire, and even if this belief produces in me an instrumental desire to go to the grocery store, that desire is derivative and not what is really motivating, but is rather the state of being motivated. 68 Because I believe that in the future I will be hungry, and because I believe that going to the grocery store will satisfy this desire, I gain a desire to go to the grocery store. It has to be my belief about what I will desire in the future that motivates, Nagel thinks, because there is no desire that is currently positioned correctly to be my motivation.

One obvious response to Nagel’s argument is that we have a standing desire to do what is necessary now to fulfill future desires that we know we will have, which Nagel thinks is plausible. 69 This desire could explain my going to the grocery store in terms of the HTM, as I would be motivated to go to the grocery store by my desire to do what I need to now to satisfy the desire I will have in the future to deal with my hunger. There may be, Nagel allows, such a desire to fulfill my future desires in general.

68 p 29.
69 p 28.
including to eat in the future, which is present when I go to get food, but he argues that the only way we attribute such a desire to people is by observing that they are indeed motivated to do so, and therefore “That I have the appropriate desire simply follows from the fact that these considerations motivate me”\(^{70}\) and “it is not necessary either as a contributing influence, or as a causal condition.”\(^{71}\)

There’s no reason that the Humean needs to accept such a conclusion however. It’s true that our beliefs that we will have certain desires in the future, and that doing certain things now will satisfy these desires, will play a role in our action, but we need only view this in terms of the usual one of guidance accorded to beliefs by the HTM, if the Humean is granted that we have a standing desire to satisfy our future desires. If I desire to own a car, my belief that going to a car dealership is the thing to do will combine with the desire and lead me to go to a dealership. Likewise, since I desire to satisfy my future desires, my beliefs that I will have a desire to eat in the future, and that going to the grocery store will take care of this desire, will direct me to go to the grocery store. These are, I think, perfectly plausible accounts of motivation, and ones which accord with the HTM. Humeans can even allow that the desire specifically to go to the grocery store is derivative and produced by the belief that we will have such future desires and that this is how to satisfy them, because ultimately it is the desire to satisfy future desires which is doing the motivating. We can therefore recast the activity of going to the grocery store not directly as a way of satisfying a future desire to eat, but rather as a way of satisfying a present desire to do what is necessary to satisfy my future desires, and can allow that beliefs give rise to the more specific desire and motivation to go to the grocery store. So it’s true that my beliefs are important in determining which actions my desires will result in, but then the HTM

\(^{70}\) p 29.  
\(^{71}\) p 30.
already accepts that, and we can therefore see that the HTM has no problem explaining such Nagelian cases of future desires. Certainly it is true that how we attribute this standing desire to satisfy future desires to anyone is a result of seeing that they are motivated to act accordingly, but that alone simply doesn’t establish that the desire is not what is really motivating. How we attribute the existence of any mental states at all to someone is by observing his external actions (leaving aside brain scans and the like, which are obviously not our usual method), so there’s no particular inference about the causal status of mental states that we’re entitled to draw from that method. We may only attribute such a desire to someone derivatively from looking at his actions, but that doesn’t mean that the desire only exists derivatively, anymore than we would be entitled to conclude that beliefs only exist derivatively of actions.

Additionally I don’t agree with Nagel’s assessment of when my more specific desires are present, that is, not merely the general desire to satisfy our future desires, but also desires such as to satisfy our hunger. Nagel suggests that the desire to eat cannot be motivating because it is not present right now, because while I will be hungry in the future I am not hungry right now. However, my claim is that while I might not have a feeling of hunger to deal with right now, the desire to satisfy my hunger generally is present now, even if the feeling of hunger (and the direct desire for food now) is not: i.e. we might not want to eat right now, but we may, and usually do, want take care of our future hunger (e.g. by going shopping). A desire is a disposition to feel attracted to or repelled from an object when the object is in front of us or when we are thinking about it or remembering it. Since it is a disposition, the desire still exists even when the object is not present, just as a glass is still fragile even when not broken. It’s true that we’re not actually hungry right at that moment, but if we think of being hungry in the future we could be (ceteris paribus) attracted to the idea of dealing with that hunger, which means that at that moment the representation of the object of this desire
is in mind, and this is enough to manifest the desire as a feeling of attraction towards not being hungry – but the hunger is in the future, instead of in the present. Since this desire is present it can obviously motivate us. This desire can be present even if we don’t feel attraction to food right now, for example if we just recently gorged ourselves and can’t stand the thought of eating any more, even days from now. I can have a desire to sate my hunger in the future, and thus in a sense feel attracted to the idea of eating, even if right now circumstances are such that I don’t feel attracted to the prospect of eating at the moment.

Nagel has further objections relating to this class of examples which I will deal with in the discussion of my theory of reasons, since they concern reasons more specifically than motivation.

11. Conclusion

In this chapter I have argued that beliefs have the direction of fit such that when a belief and the world mismatch, we generally try to change the belief, whereas desires have the direction of fit such that when a desire and the world mismatch, we generally try to change the world. These directions are fundamentally the result of how beliefs aim at truth and how desires include a feeling of attraction for their object. Because desires have this direction of fit, they are suitable for motivating our actions because they can explain why we would act to change the world, but beliefs are not, because they must change to match the world. The next chapter discusses the position of judgment internalism, the view that judgments are what motivates, a significant opponent to the HTM.
Chapter 3: Judgment Internalism

1. Introduction

In the previous chapter I argued for the HTM; the argument given depended on desires having a certain direction of fit and beliefs not having this direction. I’ve established that beliefs must have the belief direction of fit and not the desire direction of fit, but what still remains to be ruled out is the claim that a certain kind of beliefs must necessarily involve or be conjoined with some state with the desire direction of fit. That claim is a popular one (although it is not always couched in these terms) and an important one. The typical sort of position to take here is that of judgment internalism (JI). The judgment internalist holds that the judgment that a consideration is a reason is what motivates us, in other words my belief that a consideration is a reason necessarily leads to motivation. That is, this theory holds that we form a belief that we have X reason to do Y, and thus (somehow) become motivated to do Y. There seem to be two possibilities as to why that would happen: firstly, that the reason belief has both the regular belief direction of fit (we have a reason to do Y) and the desire direction (so we do Y); secondly, that a belief that I have a reason to do Y will be followed necessarily by a desire to do Y. The first possibility will be relatively easy to reject, but the second, more plausible and usual interpretation, will require more work to be defeated. It’s particularly common to hold that JI is true at least for moral judgments, that is, the claim that a judgment that I have a moral reason to do something will motivate me to do it. Often JI also conflicts with the theory I will argue for in the next chapter, that of reasons internalism, which holds that only considerations that motivate us can be reasons. Although JI isn’t actually a position on what constitutes reasons for action at all, just on how our views of them affect us, it has the natural tendency to reverse the relationship between reason and motivation as compared to what reasons internalism holds it to be:
that is, rather than our taking a consideration to be a reason because it motivates,\textsuperscript{72} JI holds that a consideration motivates because we take it to be a reason. To a certain extent, then, both the argument for the HTM and the arguments for reason internalism will be arguments against at least some forms of JI. I will provide arguments more directly against it in this chapter from the perspective of the HTM.

There is another kind of theory which is in some ways quite similar to JI, which I won’t be arguing against directly here, which holds that we must choose which motivations to act upon, taking those to be our reasons (and perhaps they therefore become our reasons due to our so choosing them). I have in mind the kinds of theories put forth by Korsgaard and Velleman, where our choosing to do a certain action is in some way constitutive of our being rational agents. These theories are subtle and interesting: however, depending on the way they are configured they may still accept the HTM and reasons internalism, or at the very least it’s not always clear what view they hold on the direction of fit of beliefs, and so it may not be helpful to think of them in those terms.\textsuperscript{73} Since to argue against these theories would therefore require separate arguments, and since in any case any points of disagreement between us would not be located in the terms of this current debate, I will leave aside these theories here.

2. Why Judgment Internalism?

It is perhaps accurate to say that JI has been and continues to be a popular theory, at least in moral psychology. I think there are two main reasons for this: firstly, it is simply – intuitively – felt to accurately represent how motivation works for us, since we do usually seem to be motivated to do what we judge we should do. Secondly, I

\textsuperscript{72} Being motivating is a necessary condition for a consideration being a reason on internalist theories, but it may not necessarily be a sufficient condition.

\textsuperscript{73} Korsgaard, for example, says that reasons internalism is trivially true on her theory; 2002, p 64.
think, is concern over the possible implications of the HTM, particularly what it may mean for morality. In other words, it is often an attempt to escape the very problem that I am attempting to deal with in this dissertation: that the HTM tends to imply moral theories which undermine objectivity and realism about morality. Many philosophers have felt that reasons must be connected with motivation and that if we accept the HTM and conclude that only desires can motivate us, then it seems that we are left with the conclusion that only desires can provide us with reasons. This means that unless we desire to do what is moral, we have no reason to be moral – but we don’t usually think that morality is something that can be escaped by an appeal to desire. If we can be motivated by our judgments, this means that we are motivated by our beliefs on what is moral, for our reasons judgments are based on or are expressions of what we believe to be the case. In that case, our desires do not determine whether or not we have a reason to be moral, but fundamentally it will be our beliefs which determine (or perhaps, identify) our reasons. Beliefs, as stated last chapter, aim at truth, and if our moral actions can be motivated by beliefs then this may imply that our moral actions can be based on facts, specifically whatever facts our moral beliefs and the judgments based on them are picking out. Therefore the position that beliefs can motivate tends to be friendly to the theory that there is some objective, universal truth about morality, that morality and what reasons there are in general is a question of there being some kind of relevant facts, whereas the HTM appears to more naturally tend towards (but does not entail) forms of subjectivism, or relativism, etc. My response to this concern will not come in this chapter, but in chapter 6.

This approach also has the appeal of making our practical reasoning analogous to our theoretical reasoning. When considering the formation of beliefs, the natural way

---

74 For example, Brink 1989, p 45-6, and Goldman 2009, p 144. Both identify this concern as the primary motivation for accepting some form of JI.
to think about how we form them is that we consider various pieces of evidence, perceptions, other beliefs, etc., and judge what belief they give us a reason for holding. In other words, we form a judgment that we have reason to believe a certain thing, and then we believe that thing. Analogously, our practical reason could proceed by our considering various pieces of evidence, deciding upon this evidence what reason they constitute for action, and then acting accordingly. Whether or not this is the correct picture of theoretical reasoning I will leave aside, and I won’t argue directly against this picture of practical reasoning either, as my concern is rather with the thought underlying it, namely JI. I will note though, that simply being analogous to theoretical reasoning does not, in itself, make this an argument against the HTM, as there is initially no particular reason to think that our practical reason must be analogous to our theoretical reason – since they have different objects from each other the mechanisms may well be different too. Some further argument for why we should think that these two types of reasoning are analogous would have to be provided, and that would essentially, I take it, be an argument for JI anyway.

3. Limitations on Any State Having Both Directions of Fit

When discussing the first possibility, that a belief may have two different directions of fit, we can easily dismiss this possibility, because, as Smith points out, only a certain combination of directions of fit is even conceptually possible. The first limitation we can place on such combinations is that no single, simple mental state can have both directions of fit in regards to a single thing. Let’s take some possible simple mental state, which for argument’s sake is similar in structure to a belief, in that it is a proposition attitude, and call it a “besire”. If this state had both directions of fit towards the same proposition, then we would attempt to make the attitude fit the proposition and

---

76 Smith 1995, p 118, from Altham.
also make the proposition fit the attitude, and this, as Smith says, "is just plain incoherent". The desire would pull us in two conflicting directions at once, as we would simultaneously be trying to change our desire to fit the world, while changing the world to fit our desire. If it's possible to try to change the world to fit my desire, then clearly the desire cannot accurately represent the world, otherwise there would be no need to fit the world to it. If we were to change the desire to fit the world as per the belief direction of fit, then acting to make the world fit the desire as per the desire direction of fit would be impossible, since there would be no gap between the desire and the world, and therefore there would be no action to be done. Having both directions of fit in the same mental state towards the same object is thus self-defeating. Either one or the other would have to be pursued – either we try to change our mental state, or we try to change the world. We couldn’t do both at once because there would be no reference to aim our change at – we couldn’t be changing the world to fit our desire if we were changing our desire at the same time, and vice-versa.

If no single, simple mental state can have both directions of fit towards the same thing, this still leaves open the possibility of a mental state that is comprised of further constituent states which each have their own direction of fit, which are necessarily conjoined into a large state for some reason. There are two different ways this could happen: either beliefs and desires have their own constituent parts, at least some of which have contrasting directions of fit, and thus it’s possible for a belief and/or desire to have both; or we could argue that there is some complex state that contains both beliefs and desires and thus both directions of fit within it. Taking the first option first, while perhaps desires and beliefs are not truly atomic mental states, is it plausible that beliefs and desires could be made up of further parts which have their own directions of fit? It’s difficult to see what the candidate states for constituting them

---

77 Smith, 1987. p 56, also ibid.
would be, given that these would need to have their own directions of fit, because to have a direction of fit a state could be no simpler than desires and beliefs in fact are, and so these could not be a further simpler part of a belief or desire. To have a direction of fit, a mental state needs to have something that is its object, either as part of it or something it stands in relation to, and some kind of attitude or emotion or stance etc., towards that object. This is really just to express the fairly obvious notion that a state needs at least two aspects to have a direction of fit, because to make sense of “fitting” we need to be able to fit X to Y or Y to X. You can’t fit X to itself in any meaningful sense. If a mental state is comprised only of some kind of object, then you have something like a representation or simple perception, and with only that there is nothing to make this fit with or to. If a mental state only has some kind of attitude then you have something like an emotion, and again there is nothing in this state that this can be either fitted to or have fitted to it. A mental state with a direction of fit needs at least these two elements of object and attitude then, but these two elements are precisely all that desires and beliefs involve. It’s true that desires, in addition to their objects and feelings of attraction/repulsion, are also dispositions (and perhaps beliefs are as well), but that is a modal claim, not some further, distinct constitutive element. So it’s difficult to see what further, more primitive elements could go into a desire or belief which would also have their own direction of fit. So the claim of JI could not be that a belief could have both directions of fit.

We are left then, with the possibility of a mental state composed of beliefs and desires each with their respective directions of fit. If the component beliefs only have the belief direction of fit, then the desire direction of fit of a complex must come from somewhere else than the beliefs in it, for example, from some desire. This, indeed, is what I think JI typical argues, that the mental state of some judgment is composed of a belief and a desire (or possibly more than one of each), and that the belief in a
judgment is necessarily followed by the desire: specifically, a belief that I have reason X to do Y is followed by the desire (and thus motivation) that I do Y. Considered this way, it seems that the JI is opposing the HTM in rather the same way that Nagel's objections from the last chapter did, in that the beliefs that are in the judgments are ultimately responsible for motivation, because they are necessarily followed by the desire which is our motivation – it is the beliefs which bring about this desire and thus the motivation. The desire, that I do Y, only comes about because of the belief, in this case, the belief that I have reason X to Y. What I will argue now is that such a desire does not necessarily follow from such a belief.

4. Direction of Fit of Judgments

The idea of a state like these kinds of judgments, comprised of a belief and a desire and thus two different directions of fit, is not obviously incoherent, as it is possible to try to make one aspect of our mental state fit the world, while trying to make the world fit another aspect of it. We see other examples of that, with, for example, concepts like disgust. Feelings of disgust follow certain beliefs that we form about a certain object. Note, however, that with the feeling of disgust it's fairly obvious that it is not necessary for such feelings of repulsion to follow true beliefs about the object. Many people are repulsed by spiders, yet other people dedicate their lives to studying them. It's clearly possible for both parties to have exactly the same beliefs about spiders, and yet end up with different desires relating to them. That spiders have many long legs and spin sticky webs is exactly what some people hate about them and other people find interesting. The name "disgust" is a term used for a complex state comprised of these two elements, but it doesn't mean that the one must necessarily follow the other, even if it usually, as a matter of fact, does. This is exactly the same as
with reasons judgments, as I will argue; a belief that something is a reason need not lead to motivation.

Let me discuss the belief of the judgment in a bit more detail here. I take it that the most generic type of relevant belief within a judgment involves a proposition like “X is a reason and therefore I should Y” where X is some consideration and Y is the action to be done, and therefore arguments that apply to it will apply to other possible JI beliefs. For example, that "X is good and therefore I should do it", or "X is in my interest and therefore I pursue X", and so on, seem to work in essentially the same way. Now, “X is a reason to do Y” is clearly a pretty normal belief. Whatever consideration X is, it must have certain features that lead us to conclude that it is a reason, for example, as noted previously Parfit believes that a state’s being painful is a reason to avoid being in that state, so pain constitutes a reason to do something in any situation that involves pain. If it turned out that that state would not be painful after all, then this would no longer be a reason to avoid it, in other words, our beliefs about what reasons we have would change. The belief "X is a reason to Y" is made to fit the world, clearly then, the belief that something is reason is a normal belief with the usual direction of fit: if something changes about the world, this belief adjusts to fit it.

As I said above, how we determine whether or not a consideration is a reason depends on features of that consideration, whatever they may be, and any consideration which has these features we will judge to be a reason – assuming of course we are judging correctly. It’s always possible that we could fail to detect the relevant features of a consideration or are simply outright contradictory in our judging, but these are obviously errors of various sorts and I won’t be concerned with these sorts of cases here. What precise action Y is appropriate will depend on what X is, that is, what actions we should do depend on the exact nature of the reasons before us.
Undoubtedly reasons are the kinds of things that are for being acted upon, but JI is not the only theory that can accommodate this claim. Reasons internalists think that that feature of reasons is best explained by limiting what considerations can be reasons to only those considerations which can motivate us. The positive arguments for that position must wait until the next chapter, but to some extent the arguments given here against JI will also work as arguments for that position, and vice-versa. In any case, underpinning each particular reasons judgment we probably have a general belief that all reasons for action are, *ceteris paribus*, to be acted upon, otherwise the fact that X was a reason would not be viewed as requiring action from us. Most likely this is believed to be true of reasons by definition: reasons for action just are those considerations upon which it is appropriate to act in the right circumstances. This, undoubtedly, is part of the reason why it seems plausible to think that judging that something is a reason will lead to motivation to act on it, since in so judging we have just judged that this consideration is indeed the kind of thing that it is appropriate to go and act upon. In fact, for most people most of time, the JI picture is exactly how it does go: they judge that something is a reason and they are motivated to act as is appropriate given that reason. Why that is I will discuss shortly. Crucially for me, it is, as opponents of JI have argued, at least possible to judge that I have a reason to do something without thereby being motivated to do it, and this is what I will argue in this chapter. In other words, while the belief that “I have reason X to Y” might normally be followed by a desire to Y, it needn’t always be – there is nothing necessary about the usual conjunction of these two states. This is why JI should be rejected.

A brief comment is required about an issue I wish to leave aside here: firstly, I said such judgments involve beliefs, but we might argue that they do not involve any

---

78 Of course, this doesn’t mean that people must have correct beliefs about their reasons. Somebody could believe that they have a reason which in fact they don’t, on the reasons internalist view.
kind of beliefs, and instead are merely be expressions of our feelings about a matter, and so on. If this mental state were not a belief, we would obviously be taking some kind of a noncognitivist position on it. I cannot possibly discuss even a selection of such theories here, but it seems to me that these theories would tend to be natural allies of the HTM, for if reasons judgments are not in some sense expressions of what we believe, then they are usually taken to be expressions of what we feel positively about. Taking judgments to be such expressions naturally leads to accepting the HTM, for generally we feel positively about something if we are attracted to it, which in my terms will mean that we desire it. In that case, judgments are just an expression of what we desire and therefore if we claim to be motivated by such judgments then we are essentially accepting the HTM.\textsuperscript{79} We are motivated to Y when we judge that we have reason X to Y simply because this judgment is just an expression of the desire that motivates us. It would not be that difficult, in other words, to make a noncognitivist position compatible with my positions here.

A final note: it is undoubtedly true that people in different cultures judge their reasons differently, such as a samurai who may judge what reasons he has very differently than I would. Even if we assume the truth of JI, therefore, it’s not clear that it really helps us get closer to an objective, universal system of morality than the HTM does. If motivation varies between people according to how they judge, and they judge quite differently from each other, then this is not too dissimilar from the basis we are on with the HTM when different people have different desires and thus different motivations. JI, like the HTM, requires further argumentation to produce the kind of objective morality that does not depend on psychological facts about the agent, that was typically the goal in rejecting the HTM, as I said earlier. In other words, JI does not

\textsuperscript{79} Brink 1989, p 44.
have as large an advantage over the HTM when it comes to arguing for objective
morality as might initially be thought.

5. That People Usually Are Motivated by Their Judgments

It is undoubtedly true that most of the time, most people are motivated to do
what they judge they have reason to do, and Smith takes this to be a crucial piece of
evidence for JI. Since I hold that it is not necessary that we be so motivated, I should
explain why contingently it usually is the case that we are so motivated, and so show
that JI is not the only possible explanation of this fact. This is most naturally explained
by the HTM as taking the belief of the judgment to be a typical means belief, where a
desire for some end is actually providing the motivation. This will either be some
particular end that will be served, for example that the grocery store sells food is a
reason for me to go there if I want food, or a more general end to do with how we view
reasons. For example, the desire to be rational will generally cause us to be motivated
by those considerations we believe to be reasons, given that acting on reasons is
necessary to be rational – I generally desire to be rational, and I come to believe that
acting on a particular consideration would be rational, so I come to be motivated to act
on it. While this kind of response is particularly appealing when talking about matters of
prudence or taste, for example, that I judge that something has tomatoes in it leads to
my becoming motivated to avoid it because I dislike tomatoes (so my end is to not eat
tomatoes, and the means to do that is to not eat this food), it is less appealing with
moral judgments. This is because while judgments of taste tend to vary from individual
to individual, and so we’re used to the idea that some people dislike exactly the same
things that we like, moral judgments don’t vary so much across individuals, at least
within particular cultures.

80 1995, p 72.
Judgments about reasons tend to be based, I suspect, on our upbringing and our culture, and most of us are taught to take much the same things as important, and because of this, we gain desires relating to the same kinds of things. For example, most children are taught that the fact that an action will hurt somebody is bad and an important thing to avoid – it is a reason not to do it. Consequently, almost everyone ends up judging that causing pain to others is a reason not to do something, and as such the fact that the belief and desire could come apart is more obscured than with other cases. Svavarsdottir suggests that the appeal of moral JI is really founded in the fact that since a young age the overwhelming majority of people are trained to treat morality and being good as very important. They have a hidden desire that they are not fully aware of, perhaps because it is one inculcated in them during childhood, and it is this desire that causes them to be motivated to do what they identify as good. We come to believe that X is good, and therefore conclude that something ought to be done about it is because we already have this desire to do good and be good (and thus, when we are children, be praised). Because the desire is so well entrenched in us, people are not always aware of it, and think that they move simply from beliefs about the nature of the good to the idea that they must do something about it. In fact they have a general desire to do what is good, and act accordingly. However, against this line of argument, Smith claims that it's implausible to think that people who are morally good have such a desire. He thinks that moral individuals desire individual ends which are morally good, such as specifically to avoid hurting others, and not a generalized desire to be moral. In fact, he claims that we would view such a desire as "a fetish or moral vice." I find this response implausible, as my experience has been that most moral people do indeed have a desire to do what is right, and there doesn't seem to be

---

81 1999, p 183. Foot also makes a similar point in "Morality as a System of Hypothetical Imperatives", Virtues and Vices, 2002, among others.
82 1995, p 75.
anything morally objectionable about such a desire, in fact it's worthy of praise.\textsuperscript{83} Probably that is not the only desire they have that leads them to be moral, and they could easily also have desires like wanting to help other people just considered in itself because they want others to do well, for example, but this is no problem for my account, so long as some desire is present to be motivating.\textsuperscript{84} Likewise, there could be multiple desires that explain why someone is motivated in line with their broader reasons judgments, such as the desire to be rational, or a desire to maximize their happiness, and so on.

Finally note that reasons internalism, which I will argue for in the next chapter, will also help explain this if we are not satisfied with the claim that there is a desire to be rational, etc. Reasons internalism holds that only considerations which motivate us can be reasons, and so presumably on such a view correctly judging what reasons I have will essentially be to determine what considerations are motivating or could motivate me, and so it will not be at all surprising that I would be motivated in line with my judgments, since my judgments would be following my motivation, and not, as JI typically holds, vice-versa.

6. Amoralism

The main objection to JI, and one which has been extensively argued over, is the possibility of an amoralist, which is someone who is not motivated by her moral judgments. JI assumes that people are invariably motivated by their judgments about what they have reason to do, particularly moral judgments, but according to Brink this "seems just false to both actual and possible psychological facts. Although indifference

\textsuperscript{83} Svavarsdottir 1999, p 170, also agrees and questions why such a desire should be considered morally unworthy, p 200.

\textsuperscript{84} Ibid., p 199.
to what is regarded as moral considerations may be fairly rare, it does seem to exist.\textsuperscript{85} If amoralism is possible then JI, at least about moral reasons judgments, is false, as it will be possible to be unmotivated by one’s own judgments about what reasons we have. An amoralist is not, note, an immoralist, as she does not deliberately seek to do what is immoral, which is a different kind of motivation based on moral judgments, but is rather simply not motivated do anything about moral concerns at all, at least some of the time. She is, then, somebody who judges that she has a moral reason to do something, and then does not have any motivation to do it. This need not be a systematic position on the part of the amoralist, and sometimes she may be motivated to do what morality requires, perhaps even usually so, just so long as at least once in a while a judgment fails to motivate her.\textsuperscript{86} Obviously amoralist cases are specifically concerned with whether or not moral judgments motivate, but presumably if it is possible that a moral reason judgment can fail to motivate, it is equally possible that other kinds of reason judgments may also fail to motivate, and therefore arguments for amoralism could perhaps be used for such other analogous cases about other kinds of reasons. In any case, even if they couldn’t be, breaking the link between judgment and motivation in even one class of reasons judgments is enough to undermine JI in general, since it holds that judgments and motivations are necessarily linked. Any instances where they are not linked is enough to show that the connection is not necessary. So here I will only concern myself with moral judgments in particular, since that is sufficient to refute JI. The amoralist, of course, presents no problem for the HTM, as that theory holds that motivation is produced by desires, not judgments, and so there’s no concern about judgments which fail to motivate.

\textsuperscript{85} Brink 1989, p 46.
\textsuperscript{86} Shafer-Landau 2003, p 145.
Judgment Internalists probably often focus upon moral judgments because it is usually thought that a moral reason is a very powerful one, which most of the time will override other concerns (except when, for example, the moral reason is very small and the other concerns very large), and therefore we would especially expect people to be motivated by their moral judgments. This is another reason why it makes sense for me here to focus specifically on amoralist cases, since it is here where JI may be thought to be strongest. Note though that this objection need not concern itself with what the strongest reasons are meant to be, however, as it only claims that somebody could judge themselves to have a reason and fail to be motivated by it – it does not require that this reason be the strongest one they had. Even if they would never have acted upon this reason because there were others of far more weight, what is important is whether or not they had motivation to act on it. It is easier to determine whether or not somebody was motivated by a judgment of a strong reason, of course, since a judgment that something is a small reason which apparently fails to motivate may simply be a case of that motivation being overridden by other motivations. Even if the other motivations which have overridden the favored reason of a judgment are motivations produced by what are judged to be weaker reasons, this is still not quite the amoralist position – instead these are cases of weakness of will. Weakness of will is somewhat different from amorality, being cases where motivation to act upon what we take to be our best reason is simply defeated by lesser reasons for whatever reason. JI faces no real problem in these cases, for all it need claim is that the judgment that there is a reason will produce motivation, not that the motivation thus produced will be stronger than any other motivation or other mental force, or that the strength of the motivation will comparable to the judged strength of the reason. Claiming that it must always produce the strongest motivation would be a much less plausible claim than that of JI proper, and there’s no reason to burden judgment internalists with it. Similarly,
cases where a judgment is not acted upon because there is no way to do so, say because we judge something about an event that occurred in the past, may not seem to produce motivation, but these are not true amoralist cases either. I did not permit these cases to count against desires motivating, and so cannot count them against judgments motivating here – the same argument against such cases will work in both cases.\(^87\)

For the amoralist objection what is needed is the existence of people who judge and aren't motivated at all by that judgment, whom we will call amoralists. Sometimes it is difficult to decide if a case is a question of weakness of will or true amorality, and this is often difficult to determine because sometimes the agent herself may not be entirely sure what has happened, as a weak, overridden motivation may be difficult to distinguish from no motivation at all. Nonetheless, the psychology behind such a state seems like a possible psychological state as Brink claimed. Naturally, a number of strategies for answering this objection have been proposed by defenders of JI, but all of these are unsatisfactory, as I will argue.

Examples of such a possible amoralist might be someone who thinks that morality is something to be surpassed, or ignored in favor of other goals or values. It could perhaps be someone who is concerned with a particular goal to such an extent that he or she is willing to do *whatever* is necessary to achieve that goal, without regard to morality; for example the painter Gauguin, who abandoned his family to become a better artist.\(^88\) We have to be careful with such examples, however, as a judgment internalist may argue that such people in fact only judge their morality ignoring philosophy or their overwhelming goal as providing actual reasons, and moral concerns or whatever else are judged as not actual reasons. If that were the case then their psychologies would fall in line with what is predicted by JI, as they would be motivated

\(^87\) Chapter 2, section 9.  
\(^88\) Lawrence in *Virtues and Reasons*, p 140.
by what they judge to be their actual reasons. However, it is not really that plausible, in the case of Gaugin for example, to think that he really didn’t recognize anything other than his painting as an actual reason for him. Of course, it is open as always for the judgment internalist to argue that such people were motivated by their judgments, but their motivation was just overridden. Undoubtedly sometimes this is what has happened with people who ignore particular judgments, but we don’t seem to think that all cases must be explained in this way. If we say, for example, that Gaugin didn’t care about his family, we may genuinely be saying that he just wasn’t motivated to stay with them, that we think it’s possible that he wasn’t conflicted and that he simply didn’t care about them (whether or not that was actually the case).

The amoralist might be a person suffering from depression or some other mental condition, who still judges what his reasons are, may still accept the demands of morality and rationality and so on and believe that he should act on them, and yet be unable to make himself act, in other words, completely lack any motivation to act. Such cases are not be limited to more unusual psychologies either, but are the sort of thing we encounter in everyday life. For example, Goldman says that he recognizes that there is no good reason to eat meat, and that there are good moral reasons to refrain from eating it, and yet he continues to eat meat.89 In the end, he just doesn’t care enough about the moral reasons to stop eating meat; he remains unmoved by his judgment. It is not the case that he is motivated to stop eating meat, and yet suffers from a kind of weakness of will and so does so anyway: he really just isn’t motivated. I suspect similar things happen with many people and charity. Most people probably recognize that it would be good, in some sense or other and for a variety of reasons, to donate to charities, like those which feed starving people in Africa. Most people, however, donate very little to such charities. Even apparently very rational people who

have reasoned very carefully to this conclusion, can apparently fail to be motivated to
do what they judge to be right in this sense: I'm thinking here of Peter Singer who has
sometimes been criticized for not fully living up to his stated moral judgments.
Sometimes these people might be conflicted, but most of the time they don't seem to be
conflicted at all. They might say "I know I should do it, but I just don't want to",
indicating that our usual way of interpreting such cases is that they simply lack
motivation altogether – they don't even generally think about donating to charity. Why
do people typically not volunteer to work at a soup kitchen even though we usually think
there are good reasons to do so? For no reason other than that they lack motivation:
it's too much effort, they don't really care enough. Perhaps some people have some
motivation to go and work in a soup kitchen, and this motivation is simply overridden,
but most are not conflicted in any way, they just aren't motivated to go in the slightest.
Our judgments about any kind of reasons may fail to motivate us, not just moral
reasons. It's not necessarily the case that judgments about prudence motivate us; for
example, we can imagine somebody who fails to be motivated to do what is necessary
to maintain her health, even though she really does judge that she has a very good
reason to do so – almost everyone thinks their health is a very important good after all.

Again, the amoralist objection to JI is simply that such a person is possible. If
such a person is possible, then judgments about what reasons we have do not
necessarily motivate, since it is possible that they do not motivate at least these people:
in other words, the belief and the desire of the judgment can be separated. Clearly the
case against JI is best proven if it can be shown that amoralists do actually exist, and
preferably that at least sometimes they are that way not for physical reasons like
depression. If only the latter kind of amoralist exists, then the JI may just argue that
certain conditions like depression prevent normal operation of the mind, and that we
cannot generalize from that to conclude that reason judgments need not be motivating
in the normal operation of our minds. Many defenders of JI attempt to defuse the amoralist objection in general by arguing that the only type of "amoralist" that is possible is a sort of pseudo-amoralist who only appears to make genuine reason judgments, but is in fact making judgments which are defective or not legitimate in one way or another. I will argue against this kind of response.

7. Failing to Judge Correctly

Perhaps the most common response by judgment internalists to the possibility of amoralism is to claim that these cases are really people who are somehow not quite making the judgments that it initially appears they are making, even that they hold themselves to be making. One line of response holds that the amoralist is not really making moral judgments at all, but rather that she is using moral terms in "inverted commas", that is, mentioning them rather than using them. According to this response the amoralist isn't really judging that "this is good", but rather "this is what people view as good" or "according to the rules of society this is what is good", or something along those lines. Hare describes these judgments as using evaluative terms purely descriptively, rather than to commend. If the inverted commas response is correct, then the amoralist only thinks she is making genuine moral judgments, but in fact is failing to do so. We can generalize this line of objection for reasons in general by saying that amoralist types are judging that “this is what is generally considered a reason” or maybe “for other people in my circumstances this would be a reason but for me it is not so”, and so on. The amoralist is therefore not truly making judgments about reasons at all, but rather, factual judgments about what society or other people, etc., think about certain things, and as such, these judgments are just normal factual beliefs, and not the kind of beliefs about reasons for action that JI claims motivate. The idea is

---

90 Smith, 1995, p 69.
91 Hare 1961, p 149.
that the amoralist is failing to actually recognize this consideration as a reason, and only thinks she is doing so – she is recognizing the factual claim that promise keeping is good in most people’s eyes, but not accepting that it is good, in her own eyes. This is the sort of sense that Thrasymachus in the Republic is probably thinking when he talks about justice. 92 “this is what people call justice, but it is not true justice that I recognize.” Or perhaps the amoralist is simply mistaken about what a reason is – she doesn’t realize that this reason, like all others, is for acting upon, and so she’s not really taking it to be a reason after all, even though she may believe she is (she’s using the term “reason” incorrectly in other words). Judgment internalists seem to think that were this amoralist to make a genuine reason judgment, she would be motivated just as JI holds. For example, the amoralist is judging “other people think promises give reasons”, and if she were to judge “promises give reasons”, then she would be motivated by that judgment just as JI predicts.

It does seem like it is possible for somebody to not really believe something they claim to believe. We do sometimes think, for example, that somebody doesn’t really believe something, even if she claims to believe it (and perhaps even if she seems to genuinely think that she does believe it). Somebody who consistently fails to act in line with a belief he claims to have, for example, somebody who never takes a medicine he says he believes will definitely help his illness, even though he wants to be cured, might be someone about whom we would say doesn’t really believe the medicine will help him, depending on the exact situation. If such cases are rare, that’s no problem for JI as amoralist cases may not be very common either. We generally think that somebody who fails to act on their reasons is usually suffering from some form of weakness of will, rather than true amorality. It’s also true that at least some of the supposed amoralist cases do appear to fit the inverted commas interpretation. The

92 Hare 1961, p 149.
Nietzschean amoralist, for example, is likely to judge that certain things are good by the standards of good and evil, but doesn’t endorse these standards himself, and so the Nietzschean may be judging that these standards do not provide real reasons, at least not for him.

What is really in question however is whether this must be the case for everyone who is an amoralist; since JI holds that amoralism is impossible it must explain every case away, and there are examples where the claim that the amoralist is not making a genuine reasons judgment looks less plausible as an explanation. Does Goldman mean, when he says that he has judged he has a reason to give up eating meat, not to make an evaluative claim but merely to say that eating meat is wrong according to some standards or other, or some other kind of purely descriptive claim? This doesn’t appear to be what he means, as he says that he accepts the arguments against meat eating himself. He recognizes the arguments against meat eating and has judged that it is indeed wrong for him to eat meat – he recognizes, in other words, that there is a reason on which he ought to act, and yet is not motivated to stop doing so. Likewise, most people who don’t donate to charity probably do believe that it would be good for them to do so, not just that it is something designated as good by some authority or other, but accept that it is something they ought to do. Some of these people are undoubtedly motivated to do so and have that motivation defeated by another stronger one, but some of them, I think, truly aren’t motivated by this judgment.

Presumably it is thought that the reasons we have are generated by what is valuable: either what is actually valuable, if we are a kind of realist about value, or simply the things that we do value. There seems to be an assumed premise behind JI that judging something to be valuable entails being motivated by it, except perhaps in exceptional circumstances, and this looks to be driving the typical inverted commas
response: judgment internalists may think that if the amoralist truly did value whatever she is judging to be a reason, it would be impossible for her to be unmoved by it. Since the amoralist is not motivated by a certain consideration, she is not really valuing it after all. Goldman, however, argues that he does value the moral considerations he thinks are served by not eating meat\textsuperscript{93} – it’s because he values them that he was able to pick them out and come to the conclusion that morally he ought not to eat meat. According to the inverted commas response, Goldman must be mistaken about what he values, and perhaps about what it is to value, because if he really did value these considerations he would at least somewhat motivated to stop eating meat. Most people view volunteering for charity work or simply donating to charity to be valuable, and yet most people are not motivated to do it. It’s difficult to say for how many people this is a question of weakness of will rather than a complete lack of motivation as amoralist cases require, but most of us have surely felt the difference between, say, knowing that you should do some work right now and instead procrastinating, which might come along with feelings of guilt or self-recrimination, versus that “intellectual” feeling that you really ought to do something without feeling any inclination at all to do it. It is presumably this kind of feeling that Goldman experiences when he reads the arguments for vegetarianism. We see people saying that they feel guilty for cheating on their diet much more often than they claim to feel guilty for not donating to charity, even though they would surely tell us that charity is important and valuable; in fact they would almost certainly say that the lives they could save by donating to the developing world are more valuable than their diet. It seems possible that people can value considerations without being motivated to do anything about them, and that they can judge something to be worth acting upon, without gaining a motivation to act upon it.

\textsuperscript{93} 2009, p 168-9.
Lenman, perhaps following Hare, objects to the possibility of amorality by arguing that if there's nothing necessary about reasons judgments being followed by a desire to act accordingly, then it ought to be possible for there to be nothing but amoralists, but this, he thinks, is in fact not possible. He describes a world where certain moral scientists are very good at determining moral reasons, note them all down, but absolutely nobody is interested in doing anything about these moral reasons; that such a world is possible is, he says, a "preposterous" claim. JI can explain why this is: it's because amoralists judge “this is what is judged by some people to be good”, and so there must be some people who do indeed judge that that something is good in a non-inverted commas way, for the claim of the amoralists to make any sense. A whole world of amoralists is thus a "parasite [that] has no host." The absurdity of such a world depends, it seems to me, on what precisely we mean by the term "morality". It's true that it seems implausible to think that we would really call something morality if it were not recognized by anyone as the kind of thing that bears on action, and this is perhaps why a whole world of amoralists seems implausible in some sense. Perhaps, however, the only reason we link morality so closely to action is for the reasons given earlier: that we do indeed most of the time act upon it because we've been taught to do so. This would explain the intuitive appeal of such a claim. Whatever the case there may be, however, even if we grant Lenman's argument this doesn't truly defeat the objection of the amoralist. Even if it is true that not everybody in any given world can be amoral, it doesn't follow from that claim that nobody can be amoral, and the objection to JI only needs that some people can be amoralists, not that everybody can be one.

---

95 Lenman 1999, p 445-6. Lenman thinks that appropriate desires are necessary for moral motivation.
96 Ibid., p 453.
Greenspan suggests that the amoralist is making a mistake about the applicability of moral judgments, a slightly different, but obviously related, response to the amoralist objection which would avoid the above concerns.\textsuperscript{97} She says that amoralists are not necessarily using moral terms in inverted commas, because their mistake lies at the level of metaethics: an amoralist is genuinely judging what is moral and what isn't, but makes a mistake about the applicability of moral reasons to her. This would again make the amoralist something like a Nietzschean agent, who fails to appreciate that moral reasons apply to her.\textsuperscript{98} Greenspan accuses such an agent of failing to fully understand what moral reasons are and that they apply to her. Greenspan views moral reasons in terms of claims that others can make on us, so when the amoralist views moral reasons as being inapplicable to her, she is misunderstanding the nature of moral reasons, which is that they are reasons that others give to us. Since they are given by others and not self-given the amoralist isn't able to dismiss them because she is not those other people and has no control over the reasons they produce. The situation is somewhat analogous to how we cannot dismiss perceptions; perceptions are generated by things outside of ourselves, and so we cannot simply choose to have them not be applicable to us. We're given perceptions partly by features of the world outside our control, and the same is true of reasons. Dealing with the question of the applicability of reasons will have to wait until chapter 6, but in any case I'm not convinced Greenspan's argument would succeed as a defense of JI. The position of JI is that making a genuine reason judgment produces motivation; therefore if the amoralist is making genuine moral judgments, and yet is not being motivated by them, then this is a rejection of JI regardless of what other mistakes may be being made. In general then, JI will face the amoralist problem whenever the

\textsuperscript{97} Oxford Studies in Metaethics vol. 2, p. 185.
\textsuperscript{98} That they do, as Greenspan happily admits (p. 174), implies reasons externalism. There are reasons even if they don't motivate. I will say more on that position in the next chapter.
amoralist is viewed as making a genuine moral judgment, even if some other, non-moral, errors could possibly be identified.

To truly judge that something is a reason and yet not be motivated by it is a possible state of mind to be in. The amoralist is possible and JI is wrong that moral judgments must always motivate.

8. Reasons

There is one final objection to JI to make: JI does not tell us what a reason is. It remains for judgment internalists to fill in the account of reasons and explain what they are, and this may turn out to be a difficult task for them, because, as Darwall notes, the judgment internalist has a problem explaining how we can identify certain considerations as reasons, and why we should act upon them, if we can't appeal to their being motivating.\(^9\) He argues that if we are to make judgments that certain considerations are reasons then there has to be something about the consideration that allows us to pick it out from other considerations, and identify it and not those other considerations as being a reason for us (or for everyone). So JI must identify what it is that enables us to judge that a certain consideration and not another is a reason, in order to allow us to make a judgment about it and therefore be motivated by it. It will be difficult to find some element that all the appropriate considerations share, since our reasons are so varied, and so it will be difficult for JI to provide a unified theory of what counts as a reason. This is probably why some think that reasons are non-natural properties, and think that some considerations simply are reason-giving because they have certain non-natural properties, and that nothing more can be said about that. I objected to that kind of response in the last chapter. This is a further burden that JI has that my position does not, as will become clear in the next chapter, for the reasons

internalist can appeal to considerations being motivating as what allows us to pick them out as reasons for us.

9. Motivation Again

In this chapter and chapter 2 I have argued that desires motivate and beliefs cannot, and in chapter 1 I claimed that motivations are mental states that cause or explain intentional action. It is clear that motivations must have the desire direction of fit, as motivation is aimed at doing something, at making some change in the world (even if we don't yet know what change it would be) just like desire. We do not modify motivations solely because of the mere fact that whatever they may aim at is not the case, by a mismatch between their object and the world, unlike beliefs. I gave an essentialist account of desires, describing the features that constitute desires, and a functional account of motivations, describing them in terms of the role they play in regards to action, so to say that desires can be motivations is to say that desires, and not beliefs, can fill the role of a motivation. This situation is somewhat analogous to how a consideration can fill the role of a reason for us. Not all desires will be motivations because not all desires will be relevant to any given action or situation, since as motivations serve a causal and explanatory role for actions or potential actions, desires which do not serve that role will not be motivations at least at that time, although they might be later. The situation in which a particular desire could be a motivation may never occur at all (some might not even be able to occur), and in such a case a desire would never actually be a motivation; for example, if somebody's life is never endangered then their desire to live may never act as a motivation. All desires are potential motivations, but whether they ever act as actual motivations depends on circumstance. Beliefs, however, are never motivations because they simply have the wrong direction of fit.
One final point: it seems too that whims (urges, etc.) can also become motivations, just as desires can. Whims differ from desires in, as I have argued, not being dispositional: there might be a desire that we only feel once because it so happens that the circumstances in which the feeling of attraction or repulsion arises only occur once, but it is still a disposition in that we would feel it whenever those circumstances did occur, whereas the same does not appear to be true of whims. However, it is not the fact that desires are dispositional that makes them produce action, but the feeling of attraction or repulsion that does so. Since whims have such a feeling, they too could potentially act as a kind of explanation for why we acted. It seems then that just as desires can be motivations, so too can whims. In fact most whims would usually, although not always, be overridden by motivation from our desires – we typically only do what we have a whim to do if we have basically nothing else we want to do, in other words, no strong desires for the situation we are in (and, of course, if the whim is small and harmless). Sometimes, of course, due to weakness of will, or due to a particularly strong urge, or just because we have no particular desire working against them, we will act on our whims, and so they will act as our motivation for so acting. What this means for whims and reasons I will discuss in the next chapter.

10. Conclusion

Reasons judgments are typically thought to be a combination of belief and desire, where the desire follows necessarily from the belief. I have argued that this connection is not necessary, just typical. That judgments usually seem to be followed by action can be explained with reference to desires, taking the belief to be a means belief, but with desires providing the ends and thus motivation. The amoralist is an example of what happens if that desire is not present while the belief is, demonstrating that a person is not necessarily motivated by his or her judgments. JI also faces a more
difficult task in explaining how we can pick out what reasons there are than the alternative I am presenting.

I conclude, following on the this chapter and the previous chapter, that the HTM is true, and in the next chapter I will argue that reasons internalism is also true.
Chapter 4: Reasons Internalism

1. Introduction

In this chapter I will argue for the position of reasons internalism, which holds that reasons must be considerations which are capable of motivating us. A reason is the kind of thing that it must be possible for us to act upon, and following Williams’ arguments for reasons internalism, which I will discuss, I will argue that only reasons internalism can meet this requirement. If the only things that can motivate us are desires, as previously argued, it follows from that and reasons internalism that only considerations that can connect with our desires can motivate us, and only those, therefore, can be reasons. I will also discuss some arguments against various aspects of this position, as well as argue against the claim that we can have reasons for our desires. That position entails the incorrect relationship between desires and reasons, in that it makes desires dependent on reasons, rather than the other way around as my position requires.

2. Reasons

What I am discussing in this chapter are limitations on what constitutes a reason for action for me. The question internalism is primarily concerned with is identifying what kind of things we can potentially act on, not what kind of things we ought to act on. These cannot be quite the same question, since almost everyone appears to agree that there are some reasons that we ought not to act upon, or at the very least that there are some reasons where it is not clear whether or not we ought to act upon them. Our common sense thinking holds that, for example, the fact that I am late may be offered as my reason to run red lights, and understandable as such, without it being a justification for doing so. That I am late would make sense of my action and so be a reason in that explanatory sense, but would not justify my running a red light.
What makes a reason a reason that we ought to act on, rather than merely one that we can act on, I leave open, since it will depend on some further account of morality and/or rationality that I won’t be giving here. This distinction between reasons questions is often referred to as the distinction between explanatory reasons and justificatory reasons. I follow Dancy, as previously stated (chapter 1, section 9), in holding that there are not really two different types of reasons, but two different questions to be asked about our reasons for an action, and considerations can fill in either role depending on the circumstances, as I will discuss in a moment. Internalism is a position which seeks to separate reasons out from other considerations, not judge between them.

What is a reason? As previously stated, a reason is a consideration that makes sense of the question "why did you do that?" It is a consideration that explains why we could or did perform some action. It would be an explanation which we could expect to satisfy someone who asked us why we did something, and, unlike a motivation, is not necessarily a mental state. Of course, it is possible to perform an action for multiple (or no) reasons, and perhaps no single one of those reasons would on their own provide a satisfactory explanation of why I acted in the way I did, but together those reasons would form an explanation of why I acted. Returning to the previous issue, there two different questions to be asked when we ask "why": one "why" asks for the reason why we actually acted the way we did. Answers to that question take the form of providing what we call explanatory reasons; these are the reason or reasons that I actually did, in point of fact, act upon. We could also present such reasons that would explain how we might have acted in different circumstances, counterfactually, which could be potential explanatory reasons. The second "why" is asking for a reason that justifies doing that action - for example "why did you think that was the right thing to do?" For these questions, justificatory reasons are given as answers. These reasons explain why an
action would be the right or good thing to do, or the thing that we ought to do, usually by appealing to a certain value. There are not two different kinds of considerations that match up with this distinction, there are just these two different questions. We may act for the same consideration as both explanatory and justificatory reason, for example, that fact that somebody is my spouse can be both an explanatory and a justificatory reason for helping that person.

There is an important limitation on what can count as a reason, for both kinds of reasons, according to internalism, and that is that any reason must be capable of being acted upon. An action must be done for the sake of that reason for it to truly be our reason for action, and any consideration which is a reason must be a consideration which it is possible to act on. I will discuss Williams’ argument for what this claim that all reasons must be explanatory in this sense, that reasons must be capable of being acted upon, amounts to. It is the central argument for reasons internalism, which this chapter is arguing for.

3. Backgrounding Conditions

Firstly, let me clarify that reasons are not simply desires. Goldman notes that taking reasons to be desires produces problems with whims, in that it makes them all reasons, and also means that we fail to have reasons that even internalists might think we really do have, such as when we’re led astray by false beliefs.\textsuperscript{100} For example, it’s strange to think that somebody who believes that a glass of petrol is gin, really has a reason to drink that petrol and no reason not to drink it – we’d rather give an account of reasons that would allow us to say that this person only thinks she has a reason, and doesn’t actually, and so on.\textsuperscript{101} I accept these objections to that idea, and will discuss

\textsuperscript{100} 2005, p 509.
\textsuperscript{101} Williams 1981, p. 102. Unless otherwise noted, all references to Williams are to this work.
these cases further later in this chapter. Furthermore, there is Parfit’s question which I will address shortly, among others who have raised similar issues, about what reasons we might have for our desires. While I think this is a mistaken question, as I will explain, it doesn’t seem like a meaningless question, which it would be if desires simply were reasons. To say that we have no reasons for our desires, that we can have no reasons for our desires, is a substantive position, as is the opposite claim.

Furthermore, if reasons were desires, then the fact that we desire something would directly be our reason for acting to get it, and any considerations about the thing we desire would have only a secondary role in this picture, in giving rise to the desire. Often, though, we clearly don’t take our desire to be our reason, rather, when we think we have reason to get something, we think that it is some fact about the thing we want that is our reason for getting it – most likely whatever it is that makes us desire it. In other words, we take some consideration, not a psychological state, to be our reason. Pettit and Smith, for example, note that we often choose things because we take them to be valuable, not merely because we desire them. That is, it is the fact (whatever this fact consists in exactly) that something is valuable that we take to be our reason for acting upon it, not that we desire this valuable thing; it is the former that we would offer if asked for our reason. Likewise, if I want to drink some delicious drink, it is not the fact that I desire it that is my reason, but rather the fact that it is delicious. While this is not always the case, as sometimes we would be likely to offer a desire as a reason, allowing reasons to be considerations in general can include such reasons (for desires can also be viewed as a consideration) and also other such external states of affairs, or properties, and so on. Therefore, it’s best to take desires to be necessary background conditions for our having a reason, but not necessarily the reason itself.

102 1990, p 579.
So the relationship between desires and reasons is not one of identity, but one where desires are necessary conditions for a consideration to be a reason. This is sometimes referred to as "backgrounding" desires, in that they must be in the background for a consideration to be a reason for us, but need not be the reason itself.\footnote{E.g. Pettit and Smith, 1990; Schroeder, 2007, etc.} Reasons, then, are considerations. I will argue that they are some consideration which connects with our desires either directly or has the potential to do so after sound deliberation, and therefore which does motivate or could motivate after deliberation, which is to say that it is because of the desires we have that this consideration has the potential to cause motivation in us. For example, if I have a desire to look at red things, and I become aware that a painting is red, the fact that the painting is red will be a reason for me to look at it.

4. \textbf{Williams' Argument for Internalism}

My reasons internalism, the position that only considerations that can motivate us can be reasons (with some qualifications), is based on Williams' arguments for the position.\footnote{1981 "Internal and External Reasons", p 101-113.} He holds that what is a reason for somebody depends on what he calls the subjective motivational set, because only considerations which have some connection (even if we're not aware of this connection) to members of this set can motivate, and thus be reasons for action. Reasons externalism need be no more than the denial of internalism: that a consideration need not be able to motivate us to be a reason. Reasons internalism then is asserting a limitation on what can count as a reason, and externalism is simply the denial of that limitation.

Williams gives a very broad, and (deliberately) relatively vague account of this "subjective motivational set", as he means simply anything that motivates us. He views
the relative vagueness of parts of his account to be a strength, taking
indeterminateness to be inherent to our concepts of practical reason, and probably also
wishing to make his argument as broadly appealing as possible. By "subjective
motivational set", Williams primarily means desires, but also wants to include things like
emotions, personal loyalties, projects, and so on. In fact he suggests that in some
sense all these may be classed as desires, but leaves this point relatively open;\textsuperscript{105} however my arguments will be solely concerned with desires as I have previously
defined them. Clearly these desires need not be fundamentally egoistic, as we can
desire that somebody else do well for their own sake, for example. For a consideration
to be a reason, we must be able to become motivated by it after deliberating to it from
members of the subjective motivational set. In other words, any consideration that we
cannot become motivated by, if we have no member of our motivational set (our
desires) that it would appeal to, will not be a reason for us. This does not mean, of
course, that only considerations that we actually act on can be reasons. It is possible
that I may be motivated to act in a particular way, but do not act because other
considerations move me more strongly. This does not matter for the reasons internalist
just as the analogous situation was not a problem for the judgment internalist, where
judgments didn't lead to action. Just as a motivation does not actually have to produce
action to be a motivation for that action, so too a reason does not have to actually
produce action to be a reason. It must simply be a consideration that does, or would if I
were aware of it and understood it, produce motivation to act a certain way. In short,
Williams denies the claim that there are certain reasons that everybody has no matter
what the members of their subjective motivational sets, which he takes to be the claim
that there are "external" reasons.\textsuperscript{106}

\textsuperscript{105} p 105.
\textsuperscript{106} p 101.
Of course, these claims must be qualified, as Williams immediately notes. Even for the reasons internalist, a consideration which fails to actually motivate me may still count as a reason for me if it would be true to say that there's some sense in which it would "normally" motivate me or if it could motivate me if I were aware of it and had correct beliefs concerning it.\(^{107}\) I will fill in this account later in this chapter. He argues for this because we can easily fail to act on considerations that we would act on if we had accurate beliefs about them; for example, if I am normally motivated by the idea of earning a lot of money, it is natural to say that the fact that a job pays well is a reason for me to take it, even if I am unaware of the existence of this job and thus not motivated to do anything about it. Ignorance should not be the arbiter of what reasons we have. What reasons externalism really holds, then, is that a consideration may still be a reason for me even if it would not motivate me even if I were fully and correctly aware of it.

Williams argues that only considerations which are capable of motivating, meaning only those which do motivate or could motivate after sound deliberation from members of our subjective motivational set (for my purposes, desires) can be reasons because reasons for action are, in a fundamental sense, for acting upon.\(^{108}\) That is why we call them reasons for action. Reasons for action are reasons in that they answer the question why we did what we did, or why we might or could do something if we were to do it – so they, in some sense, explain our actions. If they cannot fill this role for a person, then they are not truly reasons for her after all; this is the basic thinking behind Williams’ reasons internalism. Therefore, no reason for action which somebody could not act for could really be considered a reason for action for him, for something which you do not act on is obviously no answer to the question of why you acted, and

\(^{107}\) p 102.

\(^{108}\) Williams is certainly not the only person to think this. For example, Darwall, 1983, p 35.
something which you cannot act upon would never be an answer to the question of why you acted.

Motivation is what you directly act upon when you act, what precipitates your action, and it is therefore necessary in any explanation of why you act the way you do. No explanation of why you intentionally acted the way you did would be satisfactory without the inclusion of your motivation for acting as you did. If, when you acted, there was a consideration that didn’t motivate you, then you would have some other consideration that did motivate you, and that would be what you actually acted upon, and in that case the first consideration which didn’t motivate you wouldn’t figure into the explanation of why you acted as you did. Therefore, you wouldn’t have acted because of this first consideration, so it couldn’t be your reason for action for that particular action, although it could still be a potential reason for you if you could become motivated by it after sound deliberation, and so we could still call it a reason for you, just not one you acted upon at that time. If, however, the consideration is something that could not motivate you even after sound deliberation, then it is a consideration that you cannot act upon, and this consideration could not, then, be a reason for action for you at all. To claim that a consideration that could not motivate us is a reason is to say that we have a reason that we cannot act upon, and is to give a reason which could not explain why you acted, which is contrary to what reasons for action are. If Bob does not want to drive in Formula 1 racing because he hates going fast, and generally has no other appropriate desires for driving in it (e.g. a desire to be a famous racer), the fact that racing allows you to drive very fast would clearly not be a reason for him to race. Bob could not be motivated by this consideration, simply because he has no desire to go fast, and so he could not be motivated by that consideration, could not act for that consideration, even after deliberating on it, and therefore it could not be a reason for his action.
It is hopefully obvious that Williams' argument holds for what are often called "explanatory reasons", but it is perhaps not so obvious for the "justificatory reasons". Couldn't we be unmotivated by a consideration, and yet still think that it justifies our action? We could think that a consideration might justify some action, but it wouldn't be appropriate to call something which we cannot act upon a reason for action. The only way a consideration can be a reason for a particular person's particular action is if they act for the sake of that reason, for that is what it means for something to be a reason for action. If we are unmoved by a consideration, then our belief that it is there is irrelevant to our action, that is, we would have acted the way we do even if that consideration was not present. This non-motivating consideration is therefore superfluous to an account of our action, because even if we believe that some consideration justifies somebody's action it doesn't explain why she acted unless that consideration motivated her, unless she acted (at least in part) for the sake of it. Consider a person who buys a lottery ticket, because she wants to win a large sum of money. Often, some of the revenues of lotteries go to charitable causes, and so we might consider this a justifying reason for buying a lottery ticket. If the ticket buyer bought the ticket only caring about the money she wanted to win and not at all about the charitable donations, it's hard to see how these charitable donations explain her action; she didn't care about them, they were not a reason for her, even if they make what she did a good thing from our point of view. The explanation of why she bought the lottery ticket needs only reference her desire to win a lot of money, not the donations. Or consider somebody who wants to fix his car because it has a flat tire. It also happens that a back door is jammed shut. He is aware of this, but doesn't care about it, and doesn't intend to have it fixed because he'd rather save the money and never carries passengers anyway: he has no desires that would be served by getting it fixed. Both of these faults in the car might have been justificatory reasons (or explanatory reasons) for taking the car to the shop, but only the flat tire
actually is a reason, because that is what has a connection to his desires, whereas the stuck door has no such connection and is thus incapable of motivating him. The fact that the back door is broken has no bearing on his action: if the door weren’t broken, he would take the car into the shop or not all the same as when it is broken, and if the tire wasn’t flat but the door was broken he wouldn't take the car in. So the jammed door makes no difference to his action and it explains nothing. It could not be, then, a reason for the car owner's action in taking his car in to be repaired, explanatory or justificatory.

5. Sound Deliberation from the Subjective Motivational Set

I turn now to the stipulation that a reason is a consideration that can motivate us after sound deliberation from the subjective motivational set – meaning, for me, desires. As Williams notes, sometimes a person isn't motivated by a consideration which we would otherwise think was a reason for her, for example if a glass does contain gin and somebody who desires to drink gin believes it to contain petrol and therefore fails to be motivated to drink it. It would be a problem for reasons internalism if it couldn't explain why the fact that it actually contains gin might count as reason for her to drink it. Alternatively, certain considerations which we might be motivated by and mistakenly think are reasons are sometimes actually not reasons, for example, if a glass somebody believes contains gin actually contains petrol, meaning that she has a reason not to drink it that she isn't aware of, even though she desires to drink gin and is thus motivated to drink it. What, then, can the internalist say about these kinds of cases given that we want to say that somebody can have reasons she is unaware of?

These are the kinds of cases where the qualification of sound deliberation is required. Generally this situation would arise when we lack accurate beliefs which, if we had them, would lead to us being motivated or not motivated by some consideration.
For example, if I desire to eat a slice of bread, and so want to eat a slice of bread in front of me which, unbeknownst to me, is actually poisoned. It seems clear that in fact I have a reason not to eat the bread, and that reason is that the bread is poisoned, and I do not wish to be poisoned. Right now, of course, I have no idea that it is poisoned, and so I’m motivated to eat the bread, even though if I knew it was poisoned I certainly would not want it since I want to live much more strongly than I want a slice of bread. It seems necessary to stipulate also that the beliefs that would lead to us being motivated are true beliefs, in order to rule out the following kind of case: when somebody thinks that drinking petrol is excellent for her health, we would not take this to be a real reason for her to drink a glass of petrol because petrol is not actually going to satisfy the desire she believes it will. Bringing truth in here no doubt brings up other issues, but all I can say here is that from the point of view of an outsider, considerations which only motivate a person due to what we take to be her false beliefs about them do not seem like real reasons to us – they are not the result of sound deliberation. These sorts of examples are one of the most important strengths of the case for reasons externalism, as it most likely has no theoretical problem here. Even though externalism would still need to explain why these things are reasons in the first place, whatever that explanation is, it would not be dependent on a consideration being able to motivate, and so the issue of certain considerations which appear to be proper reasons motivating or not is irrelevant to the externalist.

A brief note here: it’s also possible that certain mental states, such as depression, might cause us to not be motivated by a consideration that we would normally be motivated by. Cases of nonstandard psychology like this will always be problems for any account of practical reason, being exceptions to whatever rules we formulate, and as such I don’t intend to address them here, as I am trying to give an account of the typical, or standard, way our practical reason works.
What can reasons internalism say about such cases as those above? The conclusion we want is clear: that they are, or would be, reasons for us under the correct circumstances. This is the point of Williams stipulation that a consideration is only a reason for us if it is capable of motivating us given our desires, that is, there must be a deliberative route from our subjective motivational set to a consideration, such that we could, by reasoning correctly, come to be motivated by it – but that doesn’t mean we actually need to take that route, or even, perhaps, that we are capable of deliberating soundly. If we have false or missing beliefs, for example, this would not count as reasoning correctly, and so while certain considerations will not actually motivate us, this is only because of those missing beliefs. They would motivate us if we had the proper beliefs about them and if we then reasoned soundly from those beliefs, because they do, in fact, have a connection to our desires. Williams is vague here, again, but the general idea is clear: if this condition can be fulfilled, then the reasons internalist can allow that a consideration that fails to actually motivate us can still in some sense be a reason for the person. It’s something that would be a reason if we reasoned correctly about it, but only if it motivated us after correct deliberation, so the condition of being able to motivate is still demanded, and the reliance of reasons upon our desires is still maintained. Therefore, this account only extends internalism to cover cases that should motivate us, given our desires, and does not extend it to considerations which have nothing to do with our motivations, and is still an internalist account.

If we take desires to be dispositions, as I have argued, then the account of these possible reasons is as follows: a consideration is a reason for me if it would connect to my desires, meaning that I would be attracted by something about it, and I would become motivated by it, if I were aware of it and had correct beliefs about it. If I have a desire to drink gin, then the fact that this is gin will be a reason for me to drink it because I have the desire to drink gin. However, if it is not the case that this is gin, then
I only think I have a reason to drink it, but I really don't, because it is not in fact the stuff that I desire. Internalism can make this move if it does not hold that desires themselves are reasons, but rather considerations, as then it matters what the consideration is, not simply that we have certain desires and beliefs. Otherwise merely the fact that I mistakenly desired it would be enough for me to have a reason to drink it, and that seems wrong. What the truth is about a consideration matters for our desires because only certain considerations are actually of the type desired, even if we don't realize which ones are of that type, and therefore it's reasonable for the internalist to conclude in turn that the truth about a consideration also matters for our reasons.

These possible reasons do not make reasons dependent on beliefs, despite their reference to our having correct beliefs; instead, our reasons are dependent on our desires and what the facts of the matter are. These are cases of our being mistaken about what is the case, and so forming mistaken views which in turn affect which desires are manifested, and it is because of this that we appear to have certain reasons. This is part of how beliefs direct our motivations, but do not ultimately determine what motivations we have. Our mistaken or missing beliefs may direct our desires towards certain considerations, and so we may appear to have reasons relating to them because we are making such a mistake, just as we may mistakenly think we desire something because we are wrong about what exactly it is, but this doesn't mean that we really do have such a reason, or that we really do have such a desire. I have a reason to drink gin, and this glass does not contain gin, I only think it does, and so I only think I have a (derived) reason to drink this glass of liquid. The consideration which would be my reason, that it is gin, is not present, I only think it is, and so I don't really have a reason here: just as something can appear to be a reason for belief, and thus we might form beliefs in accordance with it, without actually being a real reason for
belief. Beliefs clearly have an important role, but it is once again a role primarily of correctly aiming our motivation.

Why does Williams make the stipulation of sound deliberation from a member of our set of motivations? Williams thinks that a consideration which has no connection to your subjective motivational set, or in my argument, your desires, isn't able to motivate merely after sound deliberation. He says, "there is no motivation for the agent to deliberate from, to reach this new motivation"\textsuperscript{109} if the consideration does not connect to your motivations. Williams says that the externalist claim here must be that, regardless of her original motivations, the agent can rationally deliberate and come to be newly motivated by this new consideration, because they recognize it as a reason. The externalist thinks that we can become motivated by a reason without already having a related motivation that bears on it. Therefore what the externalist must hold, Williams argues, is that the new motivation is gained specifically due to the recognition of this consideration as a reason, due to a belief that this consideration is a reason, because there is nothing else that could produce the new motivation – in other words, this is JI again. Williams doesn't see how somebody not previously at all motivated by a consideration can come to be motivated by that consideration merely by a belief. I argued against that possibility in the previous chapters; belief alone cannot cause motivation, but desire can.\textsuperscript{110} The amoralist argument of the last chapter is also relevant here: it may be that people often are motivated once they recognize that something is a reason, but that is because they have standing desires that mean they want to act on what they take to be reasons, not because there is a necessary connection between belief in reasonness and motivation. There must be some desire that we have in order to be motivated by a consideration, even after sound deliberation.

\textsuperscript{110} I cannot discuss the possibility that mental states other than desires or beliefs could motivate here, simply because that is too large a project to take on.
Of course, people can gain new members of their subjective motivational sets which then motivate them in new ways, including immediately when faced by something the externalist would call a reason, and thus come to be motivated by a consideration towards which they had no motivations prior to experiencing it. Williams mentions being convinced by moving rhetoric and coming to be motivated because of it, because, presumably, it produced new desires in us, for example. The natural internalist account of such cases is surely that the person gains a new desire in such cases, which then motivates her. There is, as I've previously said, no reason that we can't gain new desires after experiencing new things (we can easily gain a desire to rock climb after experiencing it for the first time, for example) or even simply by being persuaded by someone, and thus still be reasons internalists in such cases by attributing motivation to these new desires, rather than any new beliefs about the considerations in question. Since beliefs cannot motivate, as previously argued, it must be these new desires which are motivating.

Reasons, as I stated, are answers to the question why we did something. A consideration that we can't act upon would not be one that we could accurately offer as an answer to this "why" question, because it could not be why I did what I did. It might have been something that could have been a reason for me if my desires were different, or could be or actually was a reason for somebody else, but it was not a reason for me. Williams notes that it's possible that somebody could have a general desire to act on those considerations that are taken to be reasons. Therefore, it is possible to judge that something is a reason, and then be motivated to act upon it, and this would be compatible with the HTM and internalism since the motivation proceeds

111 p 108. Such cases could probably often be explained by reference to existing desires like the desire for glory, or to do one's duty, and so on. They needn't necessarily be, however.

from the desire to act upon reasons, that is, from the subjective motivational set which includes a desire to act upon those things we take to be reasons.

6. Unwanted Desires and Whims

It might be thought that there's an objection to internalism concerning those desires that we don't want to have and repudiate, that we have a desire to be rid of, specifically that such desires don't or shouldn't give rise to reasons. Certainly such desires exist: such as a drug addict's desire for the drug he is addicted to. This is a genuine desire: he has a very strong feeling of attraction towards the drug, and consistently feels this pull whenever the object, the drug, is in mind or in front of him. It's probably one of his most consistently manifested (and strongest) desires, in fact. All of this is to say that his mental state meets my definition of a desire, given in the first chapter. He has a desire for the drug, and so is motivated to take it, and yet we may think that this desire does not really give him a reason to take the drug. It's also obvious that whims can motivate; is it not possible, then, to gain reasons from whims? Yet this seems problematic too, as we might wish to say that a whim to jump off a tall building (to my death) would not actually give me a reason to do so. All there is for me to say here is that internalism provides a limitation on what can be a reason: it must be a consideration which is capable of motivating. It doesn't follow from this position that this is all that being a reason consists in. It may be that there is some further limitation or limitations on a consideration being a reason that have yet to be filled in, and that this would rule out such desires and whims from being reasons. I don't propose to provide such an account here. This is not much of an advantage to externalism, since it must still explain how considerations are to be identified as reasons at all, whereas internalism already provides a stipulation that allows us to exclude those considerations
which definitely could not be reasons, and so externalism still has a heavier explanatory burden than internalism.

Having argued for reasons internalism, let me now turn to the case for reasons externalism. Arguing against externalism is, of course, another way to argue for internalism, since externalism is simply the rejection of internalism. The primary argument for internalism is Williams’ argument, but it is worth looking at some objections to rejections of that argument.

### 7. Reasons Externalism

I will first discuss some general points about externalism, before turning to more specific accounts of it. There are two issues which any account of externalism would seem to face that internalism does not. The first I mentioned in reference to Ji in the last chapter, and that is the difficulty that externalists would face in trying to pick out which considerations were reasons if they cannot appeal to motivation. The externalist may indeed believe that reasons often actually are motivating, or even that, for whatever reason, they always are; externalists simply reject the claim that a consideration not being able to motivate (after sound deliberation) means that it is not a reason. This can give externalism problems in trying to identify what reasons are, as Darwall notes. He suggests that the fact that a consideration is or would be motivating is the most natural factor that would allow us to identify it as a reason, and if that’s not how we pick it out, then it would probably be by some strange property of “reasonness”, or perhaps something like the claim that it’s intuitive which considerations are reasons: “The internalist move seems the only alternative to treating the property of being a reason as a further, non-natural property that is itself the object of a special sort of
Indeed, that latter is the direction that Parfit appears to favor. I will discuss a few of Parfit’s arguments later in this chapter, as well as some of McDowell’s, who thinks that the virtuous person knows which considerations are reasons because they are virtuous, and it isn’t directly a question of their desires. Neither of these accounts succeed, as I will argue.

Reasons externalism usually faces a further lacuna in its account of practical reason that reasons internalism doesn’t face, which is closely related to Williams’ argument above. Once the account of what reasons are has been given, it still remains to explain why we would act on them, something which can prove problematic for this position. In reasons internalism, why we act on our reasons is essentially a closed question, as we are motivated by them. We already have an answer as to why: we are motivated to do so. This is not to say that we can’t question whether we should act on the considerations that motivate us – that is, whether particularly reasons justify a particular action, but this is a separate issue. If reasons externalism is true, then it’s possible to ask why you should care, at all, about these considerations that are claimed to be reasons for you. That would be a strange question to ask about internal reasons, since you have to care about them for them to be reasons for you in the first place.

Neither of these issues on their own prove the case for reasons internalism. They do, however, demonstrate some advantages that internalism has over reasons externalism. Williams suggests, and I agree, that claims which appear to be appeals to external reasons can really be understood in senses which are compatible with internal reasons. That is to say, the sense of statements which appear to be external reasons statements that we want to keep, can still be maintained on internalism. I will

---

113 Darwall 1983, p 56. Velleman also notes this general issue for externalism, 2010, p 175.
114 p 111.
return to this issue in chapter 6, as this issue relates to how I propose to reconcile the possibility of objective morality and the HTM and internalism.

8. McDowell's Objection

McDowell’s argument concerns moral reasons only, but it is worth considering as an argument for at least limited externalism. McDowell's primary position is that the virtuous agent recognizes certain considerations as being moral reasons, because he or she is virtuous and the virtuous person can do this, and moral reasons are the kinds of considerations that virtuous people recognize as reasons.\(^{115}\) Fundamentally, as McDowell himself states, this is the position that the belief that there are certain reasons can motivate us.\(^{116}\) It is then a rejection of the HTM, since it holds that beliefs can motivate us, and also a rejection of internalism, since it holds that certain considerations are reasons before they motivate, and therefore a consideration does not have to be capable of motivating to be a reason. The arguments for the HTM given in the previous chapters will suffice here when combined with Williams’ argument. Beliefs, as I've argued, cannot motivate, and therefore beliefs are not sufficient by themselves to explain why we acted as internalism requires reasons do. This position is to be rejected.

There is an objection to internalism inspired by some issues that McDowell mentions that is worth discussing here, as it raises an important point. McDowell suggests that one reason to favor externalism is that internalism implies that desires can be “aggregated just anyhow”\(^{117}\) – in other words, they could be incoherent. If this were the case, then it looks like internal reasons likewise could arise in an incoherent

\(^{115}\) 1998, p 80-1.
\(^{117}\) 1998, p 106. He believes this is incorrect. His argument in this passage is not quite the objection I am outlining here, although what argument it is precisely is not entirely clear to me.
jumble, and that might mean that we could act on reasons and yet also incoherently. It's hard to see how acting incoherently could be rational, and yet we are indeed doing what we have reason to do; therefore internalism would lead to the possibility of an agent acting irrationally while acting on genuine reasons. However, I don't think this is a problem for internalism, because it is not obvious to me that there's something fundamentally irrational about having conflicting reasons for action. We probably have conflicting reasons for action almost all the time, without thinking that this makes us irrational. Generally there is more than one course of action available to me for the next hour, for example, yet clearly I cannot act on every reason that I have for actions that I could do in the next hour because it's simply not possible to complete the whole range of available actions in that time frame. Insofar as an objector thinks that internal reasons could be incoherent, this situation, where I can only act on one or the other reason but not all of them, must surely count, just as an incoherent set of sentences is a set where some, but not all, of the members can be true at the same time. Is it irrational to get into a situation where I don't have time to act upon all my reasons? I don't think it is. If it were, it seems like we'd be irrational all the time. Therefore there's nothing necessarily irrational about having a set of reasons that cannot all be satisfied.

9. Reasons Prior to Desires

One common argument given in support of externalism is that we can, and should, have reasons for our desires. If this is correct, then the combination of reasons internalism and HTM cannot be true, for there would have to be at least one reason which came before any desire, and would therefore be prior to any possibility of motivation, and be an external reason. Parfit has a view like this, and I've already discussed the fundamental concerns with his project (chapter 2, section 6) that lead me
to reject his account of how this might work, and those objections hold against arguments for desirability as a non-natural property in general.

While Parfit suggests that the fact that something really is desirable in some way is a reason to desire it, Brink suggests instead that it is a belief that something is valuable that is a reason to desire it and causes us to desire it.\(^{118}\) He believes that there being some kind of reasons for desires (and value) is necessary to explain why we recommend or condemn certain things to other people: "an appraiser’s audience [would not] have much reason to share his attitude or heed his recommendation if it had no reason to regard...[what] he was recommending as correct or valuable."\(^{119}\) On the face of it, it does seem like there could be a problem for the reasons internalist here – if my reasons depend on my desires, how could I claim that these reasons for condemning or valuing something are reasons for other people to do likewise, and how, therefore, could I claim that they ought to do what I say they should? My response to such issues will come in chapter 6, but in short I deny that people do necessarily have the same reasons to value or condemn the same things that I have reasons to value or condemn.

Those who argue for there being reasons for desires will no doubt want an internalist like me to answer the following question: if there are no reasons to desire what you do in fact desire, why not simply abandon your desires? Why should desires give rise to reasons, if there’s no reason to have desires in the first place? It would seem equally appropriate to simply abandon them rather than attempting to satisfying them, as if desires are not based on reasons then there’s no reason not to abandon them rather than do what they instruct us to do. My reply is that while it might be true that there is no reason not to abandon your desires, this is because there are no

\(^{118}\) 1989, p 64.
\(^{119}\) 1989, p 79.
reasons prior and external to motivation and thus desire. Therefore, while there may be no reason not to abandon them, there would be no reason to abandon them either. If we have reasons to satisfy our individual desires, then obviously we have more reason to satisfy them than to do something which we have no particular reason to do, that is, abandon them. Insofar as it makes sense to talk about reasons for desires at all (which I will discuss next), the "reason" we have to satisfy them is nothing other than that is how we react to them, to make the world fit them, and this is just how we are so constituted. We try to get whatever we are attracted to.

Heathwood\textsuperscript{120} suggests that there is a further problem with desires being the source of our reasons if they are ungrounded on reasons themselves, and therefore concludes that there must be reasons for desire. If it were the case that there can be no reasons for desire, and that our desires somehow give rise to reasons (as the combination of the HTM and reasons internalism seems to conclude) then our reasons would, in some sense, be arbitrary since they would arise due to mental states which themselves are arbitrary, being founded for no reason, and this is unsatisfactory, particularly for justificatory reasons. We can see the appeal of such an objection when we consider desires which seem highly irrational, for example, somebody who has a desire to not step on cracks in the sidewalk, which we typically probably think doesn't give rise to any kind of justification for avoiding stepping on cracks. However, if we are talking about the fundamental source of reasons, then there simply couldn't be any reasons for those things being the way they are, since those things – whatever you want to use to fill that account in, desires or some non-natural property, etc. – are in some sense the ultimate source of reasons. Reasons arise from them, so naturally there couldn't be reasons prior to them. Whether it is desires, or normative facts about objects, or whatever you might claim as the ultimate source of reasons, either they will

\textsuperscript{120} Oxford Studies on Metaethics Vol. 6, 2011 p 87.
truly be the foundation of reasons, in which case no reasons prior to them can be located, or they will not be, in which case reasons for them can be discovered. In that latter case they are not really the source of our reasons, since clearly reasons predate them. So any claimed source of reasons will suffer from this "arbitrariness" in the sense that there will be no reason why these things and not others give rise to reasons, and what of these things are present in the first place to give rise to particular reasons. Complaining that it is unsatisfactory that there are no reasons to hold our particular desires which give rise to our particular reasons is like complaining that a reason given to care about morality in general is not itself a moral reason. It should not be since it is the basis for morality. As far as explanatory reasons go, this is all there is to say on this matter. As for justificatory reasons, which desires are "irrational" and give rise to unacceptable reasons is not self-evident but requires some further standard of judgment, since it depends on our various opinions on what kinds of consideration justify an action, and there's nothing stopping somebody from excluding such desires from giving rise to reasons on some other ground. I am arguing that reasons must be capable of motivating, not that everything which is capable of motivating gives rise to reasons, so it is open to say that there is some further factor which considerations must have beyond simply motivating in order to be a reason, and thus desires which do not satisfy such a further factor, like perhaps a desire to avoid stepping on cracks in the sidewalk, will not give rise to reasons, even if they do motivate. What that further factor could be, I leave open, but it's not necessary, in other words, to postulate the existence of reasons for desires in order to prevent these kinds of desires from producing reasons.
10. Korsgaard’s Objection

Korsgaard gives an interesting analogy in “Skepticism about Practical Reason” which is intended to work against a certain conception of reasons internalism. She claims that just as an argument doesn’t have to convince anybody to be a good argument, so too reasons don’t have to motivate (“convince”) anybody to be good (which I take to mean “real”) reasons. In other words, we might reject internalism by holding that reasons do not have to motivate us to be reasons, and draw an analogy between reasons for beliefs and reasons for action, as clearly an argument is an attempt to provide a reason for belief.

However, a good argument being convincing is not analogous to a reason merely motivating, it’s analogous to a reason actually causing us to act. If an argument convinces, then we form a belief according to that argument, and likewise if a reason for action succeeds, then we act according to that reason. Being convinced is not analogous to being motivated, because a reason for belief successfully convincing us means that that reason has “won”, whereas a reason for action being motivating does not mean it has “won”, as it may still be overridden by other motivations. Clearly the internalism requirement is not that only those considerations we actually act upon are reasons, which is what Korsgaard’s analogy would imply, but only that a consideration must motivate to be a reason. It does not have to lead to action to be a reason for action, just as a reason for belief does not have to lead to belief to be a reason for belief. It seems fair to say that Williams’ picture of reasons for action is somewhat analogous to reasons for belief: if we could potentially form, via some sound deliberation, a belief in response to a certain consideration, then that consideration is a genuine reason for belief. Even if somebody is unable to understand the significance of

121 1986, p 14-5. Korsgaard doesn’t fit neatly into the externalist/internalist mold, so it’s not helpful to think of her as espousing an externalist position in this paper.
some consideration and how it ought to lead to particular beliefs, it can still be a proper reason for belief. Likewise with reasons for action, where considerations which are capable of motivating us after sound deliberation are really reasons. It seems then that reasons internalism is analogous to reasons for belief in the way that Korsgaard wants.

11. Nagel’s Objection

Nagel suggests three unwelcome conclusions that making reasons dependent on desires would produce, however, none of these conclusions succeed in refuting internalism. Firstly, he notes that I may have a desire to have something at some point in the future, which will give me a reason right now to acquire it if desires are responsible for reasons. That is, if I am having the desire "I want to eat a persimmon next Wednesday", for example, I will have a reason to get a persimmon now. Say I also know that when next Wednesday comes I will not desire to eat a persimmon, I still have a reason right now to get the persimmon for next Wednesday, even though I know I won't be eating it when I have a current desire to get one. Nagel objects to the conclusion that I have a reason right now to do something next Wednesday, even though I have full knowledge that I will have no reason to do it on the Wednesday itself. If I were ignorant of the fact that I will actually have no reason next Wednesday, this wouldn't be a problem, as our reasons often change, and there is nothing unusual in that; but it is a strange situation, however, to think that we could have a reason to do something for a future time, when we know that when that time comes we will have no reason. Nagel is concerned that such a situation effectively pits the individual against herself over time, and thus cannot be a correct picture of practical rationality. The second unwelcome conclusion Nagel suggests is essentially the reverse of the previous case. Suppose I know that I will be assaulted by some desire in the future, which will

116

lead to my having a reason for action in the future, however right now I am not experiencing this desire, and therefore have no reason to do what I know I will have reason to do in the future. Again, this would be irrational according to Nagel. Nagel's third unwelcome conclusion is that if our expected future desires conflict with our present desires, we have reason now to promote something that we know full well we will have reason to oppose in the future.

My reply here is quite straightforward, and is similar to my previous reply to Nagel (chapter 2, section 10): the concept of desires as dispositional defuses these problems. The only way I'd know that I would have some desire in the future is because I know that I have this disposition to be attracted to certain objects. In other words, I have the desire now, as a disposition, and although this disposition may not be manifesting itself right now, it will in the future. Knowing that we will either desire or not desire a fruit next Wednesday, for example, will allow us to deliberate from whatever desire will be in play next Wednesday; since it is dispositional it should be present now, for if it is not, if it is to be a new desire we gain next Wednesday, how are we to know that it will be present next Wednesday? Therefore, there's nothing preventing a reason being based on this desire right now, since it is present. It is possible then to have reasons which are based on what we are going to desire in the future. It's true that the desire won't be motivating if we are not feeling attracted to the object at that particular moment, but utilizing Williams' stipulation of there being a sound deliberative route from your subjective motivation set to the reason will apply here. Since the desire is present, it would be possible, if we are reasoning correctly, to deliberate from what we have, the desire, to the reason, and so it's possible even for the internalist to say that we have such a reason, even if in fact we are not motivated and so don't act on it at the moment. Should an individual, then, act conflicting reasons as Nagel fears, the response for the internalist is simply that perhaps this person should engage in more sound deliberation...
to better figure out what reasons she really has. Note too that earlier I argued that having conflicting reasons is not necessarily a concerning outcome.

12. Conclusion

Reasons are considerations, not desires, however desires turn out to be necessary for a consideration to be a reason, because only considerations which do or would motivate after sound deliberation can be reasons. Internalism about reasons, the position that for a consideration to be a reason it must capable of motivating, captures a true limitation on the nature of reasons, because reasons can only be those things that we can act on, and something which could not motivate us could not be something we could act on. Therefore externalism cannot provide a properly explanatory account of reasons. Internalism can provide an account for how we could make reasons statements even in cases where we have false beliefs, which is perhaps the biggest problem initially facing the account. We’ve also seen how various objections to internalism ultimately fail to refute it.

In the next chapter, I will discuss what follows from holding both reasons internalism and the HTM, and what that seems to mean for our account of morality.
Chapter 5: The Problem

1. Introduction

I have argued for a version of the Humean Theory of Motivation, that desires and not beliefs can motivate us, and for reasons internalism, that only considerations which are capable of motivating us can be reasons. There is an apparent problem with holding both of these, a version of which Smith points out in *The Moral Problem*,\(^{123}\) namely that they appear to conflict with the concept of objective morality. As stated, my aim in this dissertation is to show that all three of these positions can be reconciled; my aim in this chapter is to discuss why and how these positions apparently conflict.

By objective morality I mean something quite general and basic, something along the lines of the claim that morality is at least not wholly relative or subjective. A crucial part of this idea is that genuine moral disagreements are possible, and that in such a situation, one party may be factually correct and the other factually incorrect. I take objective morality to be the claim that the content of morality, whatever it is, is not something that varies depending on the particular desires, attitudes, decisions, etc., that individuals or groups of individuals have, but that there are some kind of facts of the matter about what is right and wrong about which it is possible to be correct or incorrect, regardless of how people feel or think about them. It is crucially something which equally applies to everybody in the same circumstances, meaning that "If agents in the same circumstances act in the same way then either they both act rightly or they both act wrongly."\(^{124}\) I don't intend to fill this account in any further, as I wish to leave what is meant as broad, and as acceptable, as possible. The basic idea of morality being objective is one that I take it most of us understand at this general level.

\(^{123}\) 1995, section 1.3.
\(^{124}\) Smith, 1995, p. 5. Relevant circumstances may of course include the properties of individuals, such as intelligence or membership in a specific group.
Why would we think that the HTM and internalism conflict with objective morality? The problem is as follows: if reasons internalism is true, then only considerations which are capable of motivating me can be reasons for me. If the HTM is true, then desires motivate me. Combining these two produces the position that considerations which connect with my desires in some way can produce motivation, and thus reasons. In other words, if I lack particular desires then certain considerations could fail to constitute reasons for me, therefore, if I do not have a desire to be moral, I am not guaranteed to have a reason to do any particular action that might be considered the moral thing to do, e.g. give to charity. Even worse, if acting morally in general is contrary to my desires, then I may end up with no reason to care about morality at all. Therefore, it appears that the application of morality to me is not objective but merely relative to my desires, and this obviously conflicts with the view that morality is objective. That view holds that moral standards apply to everyone, as previously stated, yet I might escape those standards due to my desires. It might, in fact, so happen that the demands of morality always do end up fulfilling some desire of mine, and so in any given situation I will indeed have a reason to be moral even on the HTM and internalism, since I always have a possible motivation to do the moral thing. This doesn't solve the general worry, however, as this dovetailing of my desires and the demands of morality is not necessary, only contingent; it could just as easily be otherwise, and indeed it's implausible to think this convenient matchup would happen for everyone all of the time, but that it did would seem to be required for morality to be objective. Generally it is thought that morality is precisely that which makes demands on us that are indifferent to our inclinations. It may happen from time to time, or even all the time, that we are inclined to be moral, but this convergence of desire and morality should not be required for moral considerations to be binding, because they will
continue to be binding even when our inclinations are otherwise.\textsuperscript{125} Objective morality holds that people cannot escape the demands of morality merely by appealing to their contrary desires. It is not acceptable to murder someone simply because the desire to do so was stronger than any other desire, or because there was no desire not to murder. You cannot escape objective morality simply because you didn't feel like complying with it. The worry is that reasons internalism and the HTM combined produce the conclusion that morality is dependent on our desires, not independent of them as we might hope. If morality is dependent on our desires, then it cannot bind everyone equally, and in fact cannot be the same for everybody, for certainly not everyone has the same desires, and therefore morality cannot be objective. When two people in the same circumstances act in the same way, they may not both act rightly or wrongly, for such determinations may vary according to their desires – this is the worry.

This is, in brief, why there is thought to be a problem in holding all three of these positions. The natural question would be: why is it a problem for the HTM and internalism that the combination of the two of them conflicts with objective morality? In other words, why not simply abandon the idea that morality is objective? However, I will not argue for the truth of objective morality, as that is far too large a project to do here, but I can give two related reasons as to why it is a concern for the HTM and reason internalism if they are not compatible with objective morality.

The first is that, despite common attitudes changing somewhat in recent times, most philosophers and probably most people still believe that in some sense morality is objective, and therefore if HTM and internalism conflict with objective morality then this is \textit{prima facie} evidence against them. They would conflict with another position that we hold to be true, and this would be a reason to think that they are incorrect. Some form

\textsuperscript{125} Kantian ethics is the classic example of this position.
of this thought is, I believe, the primary motivation behind most philosophers who reject either or both of these positions. Nagel, Parfit, Raz, Scanlon, McDowell, and many others seem to have something like this at the bottom of their rejection of one or both of these theories. Because they think it is difficult to reconcile the HTM and internalism with moral objectivity, which they believe to be true, they suspect that something must be wrong with the combination of HTM and reasons internalism, and so they look for arguments against those positions. If I can demonstrate that in fact HTM and internalism can coincide with the objectivity of morality, then this worry will be avoided.

The second reason is that doing so will strengthen the case for reasons internalism against reasons externalism specifically. Externalism is not always compatible with nonobjective theories of morality (whatever theory we want to slot in as our rejection of objectivity in morality), but typically has no problem with objective morality. If reasons are external then the natural thing to think is that whatever reasons there are is a matter of fact, not something that is relative to individuals or outright subjective. In contrast, internalism is often thought to imply that morality is likely not objective, and has trouble accounting for objective morality. It seems that morality must either be objective or not objective. If it should turn out that internalism is compatible with both of those options, whereas externalism is more generally compatible with only one of them, then this is a reason to favor internalism over externalism. Most likely some externalist theories can be made compatible with nonobjective morality, in which case making internalism compatible with objective morality is simply to level the playing field between the two theories. In either case, it can be seen that arguing for internalism being compatible with objective morality is a benefit for internalism.

---

Additionally, this tension between the three theories is sometimes taken to be a problem because the HTM and internalism appear to be the more commonsensical positions, particularly the HTM which has and does enjoy wide support. Likewise, objective morality is probably commonsensical. If these three are in tension, then our common sense thinking about morality and practical reason are in tension, and that is unfortunate, and a situation we ought to try to resolve. In this chapter I will discuss some possible solutions to this issue, and argue that none of these are satisfactory. This will lead into the next, and final, chapter, where I will discuss my solution to this problem. Many of the positions I will discuss in this chapter will be inspired by particular thinkers, however most of them do not fully agree with the conclusions I have previously established, so note that these potential solutions to this problem are based on their theories, but do not necessarily represent how they would treat this issue (since it may often be, for them, not a problem at all due to other positions they hold).

A note here: it's possible for either the HTM or internalism by themselves to be compatible with moral objectivity. That is, it's much less problematic to hold the HTM along with externalism and moral objectivity, or internalism along with moral objectivity and not the HTM, than to combine all three. There are numerous ways these positions could be argued for; I'll just outline a couple of simple ways that this compatibility could be achieved. If we hold the HTM along with reasons externalism, for example, then we could quite easily accept that desires and not beliefs motivate, without worrying whether morality motivates. We would still have reasons to be moral even if we weren't and couldn't be motivated by those reasons, as that they were not motivating would be completely irrelevant to their status as reasons (although some might object to the idea that reasons to be moral would not motivate even if properly grasped). Morality could still be perfectly objective in that case, and Foot gives an example of such a position in

Natural Goodness. Likewise, if we don’t think that beliefs are incapable of motivation, but still wanted internalism, we could simply say that beliefs about moral considerations or about the reasons we have which are produced by them, motivate. In the first case, morality would still be objective since we could have reasons to be moral which held no matter how we are individually constituted. In the second, it would matter what beliefs we had, but a strong case can be made for a claim that we could all come to have the same beliefs if we reasoned correctly, and then we would come to all have the same reasons, being based on the same beliefs. So either of these possible combinations, then, can accept moral objectivity without issue. The conflict with objective morality only arises when the HTM and internalism are both held in combination with it.

2. Similar Desires, Similar Reasons

One solution to this issue is to argue that we all (or virtually all)\(^{128}\) have relevantly similar desires such that we all end up with the same reasons to be moral and do moral things. This would permit us to retain some form of the HTM and internalism while still allowing the possibility of moral objectivity. Obviously people’s complete set of desires do generally differ greatly, but perhaps there is some desire or subset of desires such that in the end a certain reason or set of reasons can end up applying to everybody. There are two different approaches to this possible solution. One possibility is that we all share some desire(s) or would all share some desire(s) if we were rational, and on these desires we can ground reasons which are then reasons for everybody. Alternatively, it could be claimed that there are certain reasons such that whatever a person’s desires happen to be, these reasons will arise from at least some of those desires. That is, this reason, or these reasons, can be based on an extremely

\(^{128}\) It wouldn’t be an objection to this claim that somebody suffering from mental illness, for example, might not share these desires. Take it as the claim that all right functioning people share a certain set of desires.
broad set of desires, broad enough that everybody would meet the criterion for having these reasons because everybody would have enough of the required desires. This second approach would allow us to accept that people do have fundamentally different desires, without having to argue that they are in some sense reducible to a similar set, and yet still argue that people all end up with the same reasons to be moral.

One proponent of the first approach is Michael Smith. He believes that perfectly rational agents will come to have a very similar or even identical set of desires, therefore, they could end up with much the same reasons, on which we can base objective morality. Smith is not, we should note, a reasons internalist, and so although he is solving a version of the problem I am concerned with, it is not quite my own version. His position is that we have reason to do what the perfectly rational version of me would advise me to do (not what the perfectly rational version of me would do in my circumstances), because we always have reason to do what the perfectly rational version of me would do. Although I have a few general doubts about this account, it is not what I am concerned with here, instead I am concerned only with his claim that if people were perfectly rational, we would see a convergence in their desires. He argues that this is required in order to establish objective moral reasons, but notes that "we have therefore said nothing to suggest that, substantively, there are any such reasons...Substantive convergence is always assumed to be available, in so far as we converse and argue about the reasons that we have. But whether or not this assumption is true is always sub judice; something to be discovered by the outcome of those very conversations and arguments." His thinking appears to be, then, we seem to accept that reasons are not individual, in other words if we accept the other

theses such as that morality is reason-providing\textsuperscript{131}, that desires give rise to the reasons we have, and that morality is objective, then we are required to believe that rational agents' desires would converge, because that is required for there to be moral reasons which apply to everyone. As I will argue in the next chapter, I am not yet convinced that morality need be necessarily be reason-providing, so I am not compelled to accept this reasoning, but that is not my immediately concern here. While detailed analysis of Smith's position is too large a project to take up here, I will consider whether or not the claim that desires would converge amongst perfectly rational agents is compelling.

I find the claim that rational agents would end up having the same desires rather implausible. It's true that the average person typically has some, perhaps many, desires in common with other people – although it depends exactly on what you take their desires to be, that is, how specific you take the objects of desires to be. The more general you make a desire the easier it is to claim that other people share it, for example, you could claim that everybody has a desire to "do the kind of things they desire to do". I think it's pretty obvious that this isn't a proper desire, rather people have various specific desires to do various specific things, and a very general desire like that is essentially meaningless. However, most people do not share all of their desires with other people, nor are there any desires that are obviously shared by everyone. The only desire that it's really plausible to think is very widespread has is a desire to be alive, and even that is not shared by everybody. It's obvious that people do not actually share the same set of desires: serial killers have wildly different desires than most people, for example. People in other cultures have often had very different desires from us, including very different views on the desire to live. It's hard for me to conceive of caring about strange rules of honor to the extent that I would kill myself, and yet some people have apparently had such desires.

\textsuperscript{131} He appears to find this obviously true: 1995, p 6.
However, the claim was that perfectly rational agents would have a shared set of desires, not that actual agents would, so none of this is yet a proper reply to this claim. Given that people don't share the same set of desires, I don't see why we should grant this supposition that we would if only we were rational: what is the basis for that conclusion? We don't have much experience that supports such a conclusion. Say we take philosophers to be examples of (somewhat) more rational people than the average person. Certainly they have a desire to study philosophy which is shared, but beyond that, they seem to differ in their desires just as much as any other members of their culture would do. If the desire to study philosophy is shared, then presumably that could be one of the rational desires, that is, a desire that the perfectly rational versions of philosophers would have, since we would think that reasonably rational people ought to have at least a few of the desires that the perfectly rational versions of them would have, and since it is the desire they all share it may be what picks them out as being reasonably rational agents. However, that philosophers have a desire to study philosophy is no more reason to think them rational than mountain climbers' desire to climb mountains is a reason to think them rational. It is not clear why we should think that the desire to study philosophy is rational, nor particularly that to be rational everyone ought to have that desire.\textsuperscript{132} If not that desire which, then, are the shared desires we would expect between rational people? There don't seem to be compelling candidates for such desires.

Smith argues that we modify our desires in order to produce the most systematically justifiable set of desires possible, and he thinks in fact that this is the most important way new desires are gained,\textsuperscript{133} and that such reasoning will lead to convergence: “everyone can reason themselves to the same desires if they engage in a

\textsuperscript{132} Unless we're Platonists of course.
\textsuperscript{133} 1995, p 158-9.
process of systematic justification of their desires. Which desire I would end up with, after engaging in such a process, thus in no way depends on what my actual desires are to begin with.\textsuperscript{134} But most experience indicates that our desires depend in part on our upbringing (by which I include our family, our culture, etc) and especially our personality, and what experiences we have, and much less, if anything at all, to do with how rational we are and any deliberation about our desires. We can predict what kinds of desires people are likely to have much more accurately by considering the circumstances of their upbringing than we can how much time they've spent rationally deliberating over their desires. Most people probably don't spend too much time deliberating about which desires they have, as compared to how to satisfy them. The reason philosophers desire to study philosophy no doubt has to do with their temperament and taste, not rationality. I imagine the same is true of the vast majority of desires, and it's not clear which desires would be the desires that would be shared by perfectly rational agents. Smith does leaves room for there to be divergence due to our varying talents, attachments, and environment,\textsuperscript{135} which is rather vague. I take it he thinks something along the lines of "if I had a talent for the guitar, then the perfectly rational version of me (PRM) would wish to play the guitar", or "if I had a child, then the PRM would wish to care for that child", and so on. If this is the sort of thing he has in mind, then it is still not clear to me why the PRM would gain a desire to play the guitar because I happened to be good at it, and so on. That does not seem to be entailed by merely having a talent for the guitar. He also allows that preferences may still vary between perfectly rational agents, but says that which of these are allowed to vary depends on "whether fully rational agents would all converge on a desire which makes the preferences relevant to her choice."\textsuperscript{136} Since this distinction presupposes the issue

\textsuperscript{134} 1995, p 173.
\textsuperscript{135} 1995, p 173.
\textsuperscript{136} 1995, p 171.
I am currently doubting right now, since the standard is what desires are converged upon, it cannot be used to help settle this issue. Perhaps we could say that only those preferences which bear in some way on moral issues (if we could isolate such desires) must converge. That too, however, is dubious. Smith claims that the fact that there is much substantive moral agreement lends credence to the idea that perfectly rational agents will have a convergence in their desires. However, I think that this moral agreement is rather overstated, and I'm not sure we can conclude much from any supposed increased in moral agreement as time progresses, which Smith calls "moral progress" produced through reasoning. Perhaps it is produced through reasoning, or perhaps the convergence is actually coming about due to the increasingly homogenized global culture: is it really a surprise that European values are dominating the world when Europe and its descendents are generally culturally dominating the world?

It's not clear how to settle this debate. I think that we would have to somehow determine who the rational people were without considering their desires, and then look at the desires they share, in order to establish what kinds of desires the perfectly rational versions of people would have and what the candidates for the desires we would converge upon could be. It would have to be without considering their desires because otherwise we would be just declaring that the rational people, as determined by their desires, have the same desires as each other. Clearly that wouldn't establish that we are picking out the desires which rational people would converge on in virtue of their being rational, just which desires we might initially think are rational. Smith doesn't give a full account of how he would determine rationality, so I can't use this method to fully rule out this claim of convergence of desires, however, the burden of proof is clearly on the claim that we would all have a similar set of desires if we were all sufficiently rational, and as stated Smith does not fully provide such a proof. Until such

---

137 1995, p 188.
proof is provided, I reject this claim as implausible, given the current state of our desires.

3. **Constitutivism: Velleman**

Might there be desires which we all must have in virtue of being human, or perhaps in virtue of being agents? Velleman suggests that we all have a desire to make sense to ourselves, although this is not a desire that we typically feel.\(^{138}\) That latter is not a problem, Velleman thinks, for after all the desire to avoid pain, he claims, often influences our actions without us being fully aware of it, and so too it is with this desire to make sense to ourselves. If we don't make sense to ourselves in this way, he suggests, we become mere spectators watching the actions that are being done by us, and clearly that is not to be an agent. A reason for Velleman is something which makes our actions intelligible to ourselves, and so satisfies this desire.

If everybody really does have a desire to be intelligible to themselves, then everybody could be motivated by considerations that help satisfy that desire, and thus this picture would be compatible with the HTM and reasons internalism, and possibly objective morality, if we could found a system of objective morality on such a desire. It's difficult to imagine that many kinds of objective morality could have all their reasons (at least partially) motivated by a desire to make ourselves intelligible to ourselves, but some kind of Kantian based theory seems at least conceivable; assuming non-contradiction is required for our actions to be intelligible to ourselves, a version of the categorical imperative may go through. That is, if we find ourselves to be acting in contradictory ways, it's plausible to think that we may find this to be unintelligible in some sense, and so our desire to understand ourselves may lead us to be motivated to act in ways which are not self-contradictory. Kant based his account of morality on the

\(^{138}\) 2007, p 37.
concept of non-contradiction, holding that actions which are contradictory, either in themselves by having a self-defeating rule of action, or which contradict other rules of action that we must hold, are wrong. Many of his resulting moral principles could possibly be motivated by a desire to make sense to oneself: for example, if we wish others to help us when we need it and so take it as a principle that people should help each other, but we don't help others who need it, this may be an unintelligible course of action that frustrates that desire. Obviously the resulting morality won't be Kant's, since this one would be motivated (at least partly, other concurrent motivations would be possible too) by that desire, but the end result would be a criterion for making moral judgments based on a desire to be intelligible to ourselves. There may be other ways of working such a desire into accounts of moral reasons too, so this position of relating moral reasons to a desire to be intelligible to ourselves may be compatible with some kind of objective morality and the HTM and reasons internalism.

This position is an example of the claim that we all share a certain desire on which we can base reasons, in this case, the desire to make sense to ourselves. So rather than seeking to establish that we all share, or could share, the same sets of desires, we could take the approach that a particular desire is shared, and is enough to found moral reasons. The problem is, even if most people really do have such a desire, it simply seems unlikely that this desire really motivates people much of the time. Most of the time it seems like we are not really aware of such a desire, do not feel it, and it does not play into whatever deliberation (if any) we experience before acting. This isn't a problem for Velleman's theory since he is happy to accept that desires can motivate without being felt, but I hold that desires must be felt in order to motivate. Some people are quite happy to think and do contradictory things, for example, which indicates that the desire to make sense of themselves probably isn't what is motivating

\[139\] 2007, p 37.
them when they act in contradictory ways. Perhaps they can later reconcile that behavior in some way (they may simply accept that they are the kind of person who does contradictory things, for example), but there’s little indication that they were motivated by such a desire in the first place. If they do have it, and if it is possible for people to make sense of themselves as being hypocritical or self-contradictory, then this desire is apparently quite easily satisfied, and in that case it’s difficult to see how any kind of moral reasons could be produced by such a desire. I could satisfy my desire and justify not helping others while expecting help myself on the basis that “well this case is different than those others I condemned”. So it looks like the objective morality sketched above could not get the reasons it requires with that desire, since it would not rule out obviously hypocritical actions after all.

Velleman explains that this desire is almost never going to produce the strongest reason why we do something.¹⁴⁰ He thinks that it need only be a kind of tie breaker, that inclines us to choose the actions which make ourselves intelligible over those which don’t when everything else is more or less equal. If he’s right about this, then this is another problem for this potential solution to my problem, as this would not be enough to found some kind of objective morality, for obviously if this desire is never the primary motivation, then it may so happen that it is always outweighed by other reasons for certain people. If that’s the case then we’re thrown back to the original problem, that sometimes people may desire something other than what is moral, and always desire those things more strongly than to be intelligible and thus whatever morality is based on that desire. Then they would not desire the moral thing, and thus have no reason to do the moral thing, as above.

¹⁴⁰ 2007, p 46.
4. Social Basis of Shared Desires

There is another way that people might end up with similar desires, or with desires that all lead to the same reasons, not based on our nature as agents, but through our society. Certainly the way we are raised has a bearing on the desires we have, because although we have many (perhaps the majority) desires which not widely shared among all members of our societies, there are nonetheless other desires which plausibly look to have been caused by our upbringing. People's desires concerning the requirements of justice seem to be responsive to our upbringings, for example, as do desires concerning honor and such. What people desire is often a result of their place in society: a king would desire a different kind of respect from others than a peasant would. To an extent then, we have some desires that are caused by our society, and presumably then these desires would be shared among members of that society. Since these desires are shared, it may be possible to establish reasons based on these desires which are sufficient to ground an objective form of morality about which it would be possible to have real disagreements and so on.

The obvious problem with such an account is that most of those who want objective morality mean they want a morality which transcends cultural boundaries. Since any reasons provided in this way are limited to those who possess the relevant desires, which may in turn be limited to members of their society, we may end up with a morality specific to a culture, in other words cultural relativism, and cultural relativism is not objective morality. By aiming for "objective morality" we usually mean to aim for moral truths which were always true throughout all cultures in all periods of history. This account, even if it could be made to work, would not provide that, and so is unacceptable as it stands. However, if we could argue that certain desires would need to be inculcated by any society no matter its particular details, then perhaps we would
have a more promising basis to begin with for grounding objective morality. Such a solution might be inspired by Hume's account of the artificial virtues, those virtues which are acquired because they make living in society possible, or the similar argument put forward by Rachels.141 A desire which makes living in a society possible at all, could ground a reason that is common to all societies and people, since all societies clearly need to inculcate in their citizens whatever desires are prerequisites for their continuing to exist.142 Rachels notes that for society to function we need basic prohibitions on lying, for example, so that we can actually communicate meaningfully with other members of our society – if everybody lied all the time, we couldn't possibly coordinate our efforts as is required to live in a society. We still might worry that, insofar as this reason would be limited to being in a society, it is not truly objective as we wish. I don't find this very troubling, for morality certainly in large part has to do with how we treat other people, and it's not counter-intuitive to think that if we never had any dealings with other people, if we didn't live in a society with them, then our moral demands might be very different,143 without this undermining moral objectivity. We cannot have duties towards others, or be required to treat them in particular ways, if we never encounter them at all. Therefore objective morality only arising within society is compatible with most views of objective morality.144

Beyond the most basic moral concepts like prohibitions on lying and murder that Rachels suggests, Hume suggests that concepts like justice and property rights, which involve the claims others can make on us, are also required for a society to

142 I don’t necessarily mean that we consciously set up our societies to do this, but rather that it will tend to happen naturally.
143 For example, as Rousseau suggests in A Discourse on Inequality. Moral demands can be objective while still being dependant on certain factors: for example, if humans were immortal, then clearly prohibitions on murder would be unnecessary.
144 Versions of Kantian ethics, consequentialism, and virtue ethics could all exist and provide obligations for a solitary person, but they would undoubtedly be much diminished, and arguably turned into forms of prudence rather than true morality.
function. The idea of respecting, in some sense, the claims of others upon us is surely central to most conceptions of morality. It seems obvious that such ideas are also needed for societies to exist, and so it’s plausible to think that society would inculcate desires relating to these things into its members. Desires such as to get what you are owed, sympathizing with others and thus wanting to help them, among others, can be explained in this way. Whether or not such desires would give us enough material to base the reasons of an objective morality on depends on what exactly we’re after in such a theory, but the situation appears promising at first. However, there are at least two major concerns with this approach.

Firstly, what exactly we are meant to respect in the “claims of others” is very vague, and is probably filled in by the society itself. That is to say, what we expect to get from other people, and what claims we make on them in the first place, depends in large part on the societies we find ourselves in. A modern American, for example, views herself as having a valid claim to have her rights respected even by those she might consider her enemies. That is not to say that she wouldn’t expect such people to do wrong by her and ignore her claim, but rather that she holds that violating her rights is always wrong, even by those whom she considers her enemies. In contrast an Ancient Greek might expect to have his enemies attack him and view this as a kind of “justice”, and in a sense think that his enemies wouldn’t have done anything wrong by attacking him, as per the idea that it is right to do evil to your enemies. Similarly other societies may have no expectation of freedom of religion, and so don’t make claims on others relating to that, whereas in others that is expected. While societies may universally produce desires to have our rights or claims respected, that doesn’t

---

145 As Polemarchus suggests in book I of the Republic.
146 In the US for example, it is expected. In an extreme contrast a 2010 survey reports that 86% of Muslims in Jordan favor the death penalty for those who leave Islam. http://www.pewglobal.org/2010/12/02/muslims-around-the-world-divided-on-hamas-and-hezbollah/
describe an actual desire, just the form of a general desire; our actual desires would be more specific, referencing the particular claim we wish to be respected, and thus variable. These desires would vary from society to society, and so to would the reasons based on them, and as such, this theory ends up still being a form of cultural relativism.

Secondly, we may doubt how many people even within a single society have these desires. No doubt the average person in a society has the various desires he or she needs in order to be a well functioning member of that society, such as to respect the various claims of other citizens, but there will always be those who don't care about laws and morality. No society will be completely successful at producing the necessary desires in all of its members, and it can survive without doing so, as long as a majority of people do have them. The problem with this is that if some people don't have the desires which give rise to the reasons which we need for objective morality, then it appears they have escaped the grasp of the morality we're trying to introduce by once again failing to have the proper desires. If that is so, then it's no longer clear that this position would really produce what was intended and wanted by the term "objective morality". It should not be something that thugs can escape simply because they don't care about it. Yet if our objective morality is based on desires inculcated by a society, this is the result that is produced. So again, this response turns out not to be satisfactory in answering the problem.

5. Hobbes

Leaving behind the claim that we have some similar set of desires, I turn now to the position that we could differ in our desires, and yet end up with similar or identical sets of reasons. This position is perhaps more compelling than the claim that we must share desires. Schroeder argues for this view, and I will discuss his ideas in the next
section; in this section I will discuss a historical take on this position, that of Hobbes’, and Gauthier’s more contemporary version of a similar position, contractarianism, where morality is based on a social contract.

It’s necessary to give a very brief sketch of the main argument of the *Leviathan* in order to talk about this position. Naturally much must be omitted in such a brief account, but it’s possible to give the broad strokes of the argument here. In the *Leviathan* Hobbes states that that which we call good is that which we desire, and conversely evil is that which repels us. ¹⁴⁷ This is about as subjective as could be, but his eventual position offers something like an objective account of morality nonetheless, because of the way the demands of morality are founded. He argues that a powerful sovereign is needed to keep people from a state of war, meaning a state lacking the security of a guarantee that we won’t be attacked by others. The point is, that even in the original state of nature before government is set up, when good and evil is solely determined by our own individual desires, we all have reasons, and in fact overwhelmingly powerful reasons, to do what we must to get out of the state of war. What we must do to get out of the state of war is essentially refrain from attacking and mistreating others, which Hobbes’ sums up as giving us the law of nature to always to do unto others as we would have them do unto us. ¹⁴⁸ Despite the fact that our desires vary greatly from person to person, we all have reason to follow this law as it will bring us out of the state of nature.

How is it that we all have reason to follow this one law, which appears to be a moral law that applies to everyone, even though we all have different desires? Although everybody does have different desires, whatever our desires are, we desire the ability to fulfill them (because we desire the means to get whatever it is that we

desire). Typically, we will need to be alive, and to be free, to fulfill our desires – by "free" I mean something quite straightforward, for example, we can't be tied up in a dungeon, or somehow otherwise prevented from pursuing our desires. A state of war is a state where there is nothing preventing other people from preying on us, therefore, it is a state where we have no guarantee from one moment to the next that we won't be placed in a position where we can no longer fulfill our desires. In fact, simply being in a state without this security will render us unable to fulfill many of our desires as we will be limited in what we can do by the need to constantly be on watch and to protect ourselves as best we can against others. In such a state there would be no point in developing anything beyond the bare necessities of life (food and shelter primarily), since it could be taken away from us at any time. We would have no leisure time to develop arts and sciences, needing to be on guard at all times, and needing to work very hard just to survive (since cooperation between people makes the bare necessities much easier to produce). Furthermore, we cannot fulfill any social desires very easily. It's difficult to approach others to talk and socialize if we think there's a good chance that they may attack us. So, whatever our desires, we will have reason to get out of the state of nature, as the state of nature very generally prevents us from satisfying our desires. How we do that is by accepting laws, limitations, on the way we treat others, on the understanding that they too will accept such limitations in the way they treat us. By accepting such limitations (and by placing a sovereign above us to enforce these limitations) we are no longer in a situation where we have no security that we won't be attacked the next minute. We will have some guarantee that we won't be killed by the next person we meet, for people waive the right to do just that. Note that it doesn't matter if this state of war never actually existed; the argument can be rephrased simply by saying that we all have reason to avoid falling out of a state of peace and into this state in the first place.
Thus, whatever your desires, you have reasons to follow the rules that lead to security and society; Hobbes claims this rule is the do unto others rule, but it could be another rule that we think more suitable. It can easily be seen how Hobbes’ basic argument could be taken and made general: if we could somehow show that morality, that acting upon moral reasons, was required for us to satisfy our desires in general, then we would all end up with reason to act morally no matter what our desires are. This is indeed what Gauthier attempts to show, by arguing that cooperation between people is only possible under a contract, which requires certain behaviors when dealing with other people, i.e. certain moral principles, which he dubs “agreed mutual constraint”.149 This cooperation is necessary in order to maximize the satisfaction of our desires.150 Gauthier leaves the details of this argument vague, presumably because he is trying to argue that whatever our desires are, cooperation will help to fulfill them; nor does he determine exactly what the contract would require from participants. So even if somebody doesn’t wish to survive, if Gauthier is right they will still have a reason to cooperate as cooperation puts us in a position where we can more effectively fulfill a very wide range of desires.

This isn’t too implausible a position as morality seems to be connected to what we need to do in order to be able to live well with others. At the very least, usually we will need the cooperation of others to fulfill most of our desires, if only insofar as we need them to not interfere with our satisfying of them. If people attack us, kill us, steal our property, etc, then clearly it is difficult for us to satisfy our desires even if we don’t directly need the contributions of others to obtain the ends we are after. Therefore it looks like we will all have reasons to get along with other people to some extent. If the only way to do this is to refrain from interfering with others, then it looks like we will

---

150 1986, p 103.
have reason to do so based on our desires in general. In other words, virtually everybody, perhaps anyone at all, will have desires such that they have such reasons as could produce a certain kind of objective morality. So far, this proposed solution may seem strong; the main problem with it is whether or not the reasons that it ultimately produces constitute a kind of morality.

However, there are problems that lead me to reject this approach to my problem. One problem with a position like Hobbes’ or one inspired by it like Gauthier’s is that the resulting morality may not be something that we really want or mean when we express a desire for morality to be objective. For Hobbes, for example, it seems like morality collapses into prudence. We treat others according to certain rules because we want them to do the same to us for the sake of satisfying our desires, in other words, because it is in our interest that we do so. It is perhaps commonsensical that whatever we mean by “morality”, we mean something that is directly opposed to or at least separate from prudence. Prichard, for example, argues that this kind of response fails completely to explain why we ought to be moral. He argues that to provide an answer in terms of prudence to the question why be moral “only makes us want to keep” moral principles, it doesn’t really explain why we should.\(^\text{151}\) We might also think that the moral rules are those rules of conduct which precisely are not those rules of prudence,\(^\text{152}\) so to argue for a system which is just a form of prudence is a failure to provide a theory of morality at all.\(^\text{153}\) Gauthier’s approach is perhaps somewhat stronger against this objection, as he links morality and rationality more closely (a move which is appealing to many), but since what is rational for him looks a lot like what we normally call prudence in that he thinks it is rational to maximize the satisfaction of our

\(^{152}\) Williams, 1976, p 250 for example, although he calls it “egoism”.
\(^{153}\) This no criticism of Hobbes, as Hobbes was not arguing for the status of morality directly, but rather arguing for a political system. So he is not failing to provide a proper theory of morality, so much as not engaging in this debate at all.
desires, this objection could, recast, still be raised against him. We might also worry that such approaches appear to make morality dependent on our membership in a society. Those who do not belong to a society have no reason to worry about what rules are vital for cooperative behavior, and in other words, no reason to treat other people morally. I am not convinced that this is a critical worry, since it seems likely that morality fundamentally has to do with relationships between people, but there are some moral theories, such as some lines of virtue ethics, where this would be a concern.

Practically speaking, this kind of approach means that if others refuse to obey these rules, we have no reason not to discard restrictions on our actions towards them, since that would be the prudent thing for us to do; but we usually think that morality requires certain duties from us even towards those who do not treat us morally. It's true that emotionally a lot of people desire extreme punishments for wrongdoers, but it is surely not moral to treat a habitual thief in just any old way they feel like simply because she was no longer part of their contract – or for the sovereign or other such empowered party such as the police to do so, even if such treatment was enshrined in law (laws can be immoral). Morality is not purely limited to whoever is still part of our social contract. Furthermore, it seems like if we can get away with breaking the laws of nature, meaning that we don’t suffer any bad consequences, then for Hobbes this is the right thing to do, as this would be the prudent thing to do, the thing that most allows us to maximize our desires. The free rider, the person who benefits from others’ placing restrictions on themselves without himself recognizing these restrictions, is prudentially very well off provided he can get away with it. Hobbes argues that we can never be certain that we will get away with it and so it cannot be prudent to act in such a way, but the fact remains that if we could be certain that we would, there would be no objection here. It

---

155 Leviathan, Chapter XV.
is not that far-fetched to think that there could be such situations, as many people do get away with crimes and small instances of free riding. Gauthier takes up this problem for Hobbes, and argues that the proper response to such an objection is that it is rational to have a disposition to be a constrained maximizer, in other words, to reap the benefits of cooperation you cannot simply approach each circumstance and decide whether it is rational to cooperate or not cooperate each time. Rather, you must adopt a disposition of following the rules of morality, and not take advantage of others.\textsuperscript{156}

Should those you deal with know that you are the kind of person who considers whether or not to keep to his agreements each time you make them, and base this decision on what is most prudent for you, then of course they are not likely to willingly enter into such agreements with you in the first place. Thus, the free rider does not gain any of the benefits of cooperation, and it is not rational to attempt to free ride. Note, however, that this kind of response requires that others are aware of the kind of person that the free rider is, and Gauthier states that cooperation will not be a maximizing strategy if we are not able to detect such people.\textsuperscript{157} It is no small matter to state that people need to be able to detect those who are not just (for Gauthier being just means having a disposition to be a constrained maximizer\textsuperscript{158}), for we don't seem to be extremely good at figuring out people's true characters. If people can't do this, then they ought, rationally, not to be constrained maximizers, on Gauthier's view. It seems likely then that, contingently, very few people will have reason to be moral on Gauthier's theory, since hardly anyone will be able to catch every free rider they encounter. Furthermore, if the free rider could get away with her behavior (as such people surely often do), then it still appears to be rational for her to act in this way – it may be rational to agree not to do so, but that doesn't mean that it is always rational (in Gauthier's sense) to continue

\textsuperscript{156} 1986, p 162.
\textsuperscript{157} 1986, p 181.
\textsuperscript{158} 1986, p 157.
to do so if others won't detect what she's done. In other words, depending on our
desires and circumstances, we may once again escape the demands of morality on this
kind of approach too.

6. Schroeder

Schroeder also argues that there may be reasons such that virtually any desire
we have could give rise to those reasons: "there might be reasons that can be
explained by any possible desire."\(^{159}\) Such reasons he describes as overdetermined:
many of our desires are sufficient to ground such reasons, so they have more than one
sufficient cause. One such reason he suggests is that we all have a desire based
reason to have correct beliefs.\(^{160}\) Schroeder notes that whatever our desires are, it's
difficult to satisfy them unless we have correct beliefs about the state of the world and
how we can interact with it to fulfill our desires. I cannot fulfill my desire to drink unless I
have correct beliefs about what will quench my thirst and where to get it from, for
example. If I believe that sand will quench my thirst, then clearly this desire will be
frustrated. The actual object of the desire is not referenced here, so the reason to have
correct beliefs is not dependent on which precise desires a particular person has; just
so long as she has any desires, she will almost certainly have a reason to have correct
beliefs.

One immediately obvious objection to this account is that virtually everybody
has incorrect beliefs, and yet they still manage to satisfy their desires well enough, so
perhaps this reason is not really grounded on our desires in general, just that certain
beliefs and are needed for certain desires. Schroeder admits that it's perhaps possible
to fulfill certain particular desires while having a certain set of incorrect beliefs; if

\(^{160}\) 2007, p 114.
perhaps I believe that water tricks my body into thinking it's not thirsty, rather than that water actually quenches my thirst, I will still be able to satisfy my desire to drink even though I believe something false.\textsuperscript{161} Furthermore, there will undoubtedly be beliefs which have little bearing on our desires, such as beliefs about places very distant in time or space from us, e.g. about the Ancient Egyptians or distant galaxies.\textsuperscript{162} It surely doesn't frustrate any of my desires if I have an incorrect belief about the diameter of some galaxy. However, it's quite plausible to think that having many incorrect beliefs would pose an obstacle to satisfying your desires, even if it so happened that some could still be satisfied, and some false beliefs will be more problematic than others (a belief that you can fly, for example, is probably going to end up frustrating many desires). Clearly at least some false beliefs hinder the satisfaction of desires at least some of the time. Furthermore, given that our set of desires typically requires a large number of correct beliefs to be satisfied, and given therefore we all have reason to have such correct beliefs, it surely follows that we all have reason to have the correct belief forming procedure. Since we don't know what beliefs we will need to satisfy any given future desire, and as we don't know what conditions we will find ourselves in, we all have reason to have correct beliefs, without knowing specifically what beliefs will be required. Therefore, whatever our desires, we all have reason to have the ability to form correct beliefs.

So it's possible to have at least this one overdetermined reason, no matter what your desires are. That shows that overdetermined reasons are possible, but perhaps this is the only such reason; it's indeed hard to see what others there could be, aside from the Hobbesian approach already discussed. Without demonstrating that there are others, we cannot simply assume there are and move on from there, and assuming that

\textsuperscript{161} 2007, p 113-4.
\textsuperscript{162} 2007, p 114.
it is the only one, would it be satisfactory for obtaining objective morality? Obviously that latter depends on what kind of objective morality somebody is after. There are some who think that morality consists of some kind of normative facts. For such a position, it is quite likely that a reason to have correct beliefs, or a reason to have the ability to form correct beliefs, will be related to their moral systems in an essential way. To be moral on Parfit's account (which I have already objected to), for example, you need to have correct beliefs about what non-natural normative facts there are; a reason to have correct beliefs at least about those kinds of facts is, in a sense, at the foundation of his morality. However, for most theories of objective morality it seems obvious that a reason to have correct beliefs will not be enough to give everybody reason to be moral.

If Schroeder's account means that reasons are based on desire satisfaction, need it be viewed as prudential, as a Hobbesian account might be viewed, and thus as nonmoral? It will depend on what exactly kind of desire and reason we would be dealing with, and that is difficult to determine when I'm not convinced that one can be provided. However, it is not convincing to simply say that the fact that a reason is "desire satisfying" means that it cannot count as moral at all, as certainly not all of our desires are for egotistic ends – some are surely for moral ends, like that somebody's life go well. To claim that satisfying such desires could still not be moral, would be to hold that the HTM just is incompatible with objective morality on principle, which, while no doubt a possible position (like Kant's), is not one I would argue for. In the next chapter I will give a position which shows that the HTM is compatible with objective morality.

The main difficulty with Schroeder's account is in extending it beyond a very few reasons that could be based on almost any desire, which means that there are only a limited number of reasons that we could all be said to share in order to get an
objective, reason-based morality. A reason to have correct beliefs is the most promising, followed perhaps by a reason to stay alive. This latter is already more limited than the former, as some people either outright desire to die, or have desires which require them to die to be satisfied, for example a concern about one’s posthumous legacy which presupposes immediate death (say if somebody wants to become a martyr for a cause), and thus they have no reason to stay alive. Beyond those two, it becomes unclear what overdetermined reason could be the basis for an objective morality. Perhaps we could say that everybody has a reason to take the means to their ends, but it’s unlikely we would be able to found morality on just that. In the end, I just don’t think there are enough reasons that we all could share for Schroder’s suggestion to work in grounding a system of morality.

7. **Constitutivism: Korsgaard**

A rather different approach that would result in our having the same reasons whatever our desires is proposed by Korsgaard. She believes that we have to endorse reasons to count as agents, and that by endorsing certain reasons, we construct identities for ourselves: however that element of her view does not concern me here. Here I will not discuss most of her highly complex position, limiting myself to the aspect of it that could be most directly useful in solving my problem. Note that Korsgaard, although she does agree with Williams’ reasons internalism as I noted last chapter, doesn’t agree with the HTM, and so she would not be directly concerned with solving the problem I am considering here, but it is possible to modify her view to produce a possible solution to this issue. The considerations we endorse, those taken as reasons, are chosen from those that motivate us, so this is compatible with reasons internalism, and this view could be combined with the HTM, even if Korsgaard herself does not do so, as the

---

163 It would end up being something very similar to a Hobbesian view even if we could.  
164 E.g. 2008, p 32-3.
broad strokes of this position so stated take no position on what produces the motivation in the first place. It would thus be possible to say that we can only choose from among our desires, and endorse certain of these. What's most interesting for me here is that she further argues that we automatically endorse the reasons of others as our natural reaction to them: "responding to another's reasons as normative is the default position." It's quite possible that such respect for others could be the foundation of some system of objective morality, again perhaps a somewhat Kantian conception, one based on treating others as morally important. So the general features of her theory might provide a potential avenue for a combination of these views with objective morality, since whatever our desires are, on her account, we endorse other people's reasons, and so although our own reasons could be based on our desires, we would still have reasons to treat others in certain ways due to the reasons that others give us.

The main problem I identify with such a response is that it's not clear that we do (or must) endorse the reasons of others, that is, why we must share them, regardless of our own desires. Korsgaard claims that this is the natural way to view them, and that it is difficult to do otherwise – "it's nearly as hard to be bad as it is to be good" – but this claim is not particularly convincing. It might be difficult for most people to be actively cruel, but to be bad in the sense of negligently indifferent to the reasons of others or callous seems to me to not require much effort at all. Many people don't have much trouble in ignoring the reasons of others: consider the average criminal who, while probably not actively wanting to do bad (in the sense that he probably doesn't think "today I will be evil"), simply doesn't concern himself with the reasons of those whom he is preying upon. The burglar often just doesn't bother to think about the reasons that

---

weigh on him; rather than having to actively dismiss reasons he accepts, he simply doesn't notice any reasons there in the first place. So he certainly isn't endorsing the reasons that other people have for him not to rob them.

Korsgaard argues that it's obvious that reasons are shared, that is to say, that this is our default position. But if they actually aren't shared, then we're left with what Korsgaard calls private reasons which can and do vary between people, and are no closer to finding how to provide reasons that are compatible with objective morality. Korsgaard offers the following argument in favor of this conception of reasons: she suggests that interpersonal relationships would be very difficult if we did not share the reasons of other. Not "difficult" in the sense that we ought to do this for our benefit, but rather in the sense that we couldn't really have interpersonal relationships, and thus the way our relationships with other people currently go suggests that we do currently treat our reasons as shared. She gives the example of a student attempting to make an appointment with his professor: if the professor viewed the student's reasons as merely individual to him, it would be difficult to see why the professor wouldn't just tell the student to skip class in order to set the appointment at the time that best suited her, rather than seeking the time that best suited them both. Korsgaard claims that if we did not share our reasons with other people, we would view the reasons of others as tools or obstacles, and so could and would use force or tricks to get rid of them.

This understates the various private reasons we have to not treat people in this way, such as, for example, that getting ahead in society generally requires the cooperation of other people, and that means treating them sufficiently well that they are willing to treat you equally well. We have, then, our own reasons for considering the reasons of others when acting, but we are still acting on the private reasons we have,
not sharing in the reasons of others. Korsgaard does consider this possibility briefly, and raises the free-rider problem: why, she asks, couldn't you respect other people's reasons just enough to get by, or only pretend that you do and then go behind their backs?\textsuperscript{169} If people did that all the time to each other, then interpersonal relationships would once again be difficult. The problem with this response, it seems to me, is that many people do just exactly that at least some of the time, and interpersonal relationships are clearly still possible. People probably don't usually do this, because you can't always know when you can get away with it, and because as I've previously said, most people have a pretty strong desire to be moral which has been inculcated into them from childhood, and because, of course, they have numerous private reasons to want effective cooperation among different people to continue and so don't want to free ride to the extent that they could endanger such cooperation. But we do sometimes see people acting in exactly the kind of ways that we'd see them acting if they weren't treating reasons as shared. There are those who are willing to throw their trash into a river if they don't face a significant chance of being caught, even though clearly everybody else around them has reasons to prefer they not do that, and even knowing and accepting the bad effects this could have. We can also see this when people take more than their share of something, or dump work onto someone else, when they think they can get away with it, and so on. If people sometimes respect the reasons of others, and sometimes don't, this suggest to me that it is not the case that we always just do treat reasons as shared, but rather that we usually, but not always, have some private reason to treat them as such (self-interest, empathy, etc), and when this reason is not present, we don't treat other people's reasons with the respect Korsgaard demands. Perhaps we could argue that we usually do treat them as shared, and it's just that sometimes through, say, weakness of will, we fail to do so. That's also

\textsuperscript{169} 2009, p 201.
possible, but I don't see any reason to prefer that explanation over the alternative. The person dumping her trash into the river probably isn't conflicted about doing this, and needn't feel that her instincts have won out over her better judgment. It's just easier and she doesn't think about or doesn't care about how it affects others.

Also note that when this position is combined with the HTM, it appears that people would need to have some desire which leads them to take the reasons of others as shared (if this is intentional), and given that people don't always take them as such, it is unclear whether we have any such desire that motivates us with any regularity. Korsgaard does not, of course, agree with the HTM, so this is not a criticism of her position, but rather a difficulty with this response for my purposes. If somebody has no desire to recognize the reasons of others, it's hard to see how she would do so, not being motivated to do so, and in that case this person would once again escape the grasp of morality.

8. Conclusion

The HTM and internalism are thought to be incompatible with objective morality because they seem to base all reasons on our desires, which leads to the conclusion that if we desire not to do the moral thing, then we must have no reason to be moral. Objective morality is not usually something which we think people can escape by merely not wanting to follow it, so this conclusion seems problematic. Regardless of whether or not you want to do what morality dictates you do, nevertheless you are still immoral if you don't do it. In this chapter I have considered some solutions for this problem, which have fallen into two main camps: firstly, the view that we all must have certain desires, which would enable us to base reasons on them that would hold for everybody. Secondly, the view that some reasons can be based on virtually any set of
desires, such that we will all end up having these reasons. I have rejected various arguments for these camps.

Note, though, that another assumption has slipped into the original argument: that objective morality must provide reasons for action for everybody, and that without a reason to be moral we cannot be criticized for not being moral. It is because of this assumption that objective morality, the HTM and reasons internalism are in tension, and it is this assumption that I reject, as I will explain in the following and final chapter.
Chapter 6: Morality and Reasons

1. Introduction

As I noted at the end of the last chapter, the incompatibility between the HTM, reasons internalism, and moral objectivity, only holds if we hold that morality is necessarily reason-giving. If we were to deny that premise, then there is no incompatibility between the three positions: if morality were not necessarily reason-giving, then the fact that sometimes somebody had no desire to be moral would mean that she had no reason to be moral, but it wouldn’t mean she had avoided moral claims altogether. Denying this is not such a simple matter, however, since it is widely held that morality is indeed reason-giving, or even consists of nothing but reasons at some level. There is, then, some resistance to this move, although arguments for this premise are given less often than it is thought to be true. I am inclined to believe that, once again, the real reason that many people think that morality is reason-giving is that they are trained from childhood to believe that they must do the moral thing, and so their default position is to think that there is always a reason to act morally. In this chapter I will argue that in fact our common sense thinking about morality is not contrary to the claim that morality needn’t be reason-giving, and that arguments for the claim fail. Therefore I will conclude that morality needn’t necessarily be reason-giving, and given this, there is no contradiction between the three positions. Most of this account will necessarily be a negative argument: the reason to think that morality is not reason-giving is to hold the three positions, two of which I’ve already argued for and the third is broadly accepted. Therefore in this chapter I will focus on objecting to arguments that claim morality must necessarily be reason-giving.

170 E.g. Scanlon in Oxford Studies in Metaethics Vol. 2, 2007, p 11. Darwall in the same volume lists some others who have claimed this, p 127, such as Williams and Gibbard.
171 E.g. Schroeder lists some influential philosophers who have held this in Slaves of the Passions 2007, p 81. These include Scanlon, Parfit, and Raz.
What motivates the view that morality must be reason-giving? One important factor appears to be the thought that we cannot say that somebody *ought* to do something unless there is a reason for them to do it, in other words, "ought" implies that we have reason. Another concern is that we don't want people to be able to escape the demands of morality purely due to their desires, and if somebody has no reason to do what is moral, then it seems like they have escaped any requirement to be moral. Related to this, objective morality requires that in moral disagreements there is actually some kind of truth of the matter; if one person was beyond the grasp of morality, they would be neither right nor wrong considered from a moral standpoint. Therefore to be objective morality must be able to apply to everyone, and it is thought that it must necessarily be universally reason-giving for that to be the case. The third sort of motivation for this thinking is that morality is somehow required to succeed as a rational agent, and since reasons are connected with rationality, morality must provide reasons.

I will discuss these three points, among others which are more abstract and less based on common sense thinking, and either reject them or show how they are still met by my theory.

2. "Ought" Being Concerned with Reasons

In large part the basis for the thought that morality must be reason-giving has to do with the word "ought". We say that somebody "ought" to do something if she has a reason to do it (more specifically perhaps if she has a suitably weighty reason to do it), and we also say somebody "ought" to do something if it is the moral thing for her to do. It is reasonable to conclude, then, that we take morality to be reason-giving. Our use of the same word in each case, however, does not entail that this is actually how we must view morality, for it may be that there are really two different senses of "ought" that are being conflated here. Note that for a realist like Parfit, who believes that morality is
based on non-natural normative facts, the way we use language would surely either coincide with those facts or not, but the way we use language would not make one of them true and the other false. In other words, even if we do use the word ought in this fashion, it doesn't mean that we are straight away entitled to think that morality must be reason providing. Arguments must still be made as to why we should view the word as being used in the same sense each time.

A famous example of the kind of view I am suggesting, that morality and "ought"s in general need not necessarily be reason-giving was argued for by Foot in "Morality as a System of Hypothetical Imperatives". In that paper she compares the rules of morality to the rules of etiquette in that, in both cases, they provide reasons for you only if they connect to certain things that you care about. In other words, both provide only hypothetical imperatives of the following sort: if you care about etiquette/morality, then you have a reason to do what they command. Obviously if you don't care about them or whatever concerns they represent, then you don't have a reason to do what they command. Somebody who doesn't care about the rules of etiquette or being rude has no particular reason to worry about whether he is using the salad fork or the dessert fork right now, and likewise, I'm arguing, somebody who doesn't care about morality has no reason to do the moral thing. Importantly though, just because somebody has no reason to care about etiquette doesn't mean that they can't be rude – in fact such a person is more likely to be judged rude than not – and likewise, just because somebody has no reason to be moral, doesn't mean that they aren't being immoral. Therefore, it's still possible to say that, morally speaking, somebody ought to do something in the sense that being moral requires it, even if they have no actual reason to do it. This I will discuss at greater length later in the chapter.

\[172\] Contained in *Virtues and Vices*, 2002. Mackie also suggests something similar, e.g. 1990, p 55-56.
One general objection to this kind of argument (not specifically to Foot), before even getting to the case for morality being reason-giving, is that the cases of etiquette and morality are not analogous, because conformity with etiquette and morality are motivated differently. Etiquette is motivated by a desire to be polite by following the rules of etiquette, that is, when we follow the rules of etiquette we are consciously trying to do so, because we desire to be polite. Eventually, no doubt, compliance simply becomes habitual, but at least the initial habit formation would have been motivated by those kinds of desires. Assuming that we are consciously aware of why we are trying to follow the rules of etiquette, it will be from a desire to follow them, or rather a desire not to be rude. In contrast, some like Smith argue that when we act morally it is not from a desire to be moral in the overall, general sense. That is, moral people are not motivated by a desire to "be moral" by following the rules of morality, but rather motivated to do particular things which are the moral things to do. Obviously there are those, like Kant, who would disagree with this claim entirely, but it's not difficult to see why we might think that this is the case. The moral person wants to help somebody who requires help, due to the desire of wanting to help this person, rather than, so it is claimed, help because that would be the right thing to do. It is claimed that to be properly moral we should concern ourselves with the good of others because it is their good, and not because it is the moral thing to do, which would be a form of rule worship. Stocker gives the example of going to visit a friend in the hospital; should we visit the hospital because visiting one's friend is the moral thing to do and we desire to be moral, our friend is in fact not likely to welcome us or think of us as a particularly good person. Instead, the reason we visit should be that we genuinely desire to visit our friend in the hospital. So where etiquette cannot be reason-giving when we lack the desire to follow etiquette, it seems that in contrast we should all, even the moral person,

---

174 In his famous paper "The Schizophrenia of Modern Ethical Theories."
lack the desire to be moral, and this doesn't affect the reasons we have to be moral, because those reasons are provided by other desires. The moral person still has reasons to be moral, which will be desires specific to the circumstances, such as that somebody needs help, rather than a desire to be moral.

However, it is surely plausible to argue that etiquette can be motivated similarly to how morality was claimed to be, via various individual motivations aimed at specific situations. For example, somebody concerned with etiquette may wish not to embarrass herself at a formal dinner, and so will conform to the rules of etiquette (say, a specific dress code), without necessarily having a general desire to do what etiquette commands, but just having a desire to do the thing it requires this time. This would be perfectly analogous to helping someone out of a desire to help them rather than a generalized desire to do good, and Foot's argument is clearly compatible with such a position. So it is not so easy to break the analogy between morality and etiquette.

Note also that this means that somebody may lack a reason "to be moral" broadly construed, yet still have reasons which amount to their having a reason to be moral. If the moral person never has a desire to be moral in general, but only specific desires for ends which are moral, and thus doesn't have a reason to "be moral" in the sense that she has no specific reason along the lines of "to conform to the standards of morality" when she lacks the desire "to conform to the standards of morality", it doesn't follow from that that she has no reason to do whatever, specifically, morality requires her to do, considered on a case by case basis. She may have a reason to help this person, to go visit her friend in the hospital, to honor her parents, and however else we wish to fill in the requirements of morality. So in every case, provided she has the right motivations, she will have a reason to do the action which conforms with morality and, in a sense, have reasons to be moral. It is analogous to how somebody could have
reasons to make a pile of rocks, even if they have no reason which tells them to "make a pile of rocks", but rather reasons to continually put rocks in one particular place: this reason or reasons amounts to a reason to make a pile of rocks. Likewise, reasons to do each individual moral thing amount to reasons to be moral.

Additionally, I do not think it is obvious at all that most, or even many, moral people really are not motivated by a desire to be moral considered generally. In many cases, I suspect that that is exactly why they consistently do the moral thing. If we were to ask people why they don't steal, I imagine some people would say that they don't steal because of the consequences if they got caught, and some that they would feel too guilty, but many would outright state that they don't steal because it is wrong. Not only that, but many would think that this is precisely the most important reason we have to not steal, and the one on which it is best to act. People often think that they must carry out their duties precisely because they are their duties, and not specifically because of whatever ends the duties are aimed at, for example. It seems likely that pacifists, vegetarians, people like that, often have no other motivation for those positions other than that they want to do the right thing. It's impossible to know for sure what motivates people in such situations, but I suspect that for at least a sizable number of people their motivation to do the right thing is precisely that: a motivation to do whatever is the right thing to do. This is especially true in cases like, for example, not breaking rules concerning "victimless" crimes (say, stealing office supplies from a large corporation which will not notice the loss). It is less likely to be the case when people have some personal connection towards the good thing to do (e.g. to save one's child from drowning). The exact role motivation should have in morality, i.e. whether or not it is, or is only, acceptable to act from such a motive of duty, I leave open. My argument does not close off either option in that question.
3. The Inescapability of Morality

Much depends on whether or not we believe that a person who has no reason to do the moral thing can still be counted as immoral; we do not, as I stated above, use "objective morality" to refer to something that somebody can escape by not having the correct desires. The demands of morality must still apply to everyone for it to count as objective, and so if that requires being reason-giving, then it must be reason-giving. I believe that this is not the case, as a person's moral status depends on whether not their behavior (or maxims of actions, character, etc) conforms to certain moral standards. If some person conforms to those standards, then she is moral, if not, then she is not moral. The comparison to etiquette makes it obvious why and how I think this would work: somebody could have no reason to care about following social rules, and yet obviously still be judged, correctly, as being rude. This makes morality a "thick" concept as Williams calls them: there is an objective component to this judgment, namely, the conformity or lack thereof to those standards identified as moral. This allows moral disagreements to have a real answer, which as I previously stated is an important component of objective morality, as whether or not somebody is conforming to certain set of standards is something that can be established even if the person in question doubts the truth or value of those standards. It may not be easy or even completely possible to determine whether somebody in particular is conforming to these standards: if, for example, conformity demands certain motivations behind actions, as Kantian ethics does, then it may be impossible for anyone else to determine if a person had the correct motives and thus has been moral. That an answer is inaccessible however, doesn't diminish the fact that there is an answer, that there is a right or wrong answer to the question of whether a given person is being moral or not. If there is such an answer, then moral disagreements are possible and real, even if we can't identify the
answer – just as there are facts of the matter about what exists beyond the part of the universe that we can observe.

Related to that, a large part of the motivation behind thinking that morality must be reason-giving no doubt has to do with the thinking that if somebody has no reason to do something, then they can't be blamed for not doing it. What is important is that by "objective morality" we mean something that people can have actual disagreements over, and something with regard to which people can be held to be making a mistake in a meaningful sense in failing to conform to its dictates. That is surely an important facet of what we mean by "objective" in this case. It might be thought, then, that people can only be blameworthy for not acting in accordance with something which they have reason to act in accordance with, and so it seems like there must always be a reason for them to act morally if we wish to criticize them for not doing so. Therefore, I have to establish that lacking a reason to do something does not mean we can't be blamed or criticized for not doing it. However, reasons are not necessary for criticism or blame. When we criticize somebody we mean to express disapproval, and also to convey that their conduct was not acceptable according to some standard. Clearly, that is still possible on this view since I hold that morality is some standard we can judge actions by, and we can judge whether actions (or people, etc.) meet these standards or not regardless of whatever reasons they have. As previously discussed, somebody could have no reason to follow the rules of etiquette and yet they can still be judged as rude or criticized for their behavior. The fact that they don't care about being rude, and don't even have a reason to care, does not change the fact that they are being rude. Likewise with morality, we could still blame somebody for not acting morally, and judge them to be immoral, even if they don't have a reason to act morally. In fact, it is likely that the reason they don't have such a reason to act morally is due to something that we might further blame in them. For example, if I don't have a desire to avoid inflicting
unnecessary pain on others, this lack is itself something that we would be likely to find blameworthy, even though its lack is why I don't have a reason to avoid causing pain in others.

Related to the above points is the objection that "ought" implies "can", that is, if somebody ought to do something, then it should be possible for them to do it. This could be a concern for my account, for obviously if somebody has no desires which would lead to a motivation to do the moral thing, either directly or via sound deliberation, then there's a sense in which they cannot do the moral thing. The motivation is not accessible for them from their motivational set. However, it seems to me that we already blame people for things which they cannot, in this sense, do, and more specifically blame them for their desires. Say somebody signs up to the military, but during a battle (and let's further stipulate that this is a just war, whatever that requires) abandons his comrades and flees the scene. I think that we might already be inclined to accept that there are some people who just are abject cowards, and that it really is impossible for them to remain under the threat of death and carry out their duties. Regardless, this person is still likely to be morally condemned for his cowardice, and a retort from him that "this is my character, it is just the way I am, and were you like me, you'd do the same" is not much of a defense. His character is what we're criticizing. Generally, we might accept that some people have some weakness of character that makes it impossible for them to do certain things, and yet not think that this excuses them from blame for not doing the things they ought to do, or from doing things they ought not to do. Somebody who is cruel to animals is often banned from owning animals, indicating that we seem to think that there is a sense in which she cannot be trusted to prevent herself from being cruel to animals, for the purpose in so banning her does not seem to be merely the punishment of being deprived from the normal pleasures of animal ownership, but rather, it is in large part for the protection of
animals. We still, of course, think that she ought not to be cruel to them: in this case then, we have an ought which does not necessarily imply a can. It is not clear that "ought" really does imply a "can" in every situation.

4. No Reasons for Desires

It may be responded here that to push the talk of blame and criticism onto desires simply raises the issue of whether or not people can be criticized for their desires, especially since there may be no reason for our desires, in turn, and since which desires we have doesn't seem to be under our direct control. We cannot simply say, however, that someone cannot be criticized for their desires if there are no reasons for them, as the very issue in question is whether we need reasons in order to be blameworthy, but my opponents don't need to beg the question in this way and can instead raise the worry about criticism in another way. Whether or not we have reason to act morally depends on what our desires are, and we don't have full control over that, it being primarily a function of our upbringing and character. Therefore there is perhaps something unacceptable in criticizing somebody for lacking certain desires, in that it seems unacceptable to blame someone for anything which was not a free choice. There are two responses to this, the first of which is that we do indeed criticize people for desires like pedophilia, cruelty, and so on, even though we don't think those desires are under their control. It's obvious that nobody chooses to be a pedophile (given the consequences for having such desires, it's incredibly implausible to think that anyone would choose to have them), but this is still a desire worthy of criticism. Typically we don't only criticize the actions which might follow from such a desire: a pedophile who never acts on his desire, or a cruel person who never acts on his desire, while perhaps praiseworthy for their self-control, are still considered to have moral failings, but surely

175 Although if we desire to do the moral thing, this could give us a reason to try and develop desires which lead to moral actions.
we don’t think that the pedophile chose to be attracted to children or that the cruel person chooses to be inclined towards cruelty. The second response is that this issue seems analogous to a certain extent to the broader question of moral luck. A great deal of morality depends on luck and things which were not a question of free choice, as Nagel among others has noted, and yet we still offer criticism about such luck-based instances of immorality (or praise for lucky moral behavior). Consider his example of somebody who lives during the Nazi regime who commits horrible actions. This person is blameworthy, even though had they merely lived in a different time they would not have committed these acts, even while still being the same person, and thus wouldn’t have been blame-worthy. There is, indeed, a problem here, as there is a tension between our thoughts that it is not appropriate to criticize somebody when what we are criticizing is due to luck; I am not denying that. What I am saying is that there is nothing special about worries over criticizing somebody for desires they did not choose to have as opposed to such analogous cases where we criticize people for external factors out of their control, such as the times in which they live (for the Nazi sympathizer, for example). It is somewhat disconcerting to blame somebody for their desires which they cannot help, but it also disconcerting to blame someone for the results of external factors which they also cannot help, and we sometimes do that. This is issue perhaps an extension of a wider problem about moral luck and not a particular failing of my theory.

I conclude, then, that our possessing a reason is not necessary in order to be criticized for failing to do something (and also that “ought” does not necessarily imply “can”). That is, just because we have a reason to act some way, or we lack a reason to act a certain way, does not mean we are not blameworthy. Furthermore, note that the reverse certainly does not hold: people can have a reason to do something, and yet still...

176 1979, p 26; Williams also wrote a paper with that title and subject, contained in 1981.
be blamed for doing it. A murderer who finds that someone is about to turn over proof of his crime to the police has a reason to kill this additional person – say the death penalty is enforced where he lives, and so doing so will quite literally save his life. Nonetheless, we still would judge this action as wrong, even though the murderer has a perfectly understandable and prudential reason for it. There is not, I think, a very close connection between the concepts of reason and criticism.

5. **Practicality**

One motivation for viewing reasons as fundamental to morality is that morality is a question of *practical* reason. Practical reason has to do with actions, and reasons have to do with actions, clearly, since they explain our actions. It might be thought that my view of morality, being something similar to a set of laws, fails to be relevantly practical in the same way that a series of reasons is. A set of standards which leaves you free to obey or not obey depending on your motivations, does not engage practical reason as immediately as reasons to act do. It would seem that the agent deliberates depending on the reasons that he has, decides what to do, and then we can judge him moral or immoral according to some (possibly external) standard, on my view. Therefore, it looks like the rules or commands of morality rest outside the actual practical reasoning process itself, being just standards we use to judge it, and so it might be thought that I have failed to give an account of morality which has to do with practical reason.

This objection holds that I have made morality external to our practical reasoning, but that presupposes that the agent is someone who does not have desires relating to morality. Should the deliberator have desires relating to morality, like the desire to do the right thing, then the standards of morality are no longer external to her deliberation, but feature as some of the considerations to be weighed when considering
how to act. Someone who has no desires relating to morality, and so does not take the laws of morality to be reasons, is someone who quite literally doesn’t care about morality or being moral. That is to say, he is probably an immoral or amoral agent, and it is surely no surprise that such an agent's practical reasoning process takes no account of morality, for that is more or less the reason he will end up not being moral. The moral agent will usually view the commands of morality as being reasons, and have a desire to be moral.

To a certain extent, claiming that morality is a question of reasons seems to be an appeal to intuitions. I am not convinced that viewing morality as based on reasons is more commonsensical than viewing morality as a set of standards. Typically, I think non-philosophically inclined people are more likely to view morality as a set of standards or commands, like the Ten Commandments for example, than as a set of reasons. There are of course reasons given to obey such standards, but these need not always be moral reason at all. For example, the threat of Hell gives us essentially prudential reasons to do as the Ten Commandments stipulate.\(^{177}\) That is not to say that there aren't intuitions we have about morality that perhaps imply a reason-giving view of it, but simply to say that neither one nor the other has a stronger case on this front. In fact, I think most people's intuitions head in the direction that there is not a large gap between being a standard and being a reason. Most people view the fact that there is a law not to do something as being basically equivalent to there being a reason not to do it. It's obvious why they think like this: a reason tells you what action to do, and so does a law or standard. It's true that reasons are typically more personal to your individual circumstances, whereas such standards are typically more impersonal, but then such

\(^{177}\) I don’t mean to imply that this is the only reason religious people follow the dictates of their religion. There are many different reasons why people do so, many of which have nothing to do with any perceived self-interest. But it seems likely to me, through my own personal experience, that a good number of them do have a fundamentally prudential view of it.
impersonality is implied by the requirement to be objective. Remember that objective
morality cannot be personal to you, but must be the same across everyone. There's
also a sense in which having an overriding reason is a "command" to do a certain thing
in a certain situation. I don't wish to push this similarity argument particularly far, since
intuitions are not particularly strong grounds for a conclusion; it may or may not be that
reasons are actually similar to law-like commands in any significant sense. What I am
defending against is the argument that the intuitions are on the opposing side of this
debate, that morality is intuitively concerned with reasons. I think that the concepts are
not entirely separate in the commonsense view, and so my view does not suffer a
burden of contrary intuitions as compared to the opposing view. That is perhaps the
best I can establish without significant further argument about this issue and a more
concrete idea of what morality might be.

Finally, I would reject the claim that morality being a set of standards rather
than reasons means that it is not fundamentally practical. Those commands have to do
with how a person should act, and so they are still clearly practical. For the moral
person (most people, to a certain extent), these standards would figure into her
practical reasoning, so it is even more the case that for her these standards are
practical. In fact morality may require that its standards be considered as part of our
practical deliberation, in the sense that one of the rules is to do just that (it seems like
Kantian ethics have this rule, for example). So I do not think my account fails to make
morality sufficiently practical.

6. Morality and Rationality

One strand that gives rise to the thought that morality must be reason-giving
arises from the linking or even conflation of morality and rationality. Rationality has
something to do with acting upon reasons, and therefore if to act morally is to act
rationally, then there must be reasons to act morally. However, this only holds if we do agree that morality has a deep connection with rationality, something that I am not convinced by and which requires a very significant amount of further argumentation. Should it be established, however, then this would be problematic for my position. I can at least say that if rationality has to do with reasons then it would be rational to act upon prudential reasons we might have (which indeed seems accurate), even if these reasons are completely amoral, and therefore, it is at least not self-evident that rationality in the sense of acting on reasons and morality should be conflated or strongly linked. Whether they actually should be or not is not something I can determine here, as this is a very large project, so I must be content merely with noting that to make this claim is to take on a very significant burden of proof, that of linking morality and rationality, something which is still a contentious issue.

7. What We Can Say About a Non-reasons Based Morality

Even if we reject the idea that morality is necessarily reason-giving, our common sense thoughts on what we want to say about morality are not really contradicted, except for the claim that morality is reason-giving, which some people do hold. We can still say, as I noted, that somebody is doing something wrong and is worthy of criticism. That judgment depends merely on whether or not they are conforming to the rules of morality, and need have nothing to do with their reasons. After all, even those who feel that morality must provide reasons to be normative, still propose to evaluate a person's actions by comparing them to the standards of morality, which is exactly what I propose we should do. Williams noted in his argument for reasons internalism that most of what we might want to say that leads us to conclude reasons externalism is true, can be said with minor modifications on internalism.\textsuperscript{178} The

\textsuperscript{178} 1995, p 39-40.
same is true if morality is not necessarily reason-giving. We might think that saying that somebody "ought" to do something means that they have a reason to do it, however it seems to me that it could simply mean that they should conform to the standards of morality. There needn't be a reference to reasons we take that person to have in order for this "ought" to make sense, as even those with a bad set of desires and no possible motivations to do the moral thing and thus no way to act on any moral reasons we can still say should be moral. As Williams suggests, we could view our saying somebody "ought" to do something to be more accurately described as "it would be good if he did this thing" or "it would be better to do this", and so on. "Duty" is perhaps more problematic since it concerns the individual's situation more directly and has to do with commitments that they have, but I think it could perhaps be fruitfully recast in terms like "it would be good for this person to do this, because of a promise she made earlier", or "morality commands her to keep her promise", and so on. Whether this captures the meaning of "duty" is, though, rather subjective, and if this does not quite encapsulate what someone takes the meaning of "duty" to be to their satisfaction, then it seems that something will be lost on this account of morality. Perhaps, though, I can reply that this simply indicates that there was something inaccurate with the way we used the word "duty".

Some feel that morality is only prescriptive if it gives reasons, and therefore if I deny morality necessary reason-giving force, I am preventing it from being prescriptive, and turning morality into a purely descriptive matter.\footnote{E.g. Schroeder 2007, p 8.} If that were true, then it would clearly be a serious problem for my account, for certainly we want to be able to say that morality is prescriptive rather than just merely descriptive. I don't think that prescription can only be based on reasons, however, so making morality not necessarily reason-giving needn't make it non-prescriptive. Consider the law of the land: the law is
certainly prescriptive in a meaningful sense, and yet on some accounts, at least, it
doesn't give reasons in the same sense that morality is supposed to, it simply lays out
what actions will be punished by the government and to what extent. When reasons
are drawn from that it is via prudence or a desire to be lawful or whatever desires the
agent is moved by, for example, the concern could be not to go to prison, or obeying
the law might be viewed as a moral duty. Our motivation may draw on what the law
says, and we may produce reasons from the law given our other concerns, but the law
itself, I think, is not reason-giving, is not based on what reasons we have. It is,
however, prescriptive, in that it lays out rules on what can and cannot be done, in that it
commands us to act in particular ways. I view morality in much the same way; it is
prescriptive, but that is a separate issue from being inherently reason-giving. There
must be desires present in the agent that the commands of morality can connect to for
reasons to be produced. Nevertheless, insofar as morality might commend certain
things and prohibit certain others, it is prescriptive. Similarly Brink suggests that it is
often thought that to have an obligation is to have a reason, and therefore to think that
we are obligated to do whatever it is that morality commands must mean that we have a
reason to do it.\textsuperscript{180} My response to this is much the same, in that I don’t think that
having an obligation necessarily has to do with reasons. It may mean that we have a
good reason to do something, in which case naturally the claim is true, or it may mean
that there is something that morality commands us to do, which is the sense I wish to
give to it. I don’t think this latter sense is any less plausible than the former.

We might want to make the claim that morality \textit{ought} to be reason-giving, that
is, that to be moral we must view morality as giving us reasons. My account can allow
this claim too. I would view this claim as a position on one of the standards of morality,
specifically that morality commands us to take itself as reason-giving, and therefore the

\textsuperscript{180} 1989, p 52.
morally good thing to do is to take morality as reason providing. Morality, on my view, is not reason-giving for a person unless it connects to certain desires that that person has, but it is certainly still open to say that the fact that it is not reason-giving for somebody is a moral failing on their account, that is, that a person who doesn't take the demands of morality to be reasons for her is failing as a moral agent. This is basically a way of criticizing the desires that somebody has, and from the point of view of morality that is certainly possible, as desires may or may not meet the standards of morality. Related to this would be the claim that for moral agents certain things don't count as reasons, and other things do count as reasons, and the moral agent is precisely the one who can perceive these kinds of reason, the sort of picture that McDowell presents.

Again, it's perfectly possible for somebody to judge that a person not seeing certain things as reasons is demonstrating a certain kind of moral failing. Furthermore, they could say that moral agents ought not to have the kind of desires that give rise to bad reasons. These claims are justified from within the point of view of morality, even if there's a sense in which they can't be justified outside of it. I don't think that's a problem, however, as it would obviously be nonsensical to think that something could be viewed as a moral failing from outside the moral point of view.

Perhaps it's helpful to think of Harman's distinction between saying that a person has reasons, and saying that there are reasons in the sense that means "it would be good if...". On my view it's not correct to say that somebody has reasons, that is, that certain considerations really are reasons for them, if those considerations do not connect with any of their desires (as per the HTM and internalism combined), but it is still acceptable to say that there are "moral reasons" if all that really means is that it would be good for somebody to do something. This is like Williams' position above, that

---

181 Foot, 2002 p 166-7.
183 2000, e.g. p 45.
to say I have a reason not to steal is to say that it would be good if I didn't steal. In that way the typical way we have of talking and thinking, when we say that somebody has a reason to do the moral thing, can be explained as simply being claims about what is moral, about what morality demands in this particular set of circumstances. This way of talking could easily be extended to other kinds of reasons too, like reasons of self-interest. Such talk does not necessarily commit us to a position where morality is always reason-giving, and I think this is a natural interpretation of what we mean when we say that there is a reason. Talk of reasons for others just reflects moral or other sorts of judgments that we are making about what they ought to be doing, and I can allow for that.

8. Further Objections

I will now consider some individual criticisms made against the claim that morality needn't necessarily by reason-giving, by Becker and Nesbitt.

Becker suggests an interesting argument against Foot's argument, one clearly inspired by Kant's thinking. He offers a reconstruction of the crucial element of Foot's argument against morality always providing reasons as the claim that once we move outside the sphere of morality, there are no longer any reasons to accept it. With etiquette, for example, within the concerns of etiquette there are reasons to do this thing or that thing. Once we move beyond etiquette and ask "why should we care what etiquette says at all?" then we cannot give an answer that is based in any way on etiquette, but we could give prudential, etc., reasons for following its rules. What's vital is that we are able to stand outside the viewpoint of etiquette and consider whether we have reasons to do as it directs that have weight for us outside its standpoint. He suggests that in contrast taking up a position outside the moral standpoint is not

\[184\ 1973.\]
possible, precisely because moral reasons are just those given without restriction. That is, when we are given a moral reason to do something, we are being given a reason to do it "all things considered", not a reason considered within some particular domain only, or from a particular point of view, but a universal reason. There are no factors that morality rules out from being taken into consideration, therefore, it is impossible to move beyond the standpoint of morality, because unlike etiquette which is very much circumscribed, morality is unlimited and unrestricted in this way. Morality is providing the most basic reasons of all, encompassing all possible standpoints, and so it is meaningless to ask if morality provides reasons which apply to those external to the moral standpoint. Becker clarifies that he does not mean that all factors must be taken into consideration every time we make a moral decision, as only certain considerations are relevant to any given decision; he simply means that nothing is ruled out on principle when we ask for moral reasons.

One objection to such an idea that Becker identifies is that we typically do not view the sphere of prudence as being subordinate within the sphere of morality, but rather as a separate, opposing standpoint. On his view, prudential reasons are a subset of all considerations contained within moral considerations, since moral considerations are all the considerations, so any further subdivisions must be within the sphere of morality. However, we typically do not view prudential reasons in this way, but rather consider them as being concerned with different considerations than moral reasons, for example, the consideration that something is in my interest. Whether or not morality could encompass my interest depends on the exact view of morality put forward – some theories, like utilitarianism, clearly include that consideration, but others don't seem to. I can't consider such possibilities here, but will instead simply note that

---

Becker does not seem to have our usual intuitions, that prudence concerns considerations that morality does not, on his side.

I am skeptical that there really would be such unrestricted, universal reasons like this, which is no doubt unsurprising given that I have argued for reasons internalism, which tends to lead to the conclusion that there aren’t such reasons. If there were such reasons then it would have to happen that all correctly reasoning agents would come to the same conclusions, and I find that quite implausible. Why I and not another person might care, for example, about getting library books back on time would no doubt have to do with various individual considerations that were true about us, such as whether or not we each care about following rules, or being able to get any more books from that library in the future, and so on. Such differences could lead to us having different reasons, even if we end up doing the same thing. If reasons are only based on facts about the world in the largest sense, then those facts would not be unique to individuals, and so we should all be able to draw the same reasons from them. Of course, some allowance would have to be made in such a picture for certain considerations individual to certain agents to be relevant, which will lead to some variance. No (at least none I’m aware of) moral system would argue the hypothetical fact that somebody is drowning in front of me is not relevant to my moral consideration, for example. Given that, however, surely we are led to the conclusion that if moral reasons are all things considered reasons, then given the same set of external circumstances, everybody who reasons correctly will reason to exactly the same set of reasons. I do not think the evidence is there to support such a conclusion.  

Judgments about what moral reasons there are often differ widely across cultures and other factors, like socioeconomic class. Perhaps large numbers of people are simply going wrong in their reasoning, but I find it more plausible instead to think that people

---

186 Mackie also doubts this, p 61, 1990.
are coming to different conclusions because what reasons there are for them depend on some things that are true about them as individuals.

Nesbitt attacks the analogy to etiquette on a couple of different grounds. Firstly he claims that questions of etiquette aren't always "should" claims as moral claims are, but rather questions of what "rules they will lay down" and what happens to you if you break them, and morality is not like this.\textsuperscript{187} This is essentially the kind of account I gave of the law earlier. A couple of responses occur in regards to that objection, the first of which is that this is primarily a question of phrasing, rather than a substantive disagreement. Some etiquette questions may indeed not be phrased as "should" statements, but others undoubtedly are, as Nesbitt himself later admits.\textsuperscript{188} Many questions of etiquette are not only most naturally stated as should statements, but have no particular punishment attached to breaking them at all (beyond the disapproval of those around you, but that surely goes equally for morality). If I go to a fancy restaurant and use my salad fork for the main course, it's extremely unlikely that I would face any kind of punishment, and if asked I would be told that I should use such and such a fork, and not "use it or get out". So Foot's analogy does indeed work. Secondly, some conceptions morality are couched in terms of rules and punishment, just as he claims (some of) etiquette is. The more subtle and complex ideas of morality that philosophers have typically produced are not like that, but the average view of morality often depends on such ideas. Typical Christian ethics, for example, are fundamentally a question of rules with punishments attached – the idea of mortal and venial sins, the Ten Commandments, and so on, and the fact that people who break those rules go to hell. People often cite prison and other punishments as reasons not to commit crimes.

\textsuperscript{187} 1977, p 220.  
\textsuperscript{188} 1977, p 222.
There's nothing obviously incorrect, then, in taking the rules of etiquette, even those elements backed up with punishments, to be roughly analogous to morality.

Nesbitt’s second objection to Foot is to claim that we do not, in fact, think that somebody has no reason to care about etiquette if etiquette does not serve any of her desires. People who make judgments of etiquette think that they do provide reasons for anybody to follow them, he claims. I would recast this as the claim that people who accept the demands of etiquette think it would be better for others to also follow them, and, as previously stated, reasons internalism is perfectly capable of explaining such claims. I find it rather implausible to claim that people who don’t care about using the right fork (or any consequences for not doing so, like offending someone) really do have a reason to use the right fork. This is perhaps more obvious if we turn to rules of etiquette that we don't have in this culture, such as using chopsticks to eat rice. It may be customary in Asia to do so, but I fail to see what reason I have to do that in North America. Nesbitt doesn't fully argue for this claim, only claiming that it isn't obviously implausible since matters of etiquette are often very nearly matters of professional ethics and honor. This appeal is a strange one to make given that the status of moral reasons statements is exactly what is in question in Foot's argument, so appealing to them will not help decide this matter. Certainly it is true that some people believe that rules of etiquette apply to everyone, but that can be permitted in exactly the same sense as I gave for moral rules.

9. Why Be Moral

The question of whether or not morality gives us reasons to act as it commands is obviously relevant to the larger question of “why be moral?”, and obviously if denying that morality necessarily gives reasons means that we cannot answer this question,

189 1977, p 223.
then we probably had better reconsider the claim. I will argue that an answer to this question, and a justification of morality, is still potentially possible, however. The question is nothing but the question of why we should act as morality commands us to act, in other words, what reasons could a particular person have to be moral? There seem to be three general possibilities in answering this question: firstly, that morality is justified by itself; secondly, that morality is justified by nothing; thirdly, that morality is justified by something outside of morality. A full discussion of these possibilities is much too large a project to attempt here, but I will attempt to show that the third possibility, which seems obviously the strongest, is a good fit for my position that morality does not itself necessarily give reasons.

Foot suggests, very plausibly, that the first possibility is unsatisfactory, that morality justifying itself would merely be a tautology. The reason for this claim is quite obvious: if you assume the moral standpoint to begin with, then obviously there is no difficulty in showing that that standpoint is justified, and likewise, if you don't accept that standpoint, then obviously no reasons from within will get purchase on you. So to provide a moral reason to be moral would only be to convince those already convinced, and leave everyone else unmoved. The second possibility is also similar in that no purchase can be made for those who reject the moral standpoint, except that no attempt is made to provide a reason to be moral. It might be that this possibility is actually the true one (if a more skeptical point of view is accurate), however it is only one we should resort to if no attempt to find a reason to be moral proves successful. Since such reasons can, I believe, typically be found for most people, it's not required to settle for this option. Obviously neither of these possibilities are particularly satisfactory, which leaves the third option, that we have non-moral reasons to be moral.

---

The third possibility is the one similar to the general position that many thinkers have pursued, seeking to give us a reason to be moral that is a non-moral reason. This is clearly the only promising tactic to take when seeking to provide a reason to be moral, for it would give a reason to be moral that somebody who has not yet accepted moral reason may still accept and be motivated by, in other words, a reason to be moral for those who are not yet moral, or those who are struggling with a particularly demanding moral requirement. For my position, such reasons must be ones that connect to possible motivations of ours. The usual approach for this kind of position may perhaps be to base morality on some kind of prudence, on the desire to do well for yourself. However, suppose that somebody doesn’t care about their own utility, for whatever (highly unusual) reason. In that case, I will have no reason to be moral until I gain the motivations required; this can presumably be extended to any other reasons that morality could be made to rest on. There, this approach does not make morality necessarily reason-giving, as we must first desire the things that the claims of morality rest on, such as our own interest (in the case of prudence), and so on, before we have any reason to be moral. Therefore this third possibility is certainly compatible with my claim that morality does not necessarily provide reasons for everybody. The answer to the question why be moral, then, would depend on what desires a particular person has, and the claim that morality need not necessarily provide reasons does not rule this possible justification out. It could be that particular people end up with no reason to be moral but that does not, as I’ve previous argued, exempt them from moral judgment and so should not be a cause of great concern.

There is one final point to consider: what might morality end up being, on my position? I wish to leave this as open as possible, but there are several different

---

191 Of course, this objection can be escaped by rejecting reasons internalism, and therefore divorcing desires and reasons. My arguments in chapter 4 for reasons internalism would hold against such a move.
possibilities as to what it might be – which is a strength. I have generally referred to
morality as a set of standards for simplicity's sake, but it needn't be overtly like a set of
rules for action or anything like that. Given what I've previously said, if we wish to
ensure that most people do end up having reasons to be moral, morality would have to
be such that it connects with a large number of potential desires that people could have.
This need not be too restrictive a requirement however, as many theories of morality
appear to be like this. A kind of virtue ethics, for example, could be compatible with the
positions that I have laid out. It could be an objective fact that certain character traits
and/or actions lead to flourishing or some other kind of exemplary life, and so a morality
based around such a conception would be objective, and of course most people will
care about how well their own life goes. Should someone not care at all about
flourishing, then such a morality would not give her reasons, but it would still be
compatible with the HTM and reasons internalism and be objective. Likewise, a kind of
contractarianism seems promising, where, on a Gauthier kind of account, as long as
somebody cares about maximizing the satisfaction of their interests (which is surely
nearly everyone), then they would have a reason to agree to abide by certain moral
practices which objectively maximize interest satisfaction, but not if they were not
concerned with getting what they want. I think some form of consequentialism could
also work. It is objective whether or not certain actions maximize overall happiness, for
example, but likewise should you not care about happiness, then you would have no
reason to do anything about it. Obviously Kantian ethics, at least as Kant described
them, would not be compatible since he doesn't allow desires to have a motivational
role in moral action, and any morality which is heavily based on reason, as in viewing
morality as the commands of reason, would at the very least require some modification
to be compatible with my view, but it is perhaps possible that a reconciliation could be
possible depending on the exact theory. These possibilities, while not giving reasons
unless people have certain desires, are founded on desires which are surely very widespread. Most people would care about flourishing, or having their interests maximally satisfied, or about being happy (at least for themselves). While somebody could possibly have no reasons to be moral on one of these conceptions of morality, then, it is unlikely that that would be true for many people, simply because most people do have the required desires. Practically speaking then, morality's reason-giving status would not be diminished by much; rather than giving reasons to everybody necessarily, it would instead give reasons to almost everybody contingently. No doubt this is unacceptable to some, but on the other hand giving desire based reasons to be moral adds some appeal to the demands of morality for the person struggling with them, which the more intellectual theories of morality cannot quite provide.

10. Conclusion

From what I've said it is clear what an internal explanatory reason is – some consideration that is capable of motivating us after sound deliberation – but perhaps it is not yet clear what an internal justificatory reason is. In other words, I need to make clear what a consideration must have, on top of what is required to be an explanatory reason, to be a justificatory reason. Firstly, a justificatory reason is a consideration which is capable of being an explanatory reason (that is, a consideration which is capable of motivating us). Furthermore, it must be a consideration that we would point to as an answer to the question why we should do something, in the additional sense that we think it makes it so that that action is a right one to do, in that taking it to be a reason and acting on it conforms to some kind of moral standard (which I mean to be construed broadly, so as to include things like virtues, and so on, if the desired moral theory is based on such things). I have primarily focused on moral reasons here, but
we may also have justificatory reasons for prudence and etiquette and other such things in the same way.

My solution to the problem posed by holding the HTM, reasons internalism, and moral objectivity, is to deny the claim that morality is inherently reason-giving. The arguments given for the HTM and internalism provide reasons to deny this in order to make them compatible with each other and objective morality, but direct arguments against the claim also needed to be replied to. I have thus argued that the arguments given to back up the claim that morality is necessarily reason-giving fail, and that we can still say what we want to say about morality and agents, even if we deny that morality must be reason-giving. Morality does not give reasons in the sense that everybody has such reasons inherently and invariably; a person only has a reason from some consideration if they have some desire which does or could potentially lead to that consideration motivating them. This is not destructive of objective morality, however, as it maintains the elements of it that we want.
Bibliography


