Externalism or Bust - Why Internalism is Incapable of Producing Moral Reasons

RICHARD ALISTAIR ST JOHN STUART

Externalism or Bust – Why Internalism is Incapable of Producing Moral Reasons

Richard Stuart

Abstract:

Consider the following two commonly held views in Ethics. Firstly, that for something to be a reason for an agent to act it must be capable of motivating them ('reasons internalism'). Secondly, that agents have reasons – most obviously moral reasons - to act in at least some ways whatever their motivations may be. These two views would at least appear to be in conflict with each other. Internalists however, maintain that they can be reconciled.

It is argued that, given the nature of moral reasons, no such reconciliation could succeed. The argument is based in part on exploring the different attempts at reconciliation offered by three contemporary philosophers within the internalist tradition – David Gauthier, Mark Schroeder & Christine Korsgaard – each of which is shown to fail. It is then argued that this failure at reconciliation is *endemic* to internalism; internalism necessarily involves imposing a flawed constraint on what normative reasons can exist, which in practice makes it incompatible with the existence of moral reasons.

Richard Stuart – Ph.D. Thesis
Department of Philosophy
Durham University - 2021

Externalism or Bust:

Why Internalism is Incapable Of Producing Moral Reasons

Richard Stuart

Table of Contents

Statement of Copyright

Acknowledgements

I would like to offer my heartfelt thanks to my supervisory team. I am very grateful to Dr. David Faraci for his warmth, encouragement and contributions to this thesis. However, I must give special thanks to my primary supervisor, Dr. Christopher Cowie. Without his kindness, insight, generosity, breadth of knowledge, tireless support and boundless patience it simply would not have been possible to complete this work, nor, indeed, to have survived the process at all.

Thank you, both.

For My Mother

& Father

Introduction

I.1 Conflicting Intuitions

When it comes to our reasons for action, there are two widely held intuitions on the subject. Firstly, it is believed that those reasons we maintain that an agent has to act must be capable of motivating them to act, or else they can't truly be said to be reasons *for* that agent to act at all. One of the most common forms this view takes is that an agent's reason for action must be connected in some way with that agent's desires, wants, life-goals, long-term projects, etc. To say that a person has a reason to do something that will bring them or promote absolutely nothing they want or care about, either consciously or unconsciously, seems simply erroneous. This view that what reasons an agent has must be at least capable of motivating an agent to act is known as *Reasons Internalism*.

As I've just alluded to, one of the most familiar forms Reasons Internalism takes, at least in common parlance, is one that holds that it is the things we want, care about or desire that motivate us. This is because it is often tacitly accepted that it is our desires, in terms of our psychological make-up, that are primarily or even exclusively responsible for motivating us toward action. However, as we shall see throughout this thesis, the role of motivating states is not believed by everyone to be confined to desires alone. Not all forms of Reasons Internalism confine themselves to the view that only desires motivate. Reasons Internalism only mandates that to be a reason for an agent to act, it must be possible for that reason to motivate that agent in some way.

The second widely held intuition is that there exists a special class of reasons for action (or inaction), which we call *moral reasons*. There are some actions we just should or shouldn't do. We hold, or at least most of us do, that affluent agents have a moral reason to give at least some of their excess income to those in desperate need and that all agents have a moral reason to refrain from torturing innocent people, as just two examples. Additionally though, aside from their existence we think that generally moral reasons have a peculiar set of characteristics that set them apart from other kinds of reasons for action that agents have. One of the most prominent of these is the sense that they have a certain independent authority to them that we think of as having sufficient clout to trump (at least to some degree) any desires we might have to act contrary to them. In other words we take it as an essential feature of a moral reason for action that

it applies or doesn't apply to an agent, entirely regardless of whether the agent actually feels motivated to act in accordance with it. The miserly billionaire just alluded to has a moral reason to give some of their surplus cash to those who are in desperate need through no fault of their own, regardless of how little they feel motivated to do so. Likewise, the aforementioned psychopath has a moral reason not to torture innocent people, regardless of the joy and satisfaction it brings them. Our attribution of moral reasons to agents then seems, at least at first glance, to make or require no reference to the actual or potential motivational states of agents for them to be true.

These two intuitions would seem at face value to be in conflict with each other. If, as the reasons internalist demand, reasons for action must by necessity be capable of motivating an agent, yet we can readily imagine agents lacking any motivation to do as we think they have moral reason to, how can moral reasons have the kind of unconditional authority they appear to have? If moral reasons do exist and do provide genuine reasons for an agent's action independently of their desires (or any other motivational state), then *Reasons Internalism* must be either wrong or be an incomplete theory.

Historically, philosophical responses to this apparent conflict have traditionally broken down into three main forms.

A) *Error Theory*: Reasons Internalism is true and this leaves no room for objectively authoritative moral reasons, whose mere truth guarantees their motivational efficacy – hence they do not exist.
B) *Reasons Externalism*: Reasons Internalism is false, or at best only provides a partial account of reasons for action. Moral reasons exist but they do not need to have any necessary connection with an agent's motivational or psychological make-up in order that they apply to those agents.
C) *The Reconciliation Project*: Reasons Internalism is sound, but the existence of moral reasons is and can be shown to be entirely compatible with it.

It is not the purpose of this thesis to take any firm position on the veracity of either option (A) or (B). If the conclusions drawn in this work incline the reader toward either or neither of them, I am equally satisfied. Instead, the current work is intended as nothing more than an outright and full-throated rejection of option (C) – the

Reconciliation Project. It is my position that any moral theory that insists that a reason for action *must* be capable of motivating an agent is incapable of providing anything that we should classify as a truly moral reason. To put it another way, Reasons Internalism is incapable of generating truly moral reasons and so the Reconciliation Project is doomed to failure. If it can be established that Internalism is incapable of furnishing us with genuinely moral reasons, we would be left with only two options. Either we would have to conclude that an externalist moral theory is the only one that has the potential to furnish us with genuinely moral reasons; or we would have to conclude that there are not and cannot be any true moral reasons – Error Theory. Hence the title of this thesis; it's 'Externalism or Bust!'

I shall do this by clearly outlining and arguing for the kinds of characteristics a reason for action has to have in order for it to legitimately be a moral reason in the first place. Furthermore I shall show why, taken together, these reasons can't be furnished us by any possible theory which places the kind of motivational burden that Reasons Internalism places on reasons for action.

To summarize briefly here, I maintain that to be a *moral* reason a reason for action must be,

1) Categorical;
2) Be of non-negligible weight (or strength)[1]; and
3) They must have the right grounding.

I.2 What are reasons for action, *per se*?

If we are going to talk about 'moral' reasons it is important to get clear, or at least clearer, on what we mean when we talk of reasons for action, more generally. There are many competing accounts. However, for the purposes of this thesis, I shall adopt a position not totally dissimilar to that of Tim Scanlon in the opening chapter of his *What We Owe to Each Other*.

> 'I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favour of it. "Counts in favour how?" one might ask.

[1] I shall be using the terms 'weight' and 'strength' interchangeably throughout this thesis as I consider them effectively synonymous in the context of this discussion.

> '"By providing a reason for it" seems to be the only answer. So I will presuppose the idea of a reason, and presuppose also that my readers are rational in the minimal but fundamental sense that I will presently explain.'[2]

I, likewise, shall take it as a given that a reason is simply some fact about the world that counts in favour doing something. The precise details as to *how* a reason counts in favour of doing something will be discussed as and where necessary, and in the context of examining the different specific internalist theories, which will take place in Chapters Three, Four & Five of this thesis. I do not wish to risk begging-the-question against Internalism by setting up a stringent definition of what reasons for action, *per se*, are or must be.

That being said, I think it important to explain briefly why I've chosen to focus on reasons for action in this work. Whereas I read Scanlon as adopting, at least in part, his reason-fundamentalist stance based on a sincere belief in the metaphysical simplicity of reasons, my reason for treating them thus is purely pragmatic. I am open to considering that they could be analyzed more deeply. However, since the conclusions of this thesis do not turn on the metaphysical status of reasons for action *per se* and I maintain that there is a distinct theoretical advantage to taking reasons as fundamental, which I will come to, I will be treating reasons in this spirit. For the purpose of the current work, the precise nature of what constitutes a reason is far less important as to whether or not they can, do or must motivate agents. Therefore, it is on their motivational efficacy that we shall focus.

So, why reasons? In the words of Julia Markovits,

> 'Moral philosophers have long been concerned about how to respond to the *amoralist* – the person who recognizes what morality requires of him, but wonders *why he should do* what morality requires. The *moral ought*, this amoralist might concede, is certainly *about* him – it *refers to* him. But it doesn't follow merely from this that it has a *proper, normative hold on him* (whatever that comes to), any more than the fact that the dictates of some old-fashioned religion – a religion that in no way reflects what I care about – refer to me entails that I have any *real reason* to comply with them.'[3]

[2] T.M. Scanlon, *What We Owe to Each Other*, The Belknap Press of Harvard University Press (1999), p17.
[3] Julia Markovits, *Moral Reason*, Oxford University Press (2014), p.4.

I have always found the tenability of the amoralist point of view to be deeply troubling. The purpose of ethics, at least in part, is that it can curtail agent action, i.e. provide reasons for us not always to do what we want to do in the interest of other people's welfare. If it did not have this characteristic, as Hume rightly observes, it would be superfluous – people would just act as they wanted to all the time. Why would we ever do what is seemingly not in our interest if we could get away without doing so, at least part of the time? It always seems to me a rational and non-trivial question to ask why we should be moral. To paraphrase Kant; the scandal of ethics is that we have no answer to the moral skeptic.

I have long held that a potential means to silence the moral skeptic could be found in the way they pose their challenge to the moralist. "What *reason*," they ask, "do I have to be moral?" More often than not, the moral skeptic takes the existence of at least some reasons for action as a given. Typically, even the most hard-nosed moral skeptic accepts that they have some reasons to carry out some actions. Acceptance of the existence of reasons for action then, is something the moralist and the amoralist, by-and-large have in common. Reason for action *per se*, is a shared reality – a *lingua franca* uniting them, if you will. If a moral theory can be formulated where sufficiently weighty *moral* reasons for action can be shown to be no more or less plausible and coherent than the reasons the moral skeptic, for the most part, already takes as being largely unproblematic, we will have provided them with the best answer we can.

The view I am outlining is most commonly referred to in the literature as *Reasons Primitivism*, or sometimes *Reasons Fundamentalism*. This position holds that most, if not all, ethical concepts can be reduced to talk of what we have moral reasons to do and not do. For example,

> 'Goodness is not a single substantive property which gives us reason to promote or prefer the things that have it. Rather, to call something good is to claim that it has other properties (different ones in different cases) which provide such reasons.'[4]

For a reasons primitivist, what is 'Good' is simply something in the world the promotion of which provides agents with reason for action. What an agent *ought* to do

[4] Scanlon (1999), p11.

is not some special *sui generis* concept, but simply the thing that one has the strongest and best reason to do out of the plethora of different actions that agent might have reasons to carry out.

To repeat what I have written above, I do not embrace Reasons Primitivism out of a particular metaphysical commitment to it. Rather, I utilize it as a working model for purely pragmatic purposes. For, if its goals can be realized then I regard it as the single best avenue for meeting the challenge of the moral skeptic.

This is enough regarding reasons *per se* for the moment.

I.3 The Structure & Purpose of this Thesis

In the words of Huw Price, 'a thesis is six chapters long'! In keeping with this sound model, the current work contains six chapters, an introduction, which you are currently coming to the end of, and my conclusions.

To state it clearly, the overall task of this thesis is twofold. Firstly, it is to argue rigorously what a moral reason has to be in order to be worthy of the label 'moral' in keeping with the three criteria I mentioned in I.1. Secondly, it is to argue that moral reasons, so understood present a persistent problem for internalist theories of reasons. The precise arguments for each of the three criteria and why, when taken in conjunction with each other, they should constitute essential requirements of moral reasons represent substantial original work on my part and application of existing arguments. Likewise, the case I shall make for an error, which is in practice endemic to all forms of Internalism, and how specifically it renders it highly improbable that any form of Internalism shall be constitutionally capable of meeting the essential requirements I have outlined, represents original work on my part, and I hope a novel line of attack on Internalist theories apart from those which have gone before.

One prong of my strategy to highlight the inadequacies of Internalism shall be to apply my three criteria to three very different forms of Internalist theories. These shall be the neo-Hobbesian Contractarianism of David Gauthier, with particular reference to his *Morals by Agreement*; the neo-Humean Hypotheticalism of Mark Schroeder from his *Slaves of the Passions*; and finally, the neo-Kantianism of the early Christine Korsgaard in her *Sources of Normativity*.

I have selected these three specifically, from among all the possible internalist theories I could have chosen for two different reasons. Firstly, they are three of the most

well-known and influential theories. For this reason, any success I have in undermining them will be of greater import. Secondly, they are very different types of theory that utilize that apply their internalist assumptions in very different ways. A line of attack that can be applied to such a diverse range of theories is therefore one that is more versatile and profound. It will expose the inherent flaws with internalist assumptions, *qua* internalist assumption. This in turn will be useful in adding support to my claim that Internalism is constitutionally incapable of yielding true moral reasons.

Chapter One tasks itself with my outlining the three criteria I believe are essential for a reason to count as a moral reason. This will include an explanation of what each of them amount to and why they are essential to moral reasons. As I consider it the most contentious of the three, the section on categoricity will also include a lengthy examination of the kinds of arguments that might be deployed to deny its vital importance to moral reasons and why I ultimately reject all of them.

Chapter Two will be a first foray into my critique of Reasons Internalism. It will take the form of elucidating why from the outset there is a *prima facie* reason to believe that a reasons internalist theory is going to have problems generating reasons for action that can meet any, let alone all of the three criteria. However, in the spirit of fair play, it will also outline the minimum requirements that an internalist theory would have to meet to provide moral reasons – none of which are necessarily incompatible with Internalism, at least from the outset.

Chapters Three, Four & Five respectively, will each be an examination of the three different and highly influential internalist moral theories by the three different individual thinkers I have mentioned above. Chapter Three will look at David Gauthier. Chapter Four will look at Mark Schroeder. Finally, Chapter Five will look at Christine Korsgaard. In each case I will outline their core arguments as fairly as possible and then explain to what extent each succeed or fail to meet my three criteria. For each of them I will conclude that they all fail to meet at least one of the criteria to a sufficient degree to make it impossible for them to furnish us with genuine moral reasons, and hence that they are not fit for purpose.

Finally, Chapter Six will make the case more strongly, that not only do all of these individual internalist moral theories fail, but that *all* internalist moral theories fail in practice because Internalism by its very nature is committed to an *Endemic Error*, already mentioned, concerning the way it constrains what normative reasons can exist.

Fittingly, and in the traditional fashion, this work will close with a conclusion. This will draw together and make explicit the consequences of all of the forgoing chapters, particularly those of Chapter Six. The hope of reconciling our intuitions that all of our reasons for action must be capable of motivating us with our intuitions that we have moral reasons for action, is unlikely to ever succeed. In the light of this, there are only two viable options. Either we redouble our efforts to find a suitable externalist theory to provide us with moral reasons, or we accept that there are no moral reasons for action and accept the Error Theory.

Chapter One

What Kind of Reasons Are Moral Reasons?

1.1 Introduction

This chapter will explain what characteristics a reason for action must have if it can truly be considered a moral reason for action at all.

As I have already mentioned in the Introduction, to be a moral reason a reason for action must be

1) Categorical
2) Of non-negligible weight; and
3) Have the right grounding.

Any reason for action that does not adequately meet these three criteria is not a *moral* reason for action – though it may remain a perfectly valid reason for action all the same. Sections 1.2-1.4 will outline exactly what each of these three criteria mean and the case for why they are indispensible to the idea of a reason being a moral one.

1.2 Categoricity

> '[T]he distinction between requirements that are binding on someone conditionally on her having a certain desire, and requirements that are binding on someone unconditionally, that is whether or not she has a certain desire. The first are the hypothetical imperatives, the second are the categorical imperatives.'[5]

The distinction between *hypothetical* and *categorical* reasons for action concerns the role of desire in the accurate ascription of reasons to a given agent. A hypothetical reason is of the type, 'if *A* wishes to get to London by noon, they have a reason to catch the 10:30am train to King's Cross'. The truth of *A* having a reason to catch the 10:30am train is dependent on *A* actually having the sincere desire to get to London by noon. If

[5] Michael Smith, *The Moral Problem*, Blackwell Publishing (2011), p.77.

they have no such desire, then they have no such reason. Hypothetical reasons are those that are dependent in some way on an agent's desires in this way, to be true.

Categorical reasons, on the other hand, are reasons an agent has whatever their desires. They are the kinds of reasons an agent can or must have regardless of anything that agent may specifically want or not want. An example might be the kind of reasons for action the law of the land provides people with. Tax-law for example takes the form that an agent or citizen has a reason to pay taxes, regardless of whether or not they want to. Epistemologically speaking, if all evidence points to Theory A be true and Theory B being false, you have a reason to accept Theory A and dismiss Theory B – regardless of which theory you would prefer to believe.

An important distinction needs to be made here. I specifically operate with this less stringent notion of categorical reasons – i.e. that they are reasons an agent has *whatever* their desires. It is what I understand by a categorical reason throughout this thesis. A stronger definition, and one that is employed by some theorists, is that categorical reasons are reasons agents have *independently* of what their desires are. This however is too strong. It would make it impossible by fiat for any internalist theory to satisfy the categoricity requirement – since internalist theories necessitate some desire (or other motivational state) for there to be a reason in the first place. I shall return to this point later. For now though, suffice to say that satisfying this less stringent standard of 'categorical' is enough for our purposes.

Immediately, I think it is clear why our most basic intuitions regarding moral reasons, and the non-negotiable authority they strike us as having over us, would incline us to class them as being categorical in nature rather than hypothetical. When we make *moral* judgments concerning agents, about what morally they *should* do, we do not take what they want to do into consideration. Our conviction that the mugger does wrong by stealing a pensioner's purse is not negated by finding out just how very much the mugger wanted to steal the purse or how little affection they had for the pensioner. Judgments about what moral reasons people do and don't have typically take the form of absolute edicts – 'Though shalt not kill, unless you really want to' would be no moral command at all! Certainly we may, when raising children or rehabilitating prisoners, for example, try to encourage agents to align their desires with accepted morality, so they genuinely grow to desire 'doing the right thing'. But we do not think the validity of the moral reason ultimately depends on the presence of such a desire. As Joyce puts it,

> '[W]hen we say that a person morally ought to act in a certain manner, we imply something about what she would have reason to do regardless of her desires and interests.'[6]

Most of us will be familiar with the hackneyed science-fiction trope of an emergent artificial intelligence that sets out on the genocidal extermination of humanity with great efficiency but no qualm. The fact that the AI seems structurally incapable of comprehending any salient reason to desist does not imply that there isn't a reason for it to desist – i.e. that it is wrong to indiscriminately exterminate human beings.

We insist that moral reasons be the kinds of reasons that apply to agents whether the agent wants them to or not, or acknowledge them or not. If they could be wheedled out of based on the mere caprice of agents they could not serve as the *absolute* constraint on action they are both purported and sincerely held (by most) to be.

To clarify then, moral reasons must be categorical. And a reason for action is categorical if it applies to an agent – i.e. the reason applies to an agent or doesn't – entirely regardless of what their desires are. However, a little more than this needs to be said first. There is more than just one way of understanding categoricity.

As Philippa Foot covers in her famous *Morality As A System of Hypothetical Imperatives*[7], certain institutions imply the existence of categorical reasons to do things and not others. Such institutions include, but are not limited to, sports, games, etiquette and, as already mentioned, the law. It is accepted, for example, that one has a categorical reason not to violate the offside rule when playing association football, or to not use a dessert spoon when eating soup. These reasons are deemed categorical since they apply to those participating in the activity or institution. The reasons they give rise to are also deemed to provide normative reasons for action – i.e. if one is playing football one *should* adhere to the offside rule. We can refer to this as institutional categoricity and institutional normativity. They are categorical relative to the institutions in that their applying to agents, or 'players' in the football example, occurs independently of any

[6] Richard Joyce, *The Myth of Morality*, Cambridge University Press (2001), p.34.
[7] Philippa Foot, *Morality as a System of Hypothetical Imperatives,* reprinted in *Foundations of Ethics*, Edited by Russ Shafer-Landau & Terence Cuneo (2007), p.287.

desire the agent has. They are normative in that they prescribe how the agent (or player) should act.

However, what normative force institutional categoricity has seems entirely dependent on the reasons an agent has to participate in the institution[8]. You only have reason to follow the rules of a game, after all, if you have a reason to play the game in the first place. This is not the kind of categoricity we typically and intuitively think of as moral reasons having. The kind of categoricity moral reasons have seems stronger. We think of them as having a, what might be called, *genuine* normativity. In what specific way this normativity is 'genuine' is hard to define. Different theorists have different ideas about it. Two things though seem to be tacitly constitutive of it. It does not seem to depend on the existence of any contrived or invented institution for its authority, and also, its authority seems to be *inescapable*. By the latter I mean that where we deem that an agent has a moral reason to act a certain way, *ceteris parabus*, the agent has no means whatever of 'opting-out' of this reason applying to them; unlike the way the player of a game might quit at any time or renounce their citizenship of a polity, and with it their obligation to obey its laws.

Precisely demarcating between institutional normative reasons and genuinely normative reasons is hard to do. Nevertheless, it can't be denied, when we *do* talk of moral reasons we seem to imbue them with what Joyce refers to as their 'practical oomph'[9]! It is an oomph that merely institutional reasons just don't have for us. I don't think the lack of an accepted definition of genuine normativity is required to make the case I wish to make regarding the essential importance of categoricity to moral reasons – only a clarification of the sense of categoricity I shall be using throughout this thesis to refer to the kind or caliber of categoricity that I take moral reasons as needing to have.

The kind of categoricity I will be referring to as essential to moral reasons is one that is genuinely normative in this stronger sense that merely institutional reasons can never be genuinely normative. It is genuinely (or perhaps, 'absolutely') categorical in that it is not dependent on the existence of any institution and it is inescapable – i.e. *no* volition of the agent could alter whether or not a moral reason applies to them, if and when it does.

[8] See J.L Mackie's discussion in his, *Ethics: Inventing Right and Wrong*, Penguin Books (1990), p.67
[9] Richard Joyce, *The Evolution of Morality*, New York: MIT Press (2006).

We will be returning to this issue from time to time throughout this chapter and the rest of the thesis, as and where relevant. However, I hope this opening discussion will help clarify more precisely what is meant by the term 'categoricity' in the passages that are to come.

So, of the three characteristics that I believe moral reasons must have in order to be genuinely moral, which I've already listed, categoricity is in my estimation the most important and the hardest internalist moral theories have to account for. While the other two, weight and right grounding are also vital for providing moral reasons, it is the failure to provide genuinely categorical reasons that will make up the lion's share of my critique of the various forms of Internalism we will be looking at.

However, before we proceed to looking at the different ways internalist theories, *qua* internalist, fail to meet this requirement, I wish to look at some theories that attempt either to deny that moral reasons need to be categorical, or purport to provide categorical moral reasons that are anything but. Attempts to undermine the categoricity of moral reasons come from many directions and in many forms. Perhaps the obvious example, and the one we will look at first, is from moral relativists like Gilbert Harman.

1.2.i Harman

In his *Moral Relativism Defended*, Gilbert Harman targets moral reasons specifically as being necessarily ones that contain a motivational component.

Harman starts by drawing a distinction between 'inner' and 'outer' judgments regarding what certain agents 'ought' to do[10]. An outer judgment would be of the kind when we condemn the actions of another individual or group, entirely without reference or regard to any goals, desires or motivations those we condemn may or may not have, entirely from the standpoint of our own moral principles. Inner judgments, on the other hand, are when we can legitimately accuse an agent of committing a palpable error – i.e. there really was something the agent 'ought' to have done, which was determined by that agent's own goals or principles, and they failed to do it. To put it another way, outer judgments are when we judge agents by our standards and inner judgments are when we judge them by their own.

[10] Gilbert Harman, *Moral Relativism Defended*, reprinted in *Foundations of Ethics*, Edited by Russ Shafer-Landau & Terence Cuneo (2007), p.85.

> 'Inner judgments do not include judgments in which we call someone […] inhuman, evil, a betrayer, a traitor, or an enemy.'[11]

On my reading of Harman, it is this latter form of palpable error, *qua* a failure of an agent to act in line with their own accepted standards or principles, which is constitutive of the immoral; not abject repudiation in accordance with some criterion that is external to the internal value judgments of the agent being condemned.

Harman proceeds to make his case by analogy. He gives four main examples of times when, though we might reject, in his outer sense, the actions of an individual or group, we could not convict them of a transgression of their own standards or principles. These are,

1) 'Intelligent beings from outer space […] beings without the slightest concern for human life and happiness.'
2) '[A] band of cannibals [who have] eaten the sole survivor of a shipwreck'
3) '[A] contented employee of Murder, Incorporated […] raised as a child to honour and respect members of the "family" but to have nothing but contempt for the rest of society.'
4) Hitler.

In each case, despite our condemnations of and/or resistance to the actions of each individual or group, Harman asserts that given that they are acting in accordance with their own goals and sets of values, it would be 'odd' to say that they ought not to act as they do or did. In reference to the contented member of Murder, Inc., who has been charged with the assassination of 'a certain bank manager, Bernard J. Ortcutt', Harman says,

> 'in this case it would be a misuse of language to say of him that he *ought* not to kill Ortcutt or that it would be wrong of him to do so, since that would imply that our own moral considerations carry some weight with him, which they do not.'[12]

Remember, for Harman, as I read him, immorality requires wrongness, which implies error, which is a violation of practical rationality. With this definition, or

[11] Ibid, p85.
[12] Ibid, p85. My Italics.

benchmark, in place there is only one thing left for immorality to be for Harman. The only thing that foots the bill for him, in terms of constituting an immoral act, is a violation of some agreement, implicit or explicit, which for him will always be relative to some group, community or polity.

Harman attempts a more formal logical formulization of his position thus,

> 'Formulating this as a logical thesis, I want to treat the moral "ought" as a four-place predicate (or "operator"), "Ought (*A*, *D*, *C*, *M*)," which relates an agent *A*, a type of act *D*, considerations *C*, and motivating attitudes *M*.'[13]

And there you have it! For Harman the 'ought' of morality, by definition, is dependent on the motivations of agents – and hence is explicitly non-categorical. However, is Harman justified in identifying what moral reasons there are for an agent and what reasons they have by virtue of their own commitments or motivations? I say not.

The first step of which we should be suspicious in Harman's argument comes when he states,

> 'We make *moral* judgments about a person only if we suppose that he is capable of being motivated by the relevant moral considerations.'[14]

In one sense this statement is simply false. We often make moral judgments about a person based entirely on their actions and with indifference to their motivations. Before anyone accuses me of begging the question against Harman here, by saying that the point Harman is making is that such judgments aren't true moral judgments – there is nothing in Harman that deals with these kinds of common, everyday categorical judgments people often make about other people without reference to motivations. He doesn't argue against them so much as doesn't seem to acknowledge they occur. Secondly, since many perfectly intelligent and philosophically literate people say that what they are doing when they do this is making a moral judgment, without further argument from Harman, we are not entitled to discount such judgments as moral judgments. Harman can't discount them by fiat.

[13] Ibid p87.
[14] Ibid, p85. My italics.

Perhaps Harman is saying that to judge that someone is of poor moral character, we must judge that they are someone who acts against moral principles they do in fact accept, and would acknowledge that they *should* be motivated by, even if they do act immorally. I think a clue to an important distinction that Harman is making comes with the following,

> 'If someone *S* says that *A* (morally) ought to do *D*, *S* implies that *A* has reasons to *D* and *S* endorses those reasons – whereas if *S* says that *B* was evil in what *B* did, *S* does not imply that the reasons *S* would endorse for not doing what *B* did were reasons for *B* not to do that thing; in fact, *S* implies that they were not reasons for *B*.'[15]

Now I'd like to say at this point, I do not much care for the word 'evil'. It is nebulous, hard to define and altogether too mired in its theological origins to be truly useful in the formation of the kind of account of ethics that I favour. However, if I were disposed to employ the term, I might say something like this – had Hitler (or someone like him) sincerely believed that what they were doing was wrong but done it anyway, I would have said that they were merely immoral (or maybe akratic). Hitler was 'evil' *precisely because* the extermination of innocent Jews, Slavs, Poles, Roma, homosexuals and all-to-many others, did not constitute a violation of his morals. However, the implications of this view in terms of questions pertaining to the attribution of blame and approbation, and the infliction of punishment, are too far-reaching for the scope of this thesis. I will therefore leave them to one side from here on.

Suffice to say, I do not think that there is any major disagreement between Harman and myself regarding what we are concerned with when we dub an agent 'evil'. Where we do disagree is what kind of judgment we are making when we do just that. I think Harman would say that the attribution of the epithet 'evil' is an example of an outer judgment, *par excellence*. But Harman states that outer judgments are not in fact legitimate moral judgments at all.

[15] Ibid, p87.

> 'If reference is made to attitudes that are not shared by the speaker, the resulting judgment is not an inner judgment and does not represent a full-fledged moral judgment on the part of the speaker.'[16]

So moral judgment *only* applies to inner judgments. When it comes to outer judgments we are confined merely to express our own moral outrage or disgust; that the actions being carried out are in violation of our own sincerely held moral principles.

> 'If *S* says that (morally) *A* ought to do *D*, *S* implies that *A* has reasons to do *D* which *S* endorses. I shall assume that such reasons would have to have their sources in goals, desires, or intentions that *S* takes *A* to have and that *S* approves of *A*'s having because *S* shares those goals, desires, or intentions.'[17]

But this is a mischaracterization of what is really going on when we state what we believe another agent ought (morally) to do. It is not a question of *S endorsing* the reasons they believe gives *A* the moral reason to do *D*, which are sourced in their shared desires. This would be only to provide *A* with the 'ought' of practical reason. *S* is stating, tacitly or otherwise, that a moral reason exists, *simpliciter*, for *A* to do *D* – no further qualification is necessary. Now, *S* may be entirely misguided in thinking that such reasons, *simpliciter*, exist. Nevertheless, this is a more accurate characterization of the form assertions of moral reasons take – and the flavour of these assertions is, at least *prima facie*, categorical. Outer judgments are made, for the most part, with conscious disregard for the motivations or desires of the transgressor to whom they are directed.

I believe Harman is guilty of begging-the-question by playing fast and loose with the word 'ought', treating what all-things-considered an agent 'ought' to do and the specific moral 'ought' as one-and-the-same without justification. There is a conceptual gap between what one ultimately has the strongest reason to do, all-things-considered, and specifically what one has moral reason to do. It is conceivable that they are not always one and the same thing. Nevertheless, my point is, even granting that inner judgments of the kind Harman speaks of are moral judgments, why are we not entitled to consider outer judgments as moral judgments of a kind also? If it can be granted that they are, then the question of what constitutes their truth-conditions becomes relevant

[16] Ibid, p88.
[17] Ibid, p87.

– i.e. is an outer judgment merely the expression of the judge's feelings regarding another's actions or, or is it more often a legitimate attempt to assert that certain types of action, in-and-of-themselves, warrant being eschewed or resisted? For my part, the form that outer judgments often take implies that, at least on some occasions, the latter interpretation is the appropriate one. Harman explicitly states that claiming that a reason to act (or not act) is a reason *for* someone to do something, it must be an inner judgment as it pertains to their practical reasons. That being the case, if an outer judgment *does* ground any kind of reason for action, it would have to take the form of a reason for action *simpliciter*.

> 'Although we would not say concerning the contented employee of Murder, Incorporated mentioned earlier that it was wrong of him to kill Ortcutt, we could say that his action was wrong and we could say that it is wrong that there is so much killing.'[18]

Harman's own examples of the invading aliens, indifferent to human life and welfare, exposes a major reason to hold that outer judgments can have potent practical import of a distinctly moral character.

> '[W]e will want to *resist* them if they do such things […] we will judge that they are dreadful enemies *to be repelled and even destroyed*, not that they should not act as they do.'[19]

In this case, the outer judgment is a call to action – action that, as was likewise the case in resisting Hitler and the Nazis during WW2, called for a willingness to kill or be killed in its undertaking. Harman also makes no claim that to put the inner 'moral' judgment to one side and act instead in accordance with the outer judgment of utter defiance would be in any way irrational of us. In these examples, not only does Harman clearly envisage the outer judgment taking precedence in determining the ultimate course of action; but also that the kind of grave, earnest, world-altering action it can inspire, clearly gives it the kind of weight and import we typically associate with the moral. If outer judgments are not really moral judgments, is it rational for us to risk so

[18] Ibid, p86.
[19] Ibid, p85. My italics.

much to fight the invading aliens or Nazi hordes? Why are we so willing to kill and die for the sake of outer judgments? If these do not count as real examples of moral judgments, at least on some occasions and of *some* kind, I don't know what could!

It has been suggested to me that the simple fact that I would want to resist something occurring – an alien invasion for example – doesn't necessarily imply that it is due to a moral objection to the invasion. Here I can only appeal to the kinds of intuitions these types of situation invoke. In the case of a direct threat to one's own life, I agree, an agent's reason might not best be categorized as moral. But there is no shortage of examples from history where individuals have undertaken tremendous hardship and danger for the cause of protecting others or for causes that would not otherwise have directly impact their lives. I can think of no better way of describing this other than standing for a principle – for the rightness or justness of the cause itself – and so as being the result of a moral judgment. To refrain from dubbing these types of judgments 'moral' would seem to me an aberration for our standard use and understanding of the word.

To make the point a slightly different way; is Harman *really* willing to allow that an outer judgment could have been a reason for someone to put themselves into harms way to assassinate Hitler in 1941, purely on principle, but that at the same time this very same principle was not, at least in *some* sense, a reason for Hitler not to have done the things he did? More formally; If *S* is justified in resisting/destroying/killing *A* because *A* persistently *D*-s, i.e. *S* has legitimate reason to resist/kill/destroy *A*; surely that must necessarily entail that *some* reason exists for *A* not to *D*. If such a reason existed, by Harman's lights it would exist independently of any inner judgment we make of *A* – hence, there would be some categorical moral reason for *A* not to *D.*

So, to summarize, Harman has given no compelling argument that,

1) Outer judgments are not a form of moral judgment,

2) That outer judgments do not provide some of the strongest reasons for action there are. (Indeed, it seems to me that he has conceded the opposite), and

3) There's any ground for rejecting that the reasons for action generated by such outer (moral) judgments are not categorical.

Perhaps Harman has raised an interesting point about the nature of subjective moral deliberation and, as I have already mentioned, the appropriateness of blame and/or punishment. However, for my purposes I only require that outer judgments

could be legitimate and significant examples of moral judgments and that the veracity of such judgments could be grounded on the existence of categorical moral reasons for action. Harman has presented no compelling case that would rule this out. I therefore feel comfortable moving on from him.

1.2.ii Street

The preceding discussion of Harman dovetails nicely with something Sharon Street attempts to establish in the latter sections of her *In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters*.

There are several interesting parallels between Street's and Harman's papers. Both get us to focus on groups or individuals that demonstrate behavior that we would typically class as being immoral, or that we would typically *want* to class as being immoral. Then they try to get us to acknowledge that there is a very real sense that it is difficult to say of such individuals that they have committed an actual *error*, if their actions are genuinely in keeping with their desires and goals. What differs between them, however, is where they consider the true realm of the moral residing. However, I will come around to this in due course.

Street begins by outlining a certain type of 'character' that is often deployed in meta-ethical debates.

> 'The characters I have in mind are purely hypothetical, and they're distinguished by two main features. First, they accept some value that is utterly unheard of, morally repugnant, or both. Second, their acceptance of this value coheres perfectly, as a logical and instrumental matter, with all of their other values in combination with the non-normative facts. Call these characters *ideally coherent eccentrics*.'[20]

She claims that the possibility of such ideally coherent eccentrics (or just 'ICEs') existing is used to imply the unacceptability or absurdity of 'attitude-dependent' theories of value and hence, the superiority of 'attitude-independent' theories. The argument goes as follows; on an attitude-dependent model, what an agent has genuine normative reason to do is determined by the things they value – whatever they be. To

[20] Sharon Street, *In Defence of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters*, Philosophical Issues, vol. 19, 2009, Section 1.

all intents and purposes those things we ultimately value are not subject to error or correction. Agents just *do* value some things and not others – 'the heart wants what it wants'. Given this, it supposedly follows that an agent does in fact have the greatest normative reason to act in the way that brings them what they value.

Now, borrowing from Gibbard, Street presents us with an ideally coherent Caligula, who deeply and sincerely values torturing people for fun. The implication is that on the attitude-dependent model, we are forced to conclude,

> '(4') The ideally coherent Caligula has most normative reason to torture people for fun.'[21]

It is Street's claim that the critic of the attitude-dependent model takes the *prima facie* absurdity/unacceptability of this conclusion as sufficient indictment of it. The critic may acknowledge that Caligula would have *some* normative reason to act in accordance with his sadistic desires. However, there must be something wrong with a moral theory that does not require that a countervailing reason of greater normative force exist for Caligula not to torture. Street argues that the critic's confidence in this assumption and, more generally, in the significance of ICEs in the metaethical debate between attitude-dependent and attitude-independent models, is misplaced.

Street's response is two-fold. Firstly, she argues that (4') is in fact not as absurd or counter-intuitive as the critic wants us to think. If we remove from the sense in which Caligula 'ought', or has greatest normative reason, to torture for fun, all *moral* connotation, the ought in (4') becomes immediately more palatable. From Caligula's point of view, if he most desires the suffering of others, then in cold, clinical and descriptive terms he *does* have the most practical reason to act in pursuit of this. For Street, the counter-intuitiveness of (4') is chiefly caused by a failure to distinguish between different yet equally legitimate uses of the word 'ought'.

Secondly, she argues that the main force behind the critic's point is that if we accept (4') we are somehow committed to a Harman-like moral relativism. However, this is where the question turns on what constitutes the realm of the moral.

You'll recall that for Harman, only so-called 'inner judgments' are full-fledged examples of moral judgments at all. For Harman holds that to be a moral judgment it

[21] Ibid, Section 1.

must have genuine normative force. Since, for Harman, the only reasons for action that have genuine normative force are those that are grounded in our desires, moral authority pertains only to inner judgments.

In a sense, Street follows in Harman's footsteps and yet subverts this last step. She, likewise, accepts that true normative force resides within these inner judgments. What she rejects is that there is any necessary overlap between what we have greatest moral reason to do and what we have greatest normative reason to do.

For Street, moral 'facts', and the reasons for action they imply, are constructed out of human interactions, needs and contracts (explicitly-formulated or otherwise)[22]. For example, an agent's *moral* obligation to not wantonly break a promise, entered into willingly and in good faith, are not grounded by their individual, practical reason for action. Instead, they are grounded by the sanction of promise keeping/breaking within society (again, tacitly or otherwise). Morality so constructed provides a domain of facts about how agents *morally* ought to act, which can be made evident from an examination of a given society. These facts need not necessarily align with the wants or desires of a given agent all the time – or in the case of an ideally coherent Caligula, ever!

In this way, what Street is arguing for, *contra* Harman, is a standard of moral truth (or aptness) that is not necessarily married to what agents have the greatest *normative* reason to do, but that nevertheless remains what they ought morally to do. By so doing, Street is not forced to accept a Harman-like moral relativism.

On this view, our ideally coherent Caligula can be normatively sound yet remain morally reprehensible. We do not need for Caligula to be behaving irrationally to say that he is behaving immorally and to have reason to curb his actions. Since *we* are, for the most part, morally observant beings – i.e. beings that care about the moral dimension to life societal norms provide – constructed moral reasons can provide *us* with normative reasons to curb the would-be Caligulas of the world.

In one sense I am in agreement with Street and in another I am completely opposed to her. In keeping with my critique of Harman, I concur with her rejecting that the domain of the ethical is exclusive to the kind of 'inner-judgments' that Harman maintains they are. I disagree with her in that in the attempt to provide moral judgments with some desired objectivity and avoid relativism, she has sacrificed one of those characteristics I maintain is essential to morality – namely, their categoricity.

[22] Ibid, p.7.

It is at this point that I must make a small but valuable digression to discuss a little further the opposition between 'institutional' reasons for action with what Joyce dubs the 'genuine normativity' of reasons we touched on earlier. This will involve a little repetition but it is an important point, not only as it applies to Street, but it will also be a recurring issue in the discussion of certain internalist strategies to ground the moral.

In a section echoing Williams' observation concerning the causal role a reason for action must be capable of playing a role in explaining an agent's actions, Joyce writes,

> '[A]n adequate account of practical rationality must not leave an agent alienated from her reasons. If a normative reason could not potentially motivate an agent, then, if presented with such a reason, an agent could say "Yes, I accept that is a normative reason for me, but so what?" – and this, I have urged, is unacceptable.'[23]

If Joyce is correct, this calls into question the authority of what I've already referred to as 'institutional' reasons. If someone is playing chess, then the rules of the game apply to the players, regardless of whether or not they want them to. Now, of course, they can move the pieces any way they want on the board, but once they start to do this, they have ceased playing chess and instead are playing some bastardized version of the game. By analogy, so long as one is engaged in the enterprise of living in communion with others, it could be said that one has certain categorical reasons not to lie, cheat, murder or steal from others as this is to violate the standard rules and laws of civilized society.

However, to repeat the point Joyce makes, 'so what'? These institutional rules lack the 'normative oomph' we require moral norms to have. There would be no inconsistency in the practical rationality of an agent – e.g. the perennially problematic free-rider or Hume's sensible knave – who emulates moral-like behavior sufficiently well enough to garner the benefits of civilized society, but where it suits their needs, flouts them.

[23] Joyce (2001), p108.

> 'The fact is that the man who rejects morality because he sees no reason to obey its rules can be convicted of villainy but not of inconsistency.'[24]

Institutional reasons, though categorical in one sense, are not categorical in the *genuine* sense we desire that moral reasons be. Their normative authority derives ultimately from whatever reason we have to adhere to the rules of the institution in question – and our reason to obey the rules of any institution will always be contingent in some way on our desires to partake in that institution. Thus institutional reasons whilst categorical relative to the institution are ultimately hypothetical. Returning to Street once more – for her this presents no problem. The hypothetical reasons for action (including condemnation or approval of other's actions) the institution of 'morality' furnishes us with are sufficient for our purposes, she maintains.

I reject Street's constructivism and institutional reasons, in general, as a viable contender for providing genuinely moral reasons. Not only can they give no legitimate reason why the ideally coherent Caligula should change his ways but, more than that, I would argue that for any sufficiently perspicacious agent, simply becoming aware that ones 'moral' reasons for action are grounded by contingent custom, undermines them.

Take *Bushidō* – a form of Japanese chivalry. Under this strict code of honor, sometimes the only acceptable action by a *Samurai* who had committed what we in the modern West would probably regard as no more than a social *faux pa*, would be *seppuku* – i.e. ritual suicide. Now, such a call to action is entirely institutional. Yet its demanded action is extreme – literally a matter of life and death. My point is, were our dishonored *Samurai* to reflect deeply and clearly on the fact that he is expected to end his own life based on nothing but posited custom, *and nothing more*, there would surely be a high probability that they would decline to kill themselves and escape, assuming the option presented itself.

Morality makes some of life's most important demands on us. Pure institutionalism is not sufficient to ground moral reasons. To have the kind of normative sway over agents we need them to have, they must be genuinely, through-and-through, categorical – else they'd be too easy to dodge without the commission of any error in one's practical reasoning. To have the authority, the raw genuine normativity that we

[24] Philippa Foot, *Morality as a System of Hypothetical Imperatives,* reprinted in *Foundations of Ethics*, Edited by Russ Shafer-Landau & Terence Cuneo (2007), p288-289.

associate with moral reasons they must have the kind of normative grounding Street seems to take for granted as being provided by the things agents just do value. Joyce's point is that where it does make sense to ask 'so what' if a law or institution says I shouldn't *φ*, it makes little to no sense to ask 'so what' if *φ*-ing will bring me what I value.

Street has escaped relativism but at too high a price. The best reasons she can provide us will always be too weak and too dependent on the willingness of individuals to follow the mores of society without question. Who is to say we would not all be a lot happier if we were more like Caligula and adapted our customs to this end? I say any acceptable moral theory must be able to rule-out torturing for fun, by necessity, in order to be a moral theory at all. I see nothing in Street's argument that excludes Caligula's way of life writ-large, as a viable option for us to aspire to. I therefore consider it unacceptable.

1.2.iii Copp

The notion of the morality or immorality of actions existing outside of the normative force of an agent's purely practical reasons leads us nicely into the teleological relativism of David Copp. In his *Toward A Pluralist And Teleological Theory Of Normativity*, Copp 'sets out the basic elements of a 'pluralist' and 'teleological' theory of *normative judgment*'. His goal is to provide a pluralist teleological account of ethical truth that is,

1) Cognitivist – i.e. consists of factual assertions that aim at articulating facts, rather than merely expressing individual feelings or societal norms.
2) Realist – i.e. consist of assertions grounded in mind-independent facts about the world.
3) Naturalist – i.e. the real, mind-independent facts that ground the assertions rely on nothing outside of the ontology utilized by the natural sciences.[25]

The pluralism of Copp's view comes from his acknowledgement that there can be many different *kinds* of reasons. In any given situation, it may be entirely true to say that an agent as more than one valid reason to undertake a given action, or a valid reason to act in diametrically opposing ways. This might seem counter-intuitive at first glance, but

[25] David Copp, *Toward A Pluralist And Teleological Theory Of Normativity*, Philosophical Issues, 19, Metaethics, 2009, p.22.

only if one approaches the issue with the assumption that, for any given situation, there is some default, overarching way the agent should act, *simpliciter*. However, Copp maintains that the grounding of *any* reason for action is always relative to *some* normative system.

> '[T]here are "many and varied… normative systems for generating requirements." For instance, "there may be normative reasons of rationality, prudence, morality, and perhaps even normative reasons of other kinds as well." The pluralism I have in mind is a generalization of this pluralism about reasons. It holds that *all* normative statuses are 'generated' by normative systems, including kinds of goods, kinds of requirement, and so on.'[26]

Copp describes his pluralist teleology as being a 'relational view' of normativity. This is to contrast it with what he refers to as the 'unitary view' – which is simply the view I have already mentioned; that all genuine reasons are reasons unqualified or reasons *simpliciter*, with no dependence on their being related 'to any particular normative system' to ground their force. Hence, the theory is pluralist in that there may be as many normative reasons as there are normative systems to ground them.

The theory is teleological, on the other hand, in that different normative systems are intended for different purposes.

> 'The teleology of the theory is a generalization of an idea proposed by J. L. Mackie. Mackie says that, like Hobbes and Hume, he views morality as a "device" needed to solve "the problem" faced by humans because of certain contingent features of the human condition" (1977, 121).'[27]

Human beings have certain needs, which include society with other human beings and ways of living together harmoniously. These problems are varied and perennial, from generation to generation, society to society, and may more generally be referred to as the problems of 'Normative Governance' – and will invariably include moral considerations. For Copp, it is *essential* to practical reasoning that it be directed toward ameliorating the ubiquitous problems of 'sociality' (as Mackie dubs it).

[26] Ibid, p22.
[27] Ibid, p22.

Copp claims that the chain of argument that led Mackie to the Error Theory was flawed. Mackie, Copp claims, based his rejection of valid normative truths and/or moral facts on the mistaken belief that the only legitimate source of normativity would be one that aligned with the unitary view. Moreover, Mackie claimed that this would necessitate the existence of some non-natural features of reality that would have to have metaphysically queer properties. Since Mackie argued for the incoherence of such properties, he ruled that there could be no objective moral values.

If however normative facts could be grounded in the mundane (as opposed to queer) facts about the normative systems, or institutions, to use Joyce's terminology, that human beings partake in; a naturalist, cognitivist, realist account of morality could be established. This is Copp's goal.

So, to clarify, according to a pluralist teleology, *all* normativity is institutional. However, while institutional norms do have categorical-like characteristics, it is not, as I have argued above, the same kind of categoricity that I maintain moral reasons must possess. On the other hand, since Copp sees the problems of normative governance as being, in turn, effectively unavoidable for human beings, he might argue that the effective inescapability of morality's demands renders them sufficiently categorical-like, in the sense of being genuinely categorical. Even so, this form of categoricity of institutional reasons would remain ultimately inescapably hypothetical (*UIHoIR*).

Now there are various extrapolations of the implications to the status of moral reasons that can be made from Copp's chain of argument. A first reading could be,

Copp-1: 1) The idea of non-institutional reasons is incoherent.

2) (From 1) Non-institutional reasons do not exist.

3) (From 2) Any reason that exists must be institutional.

4) (*UIHoIR*) Institutional reasons are not unconditionally categorical.

5) (1-3) Moral reasons are institutional.

6) (4, 5) Moral reasons are not unconditionally categorical.

For the purposes of discussion, I am happy to concede (1-4). But even if these were granted, it would not establish (5), which is necessary to establish (6). Since I maintain that the kind of reasons we demand moral reasons be, necessitates genuine categoricity, I say that the 'moral' reasons Copp says pluralist teleology can furnish us with, are not moral reasons at all.

If moral rules were merely institutional, an amoralist or free-rider would be behaving quite rationally if they were to emulate moral behaviour for the most part, but shirk it when they felt that they could get away with it. There *are*, after all, sound practical reasons to violate the rules of chess if one can get away with it – for example, if there's the a chance of 'winning' the competition prize money in the process.

So, even if Copp's argument is sound, all he has done is re-establish the Error Theory and provide us with moral-*like* institutional reasons. This is no strike against my position.

But let's try an alternative reading of Copp and the implications of his argument.

Copp-2: 1) Moral reasons serve a certain teleological function.

2) Telos-serving reasons are institutional[28].

3) (1-2) Moral reasons are institutional.

4) (*UlHoIR*) Institutional reasons are not unconditionally categorical.

5) (3, 4) Moral reasons are not unconditionally categorical.

My chief problem with this formulation is with (1). I believe there are at least two different ways we can read this premise – one strong, one weak. On a weak reading of (1), we get only that moral reasons *can* serve a teleological function – i.e. that moral reasons happen to assist in the amelioration of the problems of normative governance – not that they *necessarily* serve such function.

I do not consider this reading as being any serious challenge to my position. Recall, it is my argument that it is *essential* for a moral reason to be categorical. There is nothing to say that any action an agent has a moral reason to carry out, might not also serve the additional function of ameliorating a societal problem. In other words, there is no issue with moral reasons and the practical reasons of normative governance sharing partial or even significant overlap. This is analogous to morality prohibiting murder and a law of the land prohibiting murder. If one is inclined, as I am, toward legal positivism, then one can see that a law and a moral can prohibit the same action but be grounded in significantly different way.

My point is that as far as any reason is merely institutional, it will never be genuinely categorical. On the other hand, there's nothing to prevent a moral reason, *qua* moral reason, being categorical, but also serving a purely institutional function at the

[28] This is implied in Copp by his position that there are no innate or unconditioned ends in the world. Ends or teloi are synthesized by institutions or systems of valuing that can be large-scale or based only on an individual's values.

same time. Since the weak reading does not establish that moral reasons are *merely* institutional, it therefore does not encroach upon my own position in the first place.

The strong reading however, is different. Taken this way, (1) demands that what it is to be a moral reason *is* that it serve a certain problem-of-sociality-ameliorating, teleological function. This is most certainly a challenge to my thesis. It posits as essential to the moral the characteristic of being teleological; hence institutional; hence non-categorical.

My rejection of this characterization of moral reasons breaks down into three parts. The first pertains to the highly counter-intuitive limitations this places on the legitimate scope of meta-ethical questions.

To recap, if Copp's position is correct, the nature of morality is purely utilitarian (with a small 'u'). It is a device employed by humans for the solving of the problems *they* face in the ongoing project of building, maintaining and improving the societies they live in and are conducive to the benefit of their members. So, by Copp's lights, the value, not only of ameliorating humanity's sociality problems, but of humanity itself, is a given. On this view the only way an ethical question would make sense is within the context of being of utility to humanity. However, what of the vast swathes of questions we intuitively hold valid and interesting but that lie well outside this context?

Take for example, our relationship or responsibilities to non-human animals. When the impetus behind our actions on their behalf is grounded on nothing but compassion, regardless of any teleological role these actions might have, what moral status do they have? Are questions like these to be relegated to questions of personal affiliation or taste, rather than being subject to serious ethical consideration?

More strongly, take an example like the Voluntary Human Extinction Movement (VHEMT) – a somewhat fringe environmental movement founded in the 1970s, which calls for human beings to cease to reproduce with the end goal going extinct in order to eliminate a major source of harm to the Earth's biosphere and sustainability. Now, regardless of where one comes down on these and similar questions, it seems obviously false to say that they are inherently unsusceptible to ethical scrutiny and investigation. 'Would it be a good thing if humanity ceased to exist?' or 'Is it morally right for human beings to stop procreating?' are both legitimate ethical questions. However, if ethics were only a question, ultimately, of what serves the telos of harmonious human societal

interactions, these questions, *by definition*, would have to be mooted! That's a pretty hard bullet to bite, and I doubt Copp would be willing to do so.

My second major reason for rejecting the strong reading of *Copp-2* follows on from the earlier discussion regarding the ultimately hypothetical, qualified categoricity of institutional reasons. Not to re-hash but I reject it for much the same reasons as I reject Street's constructivism. Copp offers no original argument for how institutions ground normativity with sufficient strength to avoid the same kind of 'so what' objection.

Third and finally, Copp seems to rule out by mere fiat the idea that there may be moral reasons for agents to do or not do certain things that serve no telos whatever. Even if we accept that some practical reasons can be grounded by purpose serving ends, he presents no sustained argument that this is the only source of normativity. I think we can easily generate examples from history of societies that were riddled with oppressive or unjust institutions, where there existed sound moral reasons for those systems to end or radically amend themselves. There is no reason to believe that a reason for action must serve some end in order to be a reason. To argue otherwise is akin to arguing that the only mathematical truths are ones with practical applications in engineering or science.

Copp's argument presents no serious challenge to my position that if moral reasons exist, they must be non-institutionally and genuinely categorical.

1.2.iv Prinz

The penultimate thinker I will be looking at is Jesse Prinz and his own form of moral relativism.

Prinz begins *The Emotional Construction of Morals* by asserting the following 'master argument'.

1) Descriptive relativism is true.
2) If descriptive relativism is true, then metaethical relativism is true.
3) Therefore, metaethical relativism is true.[29]

I have no qualms with (1). It is demonstratively the case that different people and cultures, across the world and throughout history, have considered different acts to

[29] Jesse Prinz, *The Emotional Construction of Morals*, Oxford University Press (2013), p.174.

be either morally required or prohibited (e.g. slavery, blasphemy) and have held different things (e.g. life, happiness, honour, status, family, the State) to be valuable in different ways. (3) obviously follows given (1) & (2). So the real business is clearly happening with the inference that is taking place in (2). For Prinz, the fact that people value different things implies that different things have different value. This is because he takes value as being constituted by nothing over and above *being valued* by beings that are capable of valuing.

Prinz's form of relativism is broadly Humean and even more unmediated than Copp's. For the former, it is not the role a moral rule plays within a functional moral system, that's telos is the harmonious functioning of society. It is simply whether or not people value or disvalue certain things and in a certain way – i.e. whether or not it produces within them sentiments of distinctly moral character.

> 'An action is right or wrong if there is a moral sentiment toward it. A moral sentiment is a disposition to have emotions in the approbation or disapprobation range. If descriptive moral relativism is true, then people have different moral sentiments toward the same things. If rightness and wrongness depend, metaphysically, on the sentiments people have, then the existence of differences in people's sentiments entails a difference in moral facts.'[30]

Prinz does not present much actual argument for his position in *The Emotional Construction of Morals*. He seems to see the *prima facie* fact that people just *do* value certain things as being a sufficient account of how it is that those things have value. One also gets the impression that he takes it as a given that any account of a source of moral value that lies outside of this straightforward, value-from-being-valued model, bears the burden of proof.

> 'The point is that we embrace our values because they are our values [...] The critic who assumes that values must be universal to be worth having simply begs the question against relativism. Matters of taste are a glaring counterexample. Psychologically, valuing does not require universalizing. It is not obvious that moral valuing should depend on absolutist assumptions.'[31]

[30] Ibid, p.175.
[31] Ibid, p.211.

This view is in plain opposition to my position that moral reasons must be categorical. For Prinz an action being morally wrong is entirely dependent on at least one agent having a sentiment of disapprobation against it. If the world were filled with sincere philistines, for example, who went about destroying historically priceless archaeological treasures, this not only wouldn't be wrong but *could* not be wrong.

My rejection of Prinz's view as an adequate account of moral reasons stems, in the first instance, from reflecting on the experience of *moral* approbation and disapprobation. Before I get into that though, I'd like to re-iterate what is at stake here. As I have already stated and will continue to state throughout this thesis, it is not my goal to demonstrate that there are moral truths. It is only my goal to argue that *if* moral truths exist they must have certain characteristics. I am quite happy to accept that the value of anything, in any sense, is entirely dependent on its being valued by some sapient being somewhere. What I reject and strive to argue against is that such a model could provide a bedrock for anything that could be called *moral* reasons for action. Now if this means that we are left having to reject the possibility of moral reasons existing at all, so be it.

So, lets consider moral disapprobation as being distinct from all other forms of disapprobation. One key feature of a moral sentiment is surely the intensity with which we reject an act or state of affairs – wanton animal cruelty, say – but there is also the way we reject it. It is not just that we dislike something. It is that this thing *must not* be allowed. Moral disapprobation seems to be its own call to action. It is in the nature of the sentiment experienced in making a moral judgment, unlike aesthetic judgments or others of mere taste, that it contains an implicit sense that it is grounds to expect or demand that other people likewise share and experience this same disapprobation.

The inherently absolutist phenomenological character of moral sentiments are at odds with Prinz's position that we are the ultimate source of them. '[T]hat we embrace our values because they are our values' is anathema to the sense that the things we value are worth valuing. I do not assert this as an argument in favour of value being grounded by factors beyond our own psychological make-up. Rather I say that moral sentiments contain an implicit aspiration to some form of objectivity that would be undermined or even destroyed by the immediate, simultaneous knowledge that it has no such objective grounding.

A point that Prinz himself acknowledges,

> '[O]nce we discover that our moral preferences are not privileged, our *confidence* in those preferences is destabilized. Why continue to embrace our moral values if they have no unique claim to truth.'[32]

It is this 'confidence' that we must have in the grounding of morality, if it is to have the capacity to demand the kind of self-interest-transcendent action we want it to, which necessitates the categoricity of moral reasons. Yet perhaps relativism can be salvaged from what I take to be Prinz's overly simplistic argument. In her *A Darwinian Dilemma for Realist Theories of Value* Sharon Street presents an anti-realist argument for moral truth – that is an argument for moral truths that are not independent of human minds or stances – and yet retains the sense of objective factuality that would allow us to make categorical statements about what others *should* do.

> 'Consider again the old dilemma whether things are valuable because we value them or whether we value them because they are valuable. The right answer, according to the view I've been suggesting, is somewhere in between. Before life began, nothing was valuable. But then life arose and began to value – not because it was recognizing anything, but because creatures who valued (certain things in particular) tended to survive.'[33]

Street is essentially arguing that asking whether or not valuing precedes value, or *vice verse*, is to create a false dichotomy. It is not simply the case that moral truths are nothing more than a question of what people value. The process of valuing and those things that *are* valuable have evolved side-by-side, hand-in-hand. Take the analogy with colour. The fact that colour – that is, our own subjective phenomenological idea of colour – is entirely the product of human (or possibly mammalian) visual perception, does not mean that there are not *facts* about colour. Although whiteness is not something strictly speaking 'in the world', it is false to say that pure calcium carbonate ($CaCO_3$) is purple when seen directly and in daylight. Colour-facts are not objective, since they would not exist without the existence of mammalian eyes and sufficiently sophisticated brains to conceptualize them. They are inter-subjective in a way that

[32] Ibid, p.206. My italics.
[33] Sharon Street, *A Darwinian Dilemma for Realist Theories of Value*, Philosophical Studies, 127, 2006, p155-156.

provides a standard of correctness regarding the application of colour-terms. Street argues for a view of value-facts that is analogous.

My response to this is to ask how Street's argument may be applied. On the one hand, taken in and of itself, I see potential for acceptability – at least in regards to the stipulation that moral reasons be categorical. Arguable, it can be said to be categorically false that calcium carbonate is purple when viewed under standard conditions. If a person genuinely sees $CaCO_3$ as purple we do not conclude a new colour-fact, we conclude that there is something wrong with that individual's vision. The moral analogue to the colour case would be that of a sociopathic individual, lacking all compassion for other human beings. This would not be an instance of an individual with a merely different outlook, but a *defective* outlook.

However, this too would be insufficient to ground morality. If we encountered an alien species with profoundly different sensory system, we could not convict them of any error if they failed to be sensitive to colour or cognizant of colour-facts, in the way we might reasonably expect all humans to be. Likewise, we may find their moral sensibilities profoundly different and yet we would not think that the moral reasons we intuitively maintain regarding respect for human life, failed to apply to them. We still need morality to be categorical in this way – a way Darwinian and anthropocentric models like Street can never achieve.

On the other hand, if Street's argument is used to bolster moral relativism, I would still reject this. Let's return to premise (2) of Prinz's master argument – 'If descriptive relativism is true, then metaethical relativism is true'. Recall, for Prinz, multiple mutually incompatible moral judgments implies multiple instantiations of moral truth. This is because moral truth is purely a matter of a moral sentiment being directed toward some act or state of affairs. However, for those such as Mackie (1977) and Joyce (2001), these multiple instances instead imply a 'global error'. Moral judgments are not merely expressions of moral approbation or disapprobation, but sincere attempts to assert moral facts about the world. If these attempts fail there are no actual moral properties, and hence no truth-makers for moral assertions.

I believe Prinz could well use Street's line of argument. He could suggest that the assumption that moral judgments are best categorized as unsuccessful attempts to make assertions about an objective moral reality is unfounded. Instead, he could say that they are better categorized as successful attempts to assert subjective truths about

valuing-value complexes. Then, given this, it is more plausible – in that it requires less of a violation of Ockham's Razor – to accept multiple instances of subjective moral facts than it does to posit a global-scale error.

However, for the reasons I have already stated, Prinz's moral relativism presents no real threat to my thesis. If he succeeds in establishing that what we describe as 'moral' judgments are merely the expression of the sentiments of approbation or disapprobation, then they are not truly moral as they are dependent on the hypothetical dispositional states of individual agents. Awareness of this contingency exposes moral reasons to an unacceptable degree of tenuousness. If, on the other hand, his arguments were to be deployed to create a set of species-wide inter-subjective moral truths, this would still not ground them with sufficient universality to satisfy our most basic moral intuitions.

Hence, even if successful, Prinz has not shown that moral reasons are grounded by sentiment, and so are non-categorical. He has merely shown that what he calls 'moral' reasons or sentiments are not in fact moral since any reason grounded in such a way could not actually meet or maintain the basic phenomenological character or integrity we demand moral reasons must have. His view is therefore no challenge to my thesis.

1.2.v Williams

I want to round up this collection with a brief mention of Bernard Williams, with specific reference to his views on ethics rather than his views on reasons more generally, which will be discussed in far greater detail at the top of the next chapter.

The cause for brevity is twofold – firstly Williams' specific views on ethics and how he distinguished it from what he called 'morality' are sometimes hard to pin-down and refine from his various writings. Probably the place it is most clearly outlined is in his *Ethics & the Limits of Philosophy*, particularly the final chapter. Secondly, what can be discerned about his views on morality, do not overlap perfectly with his thinking on reasons more generally and therefore has limited relevance to the current discussion on categoricity. However, I maintain that there is enough of an overlay for it to make it worth our looking at.

Williams describes 'morality' specifically, as opposed to 'ethics', as 'the peculiar institution'[34]. Though he doesn't precisely define what he takes 'ethics' to mean, I think it's fair to say that in his mouth it refers very broadly to the subject-area of philosophy that deals with the nature of values – i.e. the Good, virtue, proper human conduct, etc. Morality on the other hand, he believes refers to systems, philosophies or codes pertaining to *obligation* – i.e. what actions agents should or *must* undertake, and can legitimately be blamed for if they do not undertake them.

Though his discussion of obligation is not couched in the language of categoricity, or hypotheticality for that matter, I believe that Williams sees obligations very much as being things that have independent authority over agents. In turn, these obligations apply to them independently of their psychological dispositions or desires – and are hence always taken by those who talk of obligations as being categorical imperatives to act.

It is Williams' position that morality, *qua* an obligation-centered practice or thesis, is a peculiar institution. One that is both distorting and fundamentally wrongheaded. His view of reasons leads him to think that an agent's reasons are grounded or are exclusively dependent on the elements within an agent's subjective motivational set. As such, obligations, *qua* things an agent has overriding objective, desire-independent reason to comply with, are a nonsense. To Williams' mind they are ultimately stultifying to our attempts to resolve the many other issues and dilemmas that we encounter when discussing Ethics more broadly.

In the foregoing chapters of *Ethics & the Limits of Philosophy*, Williams lays out the groundwork of what might be described as a revisionist ethics, which is again hard to pin-down, but broadly virtue-based and in a similar spirit to the Aretaicism of Aristotle[35]. As I have already indicated, the finer details are not vital to the discussion here. The key point is that Williams believes that Ethics can function perfectly well, nee thrive, without anything that plays the role of categorical imperatives, which in turn provide categorical reasons for agent action. I believe this is an ultimately untenable position for any theorist who wishes to give such a revision of ethics as a whole.

My argument for this shall likewise be brief as I will be covering it in more detail later in this chapter when I discuss the conceptual place my three criteria play in moral

[34] Bernard Williams, *Ethics & the Limits of Philosophy*, Routledge (2006), p193.
[35] Ibid, p.10.

reasons. Suffice to summarize my problem with this is that ethics is irredeemably practical in its scope. Ethics must pertain to agent action (or inaction). Whatever the ethical system holds up as paradigmatic of valuable or representative of 'the Good', it is all for naught if actions that threaten, undermine or down-right destroy those things or states of affairs that are considered valuable, do not come-out as resoundingly and overwhelmingly unacceptable. To put it another way, if we accept that the highest virtue is compassion, say, and our ethical system is based around this fundamental insight – it must surely imply in the strongest way, at least to a large degree, that actions that make the furtherance, demonstration or manifestation of compassion impossible have to be prohibited by such an ethical system. If values or virtues carry with them absolutely no practical guidance, import or implication for the actions of agents, what possible use are they? I can make no sense of something's virtue or value if it can't fulfill this condition.

One does not need to be a reasons fundamentalist, as I am, for this point to be valid. It is sufficient to say that *any* ethical system that does not provide reasons for agent action or inaction is not fit for purpose. I have no particular objection if Williams or his ilk wish to use a different, less loaded designation other than 'obligation', but it will inevitably generate reasons for action or inaction – and if it does, I maintain that those reasons must be categorical in character.

I hope my reader will forgive my very short treatment of Williams' views on ethics here. I have included it in part simply because I think it is an interesting point worth mentioning. Additionally though, it does make an excellent prelude to the discussion later in this chapter. But first we must look at my other two criteria – weight and grounding.

In this section I have attempted to outline why I believe thoroughgoing, resolute categoricity must be an essential feature of moral reasons for action. Furthermore, I have tried to show why some of the arguments of the most obvious array of thinkers who would disagree with me on this score either fail to present a significant challenge to this stance or hold positions that are untenable.

1.3 Weight/Strength

Reasons feature in our deliberations about what to do – at least as far as they are rational deliberations. We have already discussed how reasons can be seen, at least as far as the current work goes, simply as considerations that count in favor of doing something. But not all reasons play the same sort of role in deliberation and also, they may not count *as much in favour* as each other and in all instances. This measure of to what extent a reason counts in favour of doing something is typically referred to as the weight or strength of the reason.

My enjoyment of chocolate cake gives me a reason to accept a plate with a slice of chocolate cake on it when I'm offered one. This enjoyment gives me the same reason to accept a plate with two slices on it. However, if I can only pick one of two plates it might be said I have a weightier reason to choose the plate with two slices on it as this will give me more of what I enjoy. However, it might also be argued that my reason to pick either plate are both roundly trounced by my long-term goal of being beach-body ready in time for my Summer holidays, or just maintaining good general cardiovascular health. Yet still further, as weighty as my reason for maintaining good physical health is and as relatively non-weighty (or 'light') is the short-term sensual pleasure of eating some cake on this occasion; given that the consumption of chocolate cake, if done infrequently, does not seriously impede the maintenance of good physical health, then, all-things-considered, I might have the strongest reason to enjoy the cake – but perhaps only the one slice after all.

This highly simplified example of rational deliberation features reasons that seem to have different inherent weights in their own right. However, when compared the one against the other in a specific context, these reasons contribute in different ways to the agent in question (in the example given, me) arriving at what they have most reason to do, all-things-considered.

There are different types or categories of reason that interact and intersect in diverse ways throughout our day-to-day lives. There are pragmatic reasons to look after your health, for example; hedonistic reasons to seek pleasure from life, or at least avoid ennui; personal reasons to look-out for the welfare of friends or family members; legal reasons for observing the law of the land in which you find yourself; aesthetic reasons to preserve artworks or buildings against decay or destruction; intellectual reasons to

give certain evidence or arguments priority over others in establishing the truth; and it goes on and on.

In addition to all these, we also think that there are moral reasons to do and not do certain things. We have moral reasons not to lie and to keep promises/contracts when entered into freely. We have moral reasons not to steal out of sheer greed. We have reason not to murder or be violent toward the innocent, and reason to help those in desperate need, and so on. Intuitively, moral reasons have different weights in relation to each other – i.e. *ceteris parabus* it is worse to murder than to steal, and worse to steal than tell a lie, etc. Also, it is accepted by a significant number that moral reasons don't always have the greatest weight in all situations – i.e. moral reasons aren't a 'trump-card' that must be given automatic precedence in deliberation. Sometimes they can be overridden by other considerations. The classic example of course is an agent who, all-things-considered, has the strongest reason to steal food from someone who has more than enough, to feed their starving family. Another could be an airman having reason to collaterally kill innocent people in the waging of a just war.

But despite this, we think of the class of moral reasons itself, as classes of reasons go, having a *prima facie* weight to be reckoned with. By default, a moral reason for action is one that should *never* be merely dismissed as trivial, even if it is eventually outweighed through a process of deliberation. Any successful moral theory must provide a satisfactory account for why the class of moral reasons has this default weight and what accounts for the differences in the relative weight of moral reasons to each other.

I do not wish to stack the deck against Internalism from the outset. I do not insist that to be fit for purpose and internalist theory must give a thoroughgoing account of *how* the weight of moral reasons is grounded. But I do insist that a successful internalist moral theory must, at the very least, must be able to provide moral reasons that actually *have* sufficient weight so that only the very strongest countervailing, non-moral reason will be capable of outweighing them.

I will have a little more to say on the subject of weight in Section 1.5 below. However, the most in-depth examination of this feature I consider essential for providing moral reasons will come during my treatment of the work of Mark Schroeder's Hypotheticalism (See Chapter Four).

1.4 The Right Grounding of Moral Reasons

In addition to the categoricity and weight conditions, I think that moral reasons must meet an additional condition that, for want of a better term, I call the 'right grounding' condition. This is hard to express and to be honest remains the least formally defined of my three criteria. Very little rigorous or comprehensive treatment of this exists in the literature. Nonetheless, I believe moral reasons having this quality remains a *sine qua non* of any satisfactory moral theory.

Before I get into the meat of this subsection I'd like to say, on a terminological note, I took some time in deciding what precisely to dub this condition. In many ways, what I refer to as the 'right grounding' condition is strongly analogous to the *right kind of reason* point so often discussed in Epistemology. If some interlocutor is hooked up to an infallible lie-detector, has a loaded gun pointed at their head and is told that if they don't start to sincerely believe in the Lock Ness Monster they will be shot on the spot – then, all things being equal, the agent has a genuinely very good reason to believe in the Lock Ness Monster. However, we do not think this is the right *kind* of reason to believe in the Lock Ness Monster, or indeed to believe in anything. There are reasons that are appropriate for supporting belief formation and those that are not, even where the latter provide strong reasons for belief formation in another sense.

In much the same way, I draw a fundamental distinction between a reason to behave morally on the one hand, and what I think should be called a moral reason on the other. If a child is drowning in a lake there night be any number of very strong reasons that an able-bodied passer by should risk their own life by jumping into the water to save them. *Any* of these reasons will count as a reason to behave morally – assuming that it can be correctly said that saving the child is the morally right thing to do. However, my contention is that only some of the reason will count as *moral* reasons. For example, if the passer-by has a strong respect for the value of all human life, or a sense of duty to prevent harm, or wills the promotion of general happiness, well-being or flourishing regardless of whose it is – I think most of us would typically be inclined to say these are apt reasons to be classed as moral in character. If though, the passer by is motivated to risk their own life solely as a means to gaining fame, glory or some monetary reward, again, I think most of us would not consider these as being legitimate grounds to class them as *moral* reasons to save the child.

Suffice to say, I did consider calling this condition as the 'right kind of reason', but decided to call it the right grounding condition to avoid inheriting some of the intellectual baggage that might be associated with it's epistemological name-sake. I make a special point of this here only by way of dissuading the reader from reading too many metaphysical implications into the condition. The particular mechanics of the grounding of a moral reason will vary significantly from moral theory to moral theory. However, I think it fairly uncontroversial to say that for almost any given moral theory there is an appropriate distinction to be drawn between a moral reason and merely a reason to behave morally. This is typically, though not universally, brought out by a given moral theory's account of the appropriateness of praise and blame in a given situation. I am not aware of even a single moral theory that would consider an agent who risks their life to save a drowning child for no other reason than out of concern for the child's welfare, as being more worthy of moral praise – at least on *some* level – than an agent who does the same but for no other reason than financial gain. It is not just any reason that can count as any kind of moral reason – no matter how strong a reason it truly is to act morally.

Now, if I were asked to provided a more precise *ti esti* definition of the right grounding for moral reasons, I must confess to not being able to provided a hard and fast one. For this I must look to a future work. Here I can do little more than offer examples of the kinds of things we typically associate with motivations of an acutely moral character. Typically, they are earmarked by a willingness to transcend ones own self-interest. They are also often motivated by a sense of decency, duty or the demands of righteousness. They are typically associated by other-regarding sentiments like love, compassion, or empathy. The list is potentially endless and varies greatly from moral theory to moral theory.

Parenthetically however, I might tentatively offer that for any given moral theory there will be some actions or inactions that the holder of the theory in question will always want to come out as being those that agents will tend to have a strong reasons to do or not do. For example, I can't imagine a moral theory that wouldn't insist as a prerequisite that torturing a child for fun must be morally forbidden. If we delve deeper into that theory, we will surely discover some basic justification for why the given theorist always wants child-torture to be ruled out. I would suggest then that there will be something quintessential to child-torture and other actions that make them what

they are. There is something about murder specifically that makes it *murder*, rather than, say, an act of justifiable homicide. I would go on, again tentatively, that it is these quintessential details about certain actions that lead us to want to rule them out preemptively that should play a role in providing the explanation as to why an agent always has a moral reason for not doing it.

Thinking in terms of counterfactuals, I might make the above point in the following way. There are many possible worlds in which saving an innocent child's life, at little risk or cost to ourselves, does not align with our self-interest or is not mandated by the social contract. However, there is no possible world in which, all things being equal, an agent does not have a moral reason to save an innocent child's life. If the grounding that provides a reason for an agent to carry out the morally correct thing could potentially be missing in some scenario, then I would argue that this is indicative that that theory has not provided a truly moral reason, just a reason to behave morally that happens to cover a large variety of cases. To avoid any possibility that the moral reason *not* to murder someone could be absent, therefore, what is essential to making a murder a murder would have to do its own heavy-lifting, so-to-speak, in terms of providing the grounding for the moral reason that exists not to carry it out. However, the points made in this and the previous paragraph are only speculative and are not crucial to any arguments that will be made in the rest of this thesis re: the grounding condition.

As I have said, a clearer exposition of the link between the quintessential character of certain actions and the distinctly moral reasons we have to do or not do them will have to remain the subject of a future work. For the purposes of this thesis, when it come to the right grounding for a moral reason I must repurpose the principle as cited by Justice Stewart in his discussion of how to identify pornography of 'I know it when I see it'! I do not see the near-perennial and foregoing intuition that there is something wrong with a moral theory that does not forbid or permit certain actions as being anymore reliable than the intuition that only some types of reason can count as moral, rather than merely a reason to act morally. Just as any adequate metaphysical theory must, in the final analysis, provide a theory that adequately incorporates our first-order intuitions regarding the existence of everyday objects like tables and chairs, mice and stars, so must any adequate moral theory go at least some way to explaining why some agent's motivations to do the morally right thing are more worthy of either

praise or blame than others. For this reason, absent a more formal definition of the right grounding of moral reason, the intuitive test for what counts as a moral reason should at least be taken into consideration when assessing the overall success of a moral theory in having provided them.

In closing, just to get ahead of a potential criticism that may already be in the mind of the reader; it might be objected that the way I am framing the right grounding condition begs the question against Internalism from the outset. For if a certain desire is a necessary condition for a given agent having a reason to act morally might it not always be possible that an agent could lack that desire, and hence the moral reason? Here I am reminded of some of the points made by Prichard in his *Does Moral Philosophy Rest on a Mistake?*

> 'Now, how has the moral question been answered? So far as I can see, the answers all fall, and fall from the necessities of the case, into one of two species. Either they state that we ought to do so and so, because, as we see when we fully apprehend the facts, doing so will be for our good, i.e. really, as I would rather say, for our advantage, or, better still, for our happiness; or they state that we ought to do so and so, because something realised either in or by the action is good. In other words, the reason 'why' is stated in terms either of the agent's happiness or of the goodness of something involved in the action.'[36]

> 'And this process seems to be precisely what we desire when we ask, e.g., "Why should we keep our engagements to our own loss?" for it is just the fact that the keeping our engagements runs counter to the satisfaction of our desires which produced the question.'[37]

Now, while the right grounding condition is clearly more associated with what Prichard might call the intrinsic 'goodness' of the thing rather than the 'advantage' it brings to an agent, there is no *a priori* reason to believe that the internalist can't meet this requirement. All that is required for them to do so is to show that agents have motivational states that provide them with reasons to do the morally right thing that is at least partially grounded by the thing's goodness and not exclusively by some advantage to the agent.

---

[36] H. A. Prichard, *Does Moral Philosophy Rest on a Mistake?* Published in *Mind*, New Series, Vol. 21, No. 81 (Jan., 1912), Oxford University Press on behalf of the Mind Association. p22.
[37] Ibid. p23.

As we shall see in Chapter Four on Mark Schroeder, and even more so in Chapter Five on Christine Korsgaard, there are perfectly viable options open to internalists to meet the right grounding condition in the aforementioned way. In the case of Schroeder, they make some reasons for action by their very nature so difficult to avoid that an agent can't avoid having them; or in the case of Korsgaard, constitutional of agency itself. All the right grounding condition stipulates is that at least one of the perennial reasons generated by such a method go a considerable way to meeting our foregoing notion of what makes the action the type thing we always want agents to have strong reasons to do – or as Prichard might say, our notion of what makes it morally good. So, if we are so constituted that we always have some motivation to behave fairly and honestly toward others or a reason to can't avoid be desirous on some level to behave compassionately, this would be quite adequate to meet the right grounding condition. As such, I do not believe the right grounding condition does not demand too much from Internalism and is a perfectly reasonable requirement for any acceptable moral theory.

1.5 Conceptual Requirements vs. Commonly Held Intuitions

As I have already stated, I take the three criteria just discussed to be integral to any understanding of what we would or should classify as a moral reason. However, something that needs to be discussed with greater specificity is whether or not these criteria are what might be called conceptual requirements of moral reasons, or just perennially and strongly intuitively held characteristics. This distinction is not widely covered in the existing literature and the following section is based entirely on my own thinking on the subject.

This detail will be of greatest importance when it comes to deciding whether or not Internalism has some endemic feature or features that make in constitutionally incapable of meeting criteria moral reasons *must* have; or alternatively, whether there is the option to amend our intuitions so that the kinds of moral reasons internalists *can* furnish us with become more palatable. In other words, if internalists can't meet all three criteria, is that necessarily a shortcoming of Internalism that should lead to our rejection of it? Could the fault instead be with us? Are our own tacitly accepted normative concepts poorly formed, naïve or otherwise unfit for purpose, and should be changed accordingly?

Let's go through them in order, starting with categoricity. Can the notion of a moral reason be consistently formulated where it may not apply to an agent by virtue of that agent having or lacking some desire or motivational state? Striping it down to its simplest form, to believe in the non-conceptual necessity of categoricity in moral reasons you would have to accept something like,

> *Intelligible Non-Categoricity of Moral Reasons: x* could have a moral reason to φ in situation *A*, but would not have a moral reason to φ in situation *B*, where the *only* difference between situations *A* and *B* is the motivational states possessed by *x* in each.

In other words, for a given agent and a given moral reason they might have, can we envisage it ever being the case where, *ceteris parabus*, a mere change in their motivational states might mean that they lack that reason? If the answer could conceivably be yes, then moral reasons are not necessarily categorical. In which case it might well be possible for a theory to provide moral reasons that are hypothetical and thus sidestep the need to meet this criteria – at least to some degree.

To respond succinctly to the prospect of such a possibility, I say that for any reason for action we could imagine having any hope of fulfilling the basic functions and usages of moral reasons in everyday human interactions, it could not be. In society, public discourse, politics, law or philosophy, it could not be possible for a moral reason to be denied application to an agent based only on the motivational states (or lack thereof) of that agent. To see this just think how moral reasons are employed.

Citing the moral reasons agents supposedly have to act in certain ways, yet oftentimes fail to live up to, is what justifies the quintessential type of censure, blame condemnation and punishment that accompanies moral judgements when they do so fail. Our standards for the correct application of moral judgements, for example, do not take the motivational states (the 'feelings') of the judged as a factor.

The chief reason for this is what I take to be the *telos* of moral deliberation and judgement. The role of the telos in moral reasoning should not lead the reader to think that this is only true of consequentialism – my point applies to de-ontology as well. If an agent enquires *why* they have a specifically moral reason to act some way, then the answer provided is almost always couched in terms of the achievement of some goal. They are told that the fulfillment of some obligation or duty, or the attainment of some

state of affairs that itself exists quite independently of the agent or their psychological state is what is important. You should save the drowning child to secure the wellbeing of the child, for example. The wellbeing of the child is a *telos* not constituted by the agent who saves its motivational state. Hence, it is the promotion of the *telos* taken in-and-of-itself or the failure to achieve it, or even attempt to achieve it, which regulates ascription of moral reasons to agents and the corresponding moral judgement, praise or condemnation that typically goes with it. To put it another way, when speakers (or moralizers) ask themselves questions about whether the ascription of moral reasons is being done properly, when moral reason attributions have been dealt out correctly or if a mistakes have been made, they are invariably doing it with an eye fixed on some *telos* that it is quite apart from the motivational states of the agents to whom the ascriptions or attributions are made. The categorical nature of moral reason ascriptions is implicit in the way they are applied.

Because the *telos* of moral reasons are exclusively constituted by factors in which the motivational states of agents who promote them play no necessary role, the moral reasons for action that they supposedly give rise to *can't* be conditional upon them. To reject this is to totally alter the typical function of what the ascription of moral reasons is, and how they are used. Ultimately, allowing for the conceptual possibility of hypothetical moral reasons is as absurd and unworkable as voluntary laws – e.g. 'the law says you must pay your proper amount taxes… unless you personally don't feel like it'. For this reason I consider categoricity to be conceptually essential to anything we could reasonably call a moral reason, while it remains moral in any meaningful sense.

The remarks I have just made regarding the *telos* of moral reasons and their importance in considering the conceptually essential role of categoricity dovetail nicely with the criterion I've been referring to as *the right grounding* of moral reasons. I shall therefore discuss this feature next and leave the conceptual status of weight until last.

As I have already written, the right grounding criterion is the least formally defined of the three. It has little support in the current literature and is hard to pin down exactly, but I believe most reading this will understand what is meant by it. To reiterate, there is a fundamental and salient distinction between a reason to behave morally and a moral reason. The difference is the *kind* of reason we insist that a moral reason ought to be. There may be any number of reasons to behave morally such as to avoid public condemnation, shame, guilt or punishment. However, I have suggested that

the grounding of moral reasons specifically must be something that tacitly and typically involves consideration of something that is outside of purely selfish concerns or, as Prichard might say, more in keeping with the sentiments of benevolence and compassion. This is often indicated by our intuitive tendency to treat some reasons or motivations that provide reasons as being worthy of praise and those being less worthy or unworthy of praise, or even worthy of condemnation. It is implicit in the explanation of moral reasons that are given to any agent that questions the moral reasons that apply to them, any *telos* that is cited as the grounding for a moral reason must be something of this particular kind. Any old grounding simply won't do.

The right grounding is typically altruistic in character or involves invoking the duties or obligations an agent has according to standards that are fixed outside of themselves. However, the crucial point is this, when agents are deemed to have failed to act in accordance with their moral reasons, as opposed to reasons of etiquette, pure pragmatism or self-interest, it invokes in their moral judges a peculiar kind of ire, disapprobation or even censure. This peculiar kind of ire is indicative of a reverence held for, or value placed on, to use less emotive language, the telos not served or attained, which transcends the value placed merely on the agent who has failed to serve or attain it. Again this echoes Prichard's point that the moral question is typically answered in one of two ways – one that refers to some boon obtained by the agent, and another that refers to the value of the thing attained.

Think of the kinds of things moral edicts prohibit. They almost invariably serve the kinds of things people and communities ascribe the highest value to. *Moral* condemnation has the character it does because of the thing valued that has not been served, rather than the qualities or characteristics of the agent that fails to meet them – unless of course the former is the same as the latter as is the case in Virtue Ethics. However, even in the case of Virtue Ethics, the virtues, *qua* mere characteristics of agents are not in themselves virtuous. Cases made for the value of the virtues almost invariably involve arguments for their intrinsic value.

To summarize my main point; moral reasons are grounded by factors implicitly held by moralizers to have value that can be considered *apart from* the agent in-and-of themselves, when being considered only as far as their capacity as agents. That being said, while I do believe that this implicit feature of moral reason attribution and indeed, moral judgement and condemnation, is perennial to moral reasoning, I will not make

the case here that it is actually a conceptual necessity. Though I do not deny the possibility that the right grounding may well turn out to be a conceptual necessity, as I have already argued categoricity must be, I do not currently have firm enough arguments to establish it. I shall confine myself then to making what appears to be the more plausible case that it is no more than a commonly and strongly held, though often tacit, intuition in attribution of moral reasons.

In a nutshell, my argument for this is the tremendous diversity of types of groundings explicitly invoked in the citing and explanation of moral reasons. In Utilitarianism it is the maximization of the psychological states of pleasure or happiness that ground normativity. Yet in something like Divine Command Theory it is the will of God. The groundings of moral reasons are as myriad and diverse as moral theories are. Given the enormous diversity of differing types of groundings with which speakers (or moralizers) seem implicitly comfortable with when ascribing moral reasons to agents, it does not seem reasonable to argue that any specific type of grounding is actually a necessary conceptual requirement of a reason being a moral reason.

To make the case that it is a necessary conceptual requirement (which, remember, I am still quite open to) I would have to establish one of two things. Either I would have to argue that when people ascribe moral reasons to people, they have a specific type of grounding in mind – i.e. they actually equate moral reasons with utility-serving or God-serving reasons. In other words, typical moralizers and moral language users would have to be committed, tacitly or otherwise, to an identity relation with the moral reasons they ascribe and the things that ground them, in order to feel confident in their ascriptions and competent in their language use. However, there is no evidence I can see that such implicit identity relations are prevalent among ordinary moral language users. Moralizers can feel great confidence in the moral judgements they make, without it being necessary for them to have also simultaneously judged that it is because of a specific grounding. In other words, ascribing a moral reason and ascribing the presence of some grounding seem to be entirely separable cognitive and linguistic operations.

The second option for arguing that right grounding is a necessary conceptual requirement would be to argue that ordinary moralizers and moral language users are somehow aware and seized of the very kinds of arguments I am making here. While they may not have a *specific* type of grounding in mind they are aware that the veracity

of their moral reason ascriptions is somehow dependent on their being *some* grounding – i.e. that the truth of any moral ascription claim rests on there being a grounding of some kind and that this grounding will ultimately act as guarantor of their claim.

However, once again there is no evidence that this is the case for ordinary ascriptions of moral reasons to agents. Moralizers make reason ascriptions and moral judgements confidently and competently without implicit knowledge or commitment to anything like the kinds of arguments I make here. Furthermore, there is no reason to believe that exposure to arguments concerning grounding, or lack there or, would necessarily undermine a moralizer's confidence in the ascriptions and judgments they make.

For these reasons, I can't ultimately argue in the current work for anything stronger than that the stipulation that moral reasons have the right grounding is more than a very commonly held intuition. However, I would caution against dismissing the notion that the way moral reasons are grounded my have some conceptually necessary dimension that is worth further thought.

That leaves the question of weight. To put it simply, as with the right grounding condition, the sheer diversity of views concerning how moral reasons have weight and how they measure up when weighed against non-moral reasons means that I don't believe any specific account of weight should be taken as a conceptual necessity. However, what I do take to be conceptually essential to anything that could count as a moral reason is not *how* it has weight, but simply that it *have* weight – specifically that it have non-negligible weight. My position on this can be divided into two different claims; one strong, one weak. First the strong.

My strong claim is that trivial moral reasons, when taken individually, are a nonsense. There would be something profoundly wrong with the idea of a moral reason that could ever occupy a low position in the hierarchy of reasons that a rational agent should be considering when deliberating a course of action.

This is not to say, as I have earlier, that that moral reasons must always trump non-moral reasons. The strong claim is not that it is a conceptual necessity in order for something to be a moral reason it must have sufficient weight to trump *any* non-moral reason. The strong claim is that there is never a time when it would not be inappropriate to exclude ones moral reasons when weighing up a course of action. For example, if we are deciding the best way to set up a scientific experiment, it might be

entirely appropriate that some classes of reasons – e.g. aesthetic or political – whilst real, could be set to one side. They might be dismissed as irrelevant for the current task. Similarly if one were trying to create a work of art, some pragmatic or etiquette-based reasons might easily be set to one side.

It's not that these considerations are being outweighed – they're just being ignored. My strong claim about moral reasons however, is that this would never be the case with moral considerations – or at least, there is something wrong with agents that did completely ignore their moral reasons. If there is any moral consideration to a task or situation it is never appropriate to set them aside by fiat. Moral reasons are not an optional extra in this sense. Wherever moral reasons are salient to some deliberation, they must be ubiquitous. If they are significant *at all*, it could not be the case that they are not weighty – even if they do end up being outweighed. The strong claim is that there is something inconsistent about the notion of a totally trivial moral reason.

The strong claim though, is perhaps a bit too strong. Though I believe it is for the most part correct, it is all too easy to think of legitimate counter-arguments. Maybe a reason we would feel obliged to dub moral *could* in fact be genuinely trivial. An example of this might be that I have a moral reason to give a homeless person the one-penny piece I have in my pocket. Yet, given the almost negligible difference one pence will have on anyone's life in today's economy, it is legitimate for me to treat this moral reason as having negligible weight. The debate could go back and forth, and I do not wish to spend time and space here defending the strong claim when my weaker claim is far easier to defend and will serve my purposes just fine.

My weaker claim is just this; it is a necessary conceptual requirement for a theory of moral reasons, taken as a whole, that it be able to generate *at least some* moral reasons that are of non-negligible weight. There is something profoundly wrong with a theory of moral reasons that is incapable of generating at least one reason which can't be dismissed as trivial or outweighed by only the very strongest of non-moral reasons.

My argument for the weaker claim borrows from my previous arguments regarding the telos of moral reason ascriptions. A system that can't generate a single reason strong enough or weighty enough to outweigh non-moral reasons of only minor or middling weight could not secure the achievements of moral teloi with anything like the kind of regularity or security we would demand of a theory of moral reasons. A

moral theory generating only moral reasons of negligible weight would simply not be fit for purpose.

This weaker claim I think far better to the current task. I can't think of a counter example of a theory of moral reasons that exclusively fails to generate a single weighty moral reason. Secondly, that it is a necessary conceptual requirement only that a theory of moral reasons be able to generate *some* weighty reasons is a sufficient burden to place on a theory. It shows that the capacity to generate weighty moral reasons is a requirement, but does not set the bar too high from the outset. It asks of a theory only what is essential and no more.

So to sum up this section, that moral reasons be categorical and (that at least some of them) be of non-negligible weight are both necessary conceptual requirements. This means that any theorist who would seek to side step some of the problems I will be discussing with Internalism or metaethics more generally must incorporate them into their amended model in a satisfactory way.

Although I do not argue here that the right grounding condition is a conceptual necessity of moral reasons I would argue that it is so strongly held an intuition. I would also argue that any moral theory that amends its normative concepts to make it entirely redundant, would have its work cut out for it and must provide a satisfactory explanation for why this intuition, which is so strongly held, is not in fact essential.

The possibilities of employing different normative concepts to solve these problems, or suggest a fruitful new avenue to eventually do so will not be taken up again until Chapter Six.

1.6 Where we go from here

In the forgoing sections, I have tried to explain what I think a moral reason, as opposed to simply a reason to behave morally, must look like. It must be categorical, sufficiently weighty and have the right grounding. *Any* moral theory that can provide reasons that meet these criteria is acceptable to me. However, in the next chapter I will undertake the separate task of trying to explain why meeting these criteria presents a *prima facie* problem for any moral theory that considers itself internalist.

Chapter Two

A *Prima Facie* Problem for Internalism

2.1 Introduction

This chapter is short. In Chapter One I outlined what I maintain are the essential characteristics a reason for action must have in order for it to meet our basic, implicit desiderata for it to be specifically a *moral* reason. In this chapter, I am going to explain exactly why moral reasons, as I have described them, present a fundamental problem for the reasons internalist.

Section 2.2 will highlight some important distinctions and clarifications that need to be made before we proceed further into the discussion concerning Internalism *per se*. In Section 2.3 I will go through exactly why Internalism faces a *prima facie* problem meeting each of my three criteria for moral reasons, and to what extent. Section 2.4 will outline exactly what an internalist moral theory will have to achieve in order to have successfully, or even have a shot at, providing genuinely moral reasons.

2.2 Why Are Some People Internalists?

We generally accept that agents have reasons for doing some things and reasons for not doing others. The Internalism/Externalism debate concerns one dimension of theorizing about the reasons agents can and do have and why they have them – i.e. why does one agent have a reason to do something where another agent might not? Though there are many different aspects to philosophical theorizing about reasons for action, this debate is one of the most important and where a theorist stands on this issue will have profound implications for almost all aspects of their thinking about reasons generally.

It is widely accepted that the debate in its modern form began with Bernard Williams' *Internal and External Reasons*. Here Williams makes the case that in order for it to be true that some agent has a reason to do something it must be possible for an agent to act *for* that reason. In order for an agent to act for a reason it must be possible for that reason to be able to motivate them. To put it another way, it must be able to play some role in explaining an agent's action or potential action. Hence, Williams posits Internalism chiefly because of its explanative advantage over Externalism. The former offers a straightforward account for the idiosyncratic connection between agents and

their reasons. The latter has the *prima facie* more challenging problem of explaining how agents can be said to have the reasons they have.

So, for any agent's voluntary action, it is appropriate to ask *why* they acted in that way. If the answer takes the form of stating a reason that they had to do so, then it would seem that a crucial characteristic of reasons is that they play some role in accounting for action – i.e. reasons have to be able to serve as or contribute to the causation of voluntary action. Furthermore, Williams takes it that what ultimately causes agent's voluntary actions are their motivational states. Motivational states are those aspects of a person's psychological make-up that motion them toward action.

> 'If there are reasons for action, it must be that people sometimes *act for those reasons*, and if they do, their reasons must figure in some correct explanation of their action (it does not follow that they must figure in all correct explanations of their action).'[38]

> '[N]othing can explain an agent's (intentional) actions except something that motivates him so to act.'[39]

For Hume, motivational states were limited strictly to desires. Famously, and highly influentially, he held that it was desires alone, as opposed to beliefs, that could actually motivate action. Williams makes no such narrow commitment and acknowledges that there are a broad array of items in a persons psychology that may motivate them to act. He dubs the totality of items within a person that could potentially motivate them to act as the agent's 'subjective motivational set' (SMS).

According to Williams, an agent's subjective motivational set can also include 'dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent'[40]. He refers to this picture of action motivation as the 'internalist model' and sums it up thus,

---

[38] Bernard Williams, *Internal & External Reasons*, re-printed in *Moral Luck*, Cambridge University Press (1999), p.102. My italics.
[39] Ibid, p.107. My italics.
[40] Ibid, p.105.

> *'A* has a reason to *φ* iff *A* has some desire the satisfaction of which will be served by his *φ*-ing. Alternatively, we might say… some desire, the satisfaction of which *A* believes will be served by his *φ*-ing;'[41]

> 'Basically and by definition, any model for the internalist interpretation must display a relativity of the reason statement to the agent's *subjective motivational set*, which I shall call the agent's *S*.'[42]

Williams uses the now famous example of someone thinking that the glass in front of them contains gin & tonic, when in fact it contains petrol. Because the agent in question desires to drink a gin & tonic this desire could motivate them to drink it. However, William's does not wish to say that they have a reason to drink the contents of the glass because the agent wants to drink gin & tonic, and the glass actually contains petrol. Williams' solution is to say that the only reason they have this desire is because they are in possession of a false belief about what the glass contains. If they had no false beliefs and were deliberating soundly (i.e. thinking and reasoning about their actions clearly and correctly), they would cease to be motivated to drink what is in the glass. For Williams, motivation to act *simpliciter,* is not enough to give an agent a reason to act – only motivations they have in the absence of false beliefs and when deliberating soundly.

Furthermore, it is not necessary, by Williams' lights that agents actually *are* motivated to act as they have reason to – only that they would be motivated to act that way if they were, again, deliberating soundly and were not in possession of any false beliefs that were salient to the matter they are deliberating about. An agent might have a reason to act that they aren't even aware of. The crucial point is that what could motivate them under the right circumstances is what makes something a reason for the agent to do it.

To be fair to Internalists, the capacity of reasons to play a role in *causing* action is not universally accepted by internalists. Mark Schroeder uses the example of someone who loves surprise parties who, as a result, would therefore have a reason to go somewhere where one is going to be sprung on them. By its nature this reason can't cause the agent to go to the party, for the reasons force is constituted, at least in part, by

[41] Ibid, p.101.
[42] Ibid, p.102.

them not being aware of it. It is a reason they have to go to the party that they can't act *for*. What is essential is only that the reason bears some relation to their subjective motivational set.

So, I take Williams' basic argument to run as follows,

1) For something to be a reason for action it must be capable of playing a role in explaining an agent's actions,
2) To be capable of playing a role in explaining an agent's action, this something must be capable of playing some causal role in an agent's actions – i.e. be capable of motivating an agent to act,
3) For something to be capable of playing a causal/motivational role in an agent's actions, it must be or be connected with an agent's desires (or some other contents of their subjective motivational set).
4) For a reason to be categorical it must be possible that an agent has that reason whatever the agent's desires (or some other contents of their SMS) are.
5) (1-4) No reason for action can be categorical.

I will return to the implications of (5) later in this chapter. But I see the fundamental motivation of Williams' account and indeed the most appealing feature of Internalism over Externalism is what might be called its drive toward the demystification of reasons from more or purely abstract entities, to ones more firmly grounded in the natural, palpable facts regarding human psychology.

> 'If something can be a reason for action, then it could be someone's reason for acting on a particular occasion, and it would then figure in an explanation of the action. Now no external reason statement could *by itself* offer an explanation of anyone's action.'[43]

A key problem for the externalist is telling a convincing tale as to how precisely agents come to have the reasons they do. The defining trait of an externalist theory is that the reasons an agent has is in no way limited to what can motivate them. An agent could, in principle, have a whole host of reasons for action that the agent in question has absolutely no interest in fulfilling, and never would do. Yet for the externalists they are reasons for that agent nonetheless. They are comfortable with the notion that agents

[43] Ibid, p.106.

can just have reasons, regardless of anything whatsoever contained in their subjective motivational set.

Different externalists offer different accounts of exactly how an agent can have reason to do something even if that reason would not facilitate or promote anything an agent might want or care about. Indeed, some externalists are happy with the prospect of agents having some reasons that are diametrically opposed to the things an agent wants. However, a common feature of externalist theories is that there is some feature of the reasons themselves that enables them to be had by agents. They are the guarantors of their own normative status, so to speak. Some even just state that reasons for action are simply brute facts about the world such as some logical or mathematical truths are just true in some fundamental *sui generis* way, which stands in no need of further explanation. A common argument from analogy deployed is referred to as the 'partners in crime argument' with epistemological stances taken toward the role of truth in belief formation. In epistemology it is often taken as basic that the truth of a statement *is* in-and-of-itself a reason for forming ones beliefs in alignment with it – quite independently of what any agent might want, desire or feel motivated to believe. It has been a longstanding hope that externalists could put reasons for action of a similar supposedly firm and intuitive grounding – i.e. It is just in the nature of some actions that agents have reason to carry them out, or that some states of affairs just give reasons for agents to promote them. This discussion and its hopes of success has raged in the literature for years and is still ongoing.

However, an obvious problem with this notion of reasons is what J. L. Mackie famously referred to as his 'argument from queerness'[44]. Here, Mackie argues that there is something unavoidably metaphysically *queer* or fishy about the notion of a fact or property that posits something objectively true about the world while simultaneously being capable of motivating an agent, simply by virtue of their comprehension of the fact. For example, Mackie believed the idea that any agent who acknowledged the truth of the statement 'charity is morally good', in acknowledging it could not help but be motivated to act charitably, was absurd. According to error theorists, no extant property, objective value or fact about the world could fulfill both these roles at the same time – could be both descriptive and prescriptive at the same time.

[44] J.L. Mackie, *Ethics: Inventing Right & Wrong*, Penguin Books (1990), p.38-42.

The argument is made that externalist realists are committed to such entities existing if their theories are to work, and are hence fundamentally ontologically flawed. However, many externalists are just comfortable with the idea that reasons can fulfill this metaphysical role, while others deny the premise used by Mackie and other error theorists that moral realism or Externalism are actually committed to the existence of such queer entities. A major drive toward Internalism however, is a sense of unease with reasons construed in such a way – that somehow the externalists' account is incomplete or dodging what Armstrong might call 'a mandatory question of the exam'.

Internalists favour shifting the burden of the normative impetus of reasons away from the reasons alone, and instead ground it in the relationship a reason has to the desires or other motivational states of agents. Hence, and in keeping with the spirit of ontological parsimony, of Ockham's Razor, if an internalist can account for reasons in natural, metaphysically un-(or less)-contentious features about human psychology, it would be, *ceteris parabus*, the superior theory.

By grounding the existence of reasons in natural facts about human psychology Internalism offers a simpler account of how agents come to have the reasons they do. In this way Internal reasons would align with our basic intuitions concerning agents and their reasons that remains in keeping with measurable facts concerning their make-up. It would also sidestep the metaphysical issues inevitably encountered by externalists. Internalism then, offers the prospect of comfortably satisfying one of the two commonly held intuitions I mentioned in the opening paragraph of this introduction – that reasons for action are intimately linked with the desires and long-term projects of agents. However, as you will see, the rest of this chapter will be dedicated to the *prima facie* problem it faces conforming with the other – the existence of moral reasons.

2.3 The Motivational Requirement and Some Distinctions

Put in its simplest possible form, Reasons Internalism holds that there is a necessary connection between what reasons an agent has to act and how an agent is or could be motivated to act. For the reasons internalist there is an inconsistency in the notion of an agent having a reason to carry out some action that they are not or could not find themselves being motivated to carry out. For the internalist, the truth of reason ascriptions to agents implies that they possess or are constitutive in some way of motivational efficacy – the capacity to cause an agent to act. You'll recall, from what of

Williams we've already looked at, for something to be a reason that an agent has it must be possible for an agent to act *for* that reason. A 'reason' that could not be part of the causal *explanans* of the agent's action is no reason at all, at least, according to the internalist. The stipulation that whatever reasons for action an agent has having the capacity to motivate them is sometimes referred to as the 'motivational requirement' of reasons for action.

> *The Motivational Requirement*: For there to be a reason that an agent φ, it must be possible that the agent could be motivated to φ.

However, as with almost everything in Philosophy it isn't quite that straight forward. 'Internalism' means different things in the mouths of different writers and theorists and there are nuances to bear in mind if we're going to identify the precise version or tenet of Internalism we wish to show is incompatible with moral reasons as I have described them. The first of which is the distinction between Reasons Internalism and Motivational Judgement Internalism.

Motivational Judgement Internalism is a form of Internalism that holds that there is a necessary connection between the judgements agents make about the actions they should carry out and their being motivated to carry them out. For example, if an agent arrives at the sincere judgement that they should exercise more or give more to charity, the motivational judgement internalist holds that this judgement *necessarily* implies that the agent feels motivated to exercise more or give more to charity, respectively. According to this model, a psychopath who feels *no* motivation at all to be compassionate yet claims that they can see or have judged that they shouldn't do the cruel things they do, hasn't genuinely made this judgement. The strongest *prima facie* evidence for this position being the regular and reliable correlation between the judgements agents make about what they should do and the motivations they experience.

Reasons Internalism – or Existence Internalism, as Stephen Darwell called it – is not about the judgements agents make concerning their prospective courses of action. Instead it's about the reasons they do or do not have or the reasons that actually 'exist' for them. An agent has some reasons for action and not others. They might even be unaware at times of the reasons they have, but the reasons internalist believes that

what reasons they do have, they have in virtue of it having some connection with those things that motivate them.

To illustrate, I will return to Mark Schroeder's example of the surprise party[45]. To paraphrase, Suzy loves successfully organized surprise parties. Her friend Lucy has planned just such a surprise party at her house that evening. The fact that Suzy loves surprise parties and that there is one at Lucy's house this evening is unquestionably a reason that Suzy has to go to Lucy's house. However, by definition it is not a reason that Suzy could be motivated by; for to be motivated by it she would need to be aware of the party which would negate the surprise and hence the reason to go. However, the fact that Suzy does have this reason is dependent on something she desires – i.e. to have a surprise party successfully sprung on her – not that she could be motivated by the fact that there is a surprise party. In other words, the existence of the motivational state is what makes someone have a reason to do something.

Throughout this thesis I will use the term 'Internalism', when unqualified, simply to refer to Reasons or Existence Internalism as I have outlined it above, to distinguish it from Motivational Judgement Internalism. The latter will not be discussed again in this thesis. Conversely, I shall be using the term 'Externalism', when unqualified, simply to refer to any theory of reasons that makes no stipulation that the motivational requirement be met – i.e. I take an externalist as simply being any reasons theorists who believes that the ascription of reasons to an agent does not necessarily depend on that reason being capable of motivating them in the manner Williams envisaged.

A further clarification that needs to be made is the important question concerning the strength of the claim being made by the internalist. On the one hand the internalist could be saying that,

1) If A has a reason to φ then necessarily A has some motivational state that makes it a reason for A to φ.

(1) is a very strong claim. It rules as impossible that anyone could have a reason to do something, yet lack the motivation to do it. Yet counterexamples abound. There are examples of weak-willed individuals who for reasons of fatigue, depression, addiction or some other psychological block may lack the motivations which track with their reasons. Or, as is the case with Williams' petrol/gin & tonic example, they may

[45] Mark Schroeder, *Slaves of the Passions*, Oxford University Press (2013), p.33.

lack the motivation to act as they have actual reason to, or vice versa. Are we really going to say that reasons are always lacking in the absence of actual motivations that concord with them? I don't think we should – and most internalists in fact don't.

A weaker, more defensible and more widespread claim is something like,

2) If *A* has a reason to *φ* then necessarily *A* has some motivational state that makes it a reason for *A* to *φ*, or *A* would come to have such a motivational state were they in possession of all the relevant information and deliberating on it soundly.

This is much better. It allows for the possibility of mismatch between the reasons an agent has at any given time and the lacking the actual motivational state that would make this a reason for them, whilst still meeting the motivational requirement. Since the versions of Internalism that stick to (2) – i.e. the weaker of the two claims – makes for stronger theories, I shall aim to make these my target. To do otherwise would be to severely reduce the scope and applicability of my central point – and, to a degree, strawman Internalism.

Another reason this is important is to highlight the distinction that should be made between 'Internalism' and 'Constructivism'. The majority of internalists hold that motivational states (most commonly, desires) are a necessary ingredient in accounting for why an agent has a reason to do something. Constructivists however, make the much stronger claim that the motivational state is both necessary *but also* sufficient for the agent in question to have a reason to act.

Take the example of some agent; call them Robin. Does Robin have a reason to eat the delicious slice of cake in front of them? Let's assume that Robin has a strong desire to eat the cake. The constructivist would say that Robin's desire for the cake means that Robin has a reason to eat the cake – because for constructivists, motivational states are sufficient for the existence of reasons. It is also possible, however, for the constructivist to concede that Robin might have a countervailing reason *not* to eat the cake – if they also have a desire to lose weight, for example. And so Robin's choice may come down to which reason is the weightier.

The internalist, on the other hand, can't say whether or not Robin has a reason to eat the cake based simply on the presence of Robin's desire to do so, because motivational states alone are not sufficient for them, only necessary. So, they would have to confine themselves to saying merely that *if* Robin does have a reason to eat the

cake it *must* be because of some motivational state that they have. Also, just parenthetically, the externalist holds that it is possible that Robin has a perfectly good reason to eat the cake regardless of any motivational state they might have. The crucial point here then, is that both internalists and constructivists hold motivational states necessary for the existence of reasons.

Both internalists and constructivists, even between themselves, disagree as to how exactly motivational states ground reasons, or make reason ascriptions to agents true. There are some very hard-line constructivists who would maintain, that because desires are sufficient for reasons, they can say that motivational states *are* reasons, *simpliciter*. This is an exceptionally strong claim however and not widely held. But both internalists and constructivists can hold that motivational states play some part in constituting the reasons an agent has, or reject that they constitute reasons at all, and hold instead that they merely play a role in accounting for why agents have the reasons they do.

For example, my being a mammal is constitutive of my reason to consume Oxygen. In other words, it is my *being* a mammal that makes it a reason for me to breath Oxygen. On the other hand, the reason I have the DNA I have is in part because my parents have the DNA they have. However, my parents do not constitute me.

Different internalists and constructivist theorists can hold totally different positions as to if and how and to what extent motivational states are part of reasons for agents, or are just what makes it true that agents have those reasons. They may even agree on the relations between motivational states and reasons. What is important is whether or not they consider them as necessary and sufficient, or simply necessary.

The constructivist makes the stronger, and hence narrower, claim that motivational states are sufficient to account for the existence of reasons. All that is needed for an agent to have a reason is for them have a suitable motivational state that makes it a reason for them. The internalist makes the weaker, and so broader claim that there doesn't ultimately have to be a reason for every motivational state – only that for every reason there will be a motivational state that accounts for it. This makes internalist accounts of reasons inherently more nuanced and versatile.

While we will be looking at a more constructivist theory in the form of Mark Schroeder's *Hypotheticalism* (See Chapter Four), it is Internalism that will be my main target in this thesis. This is because I believe specifically that making the existence of

moral reasons in any way contingent on the elements of an agent's psychological make-up, irretrievably undermines them. It is not my task to refute that motivational states in-and-of-themselves provide reasons for action. It is Internalism then that is my true target. Where I do attack constructivist theories it will only be in as much as the claims they make are shared by or compatible with internalist ones.

2.4 The Apparent Problem

The reader should have already spotted a *prima facie* mismatch, alluded to in the introduction (I.1), between the requirements applied to reasons by Internalism and each of the three criteria for being a moral reason as I have outlined them in Chapter One. We will go through each one in turn, starting with the first – categoricity. Let me state the problem directly.

1) What agents have moral reason to do is true or false whatever they are or could be motivated to do.
2) What agents have reason to do is invariably dependent on what they are or could be motivated to do.

At first glance it would seem that these two statements, as they stand, are in conflict with one another and that at most one of them can be true at any given time – of course, only if we are working under the assumption that there are really moral reasons.

(1) is merely what I mean by the categoricity of moral reasons. If you have a moral reason to φ then this is simply a fact about an agent and/or the world. As such it would be a fact that an agent could sincerely accept as being true of themselves, while all the time feeling wholly unmotivated to act in accordance with it and without incurring the charge or practical irrationality.

(2) is an essential and indeed core tenet of Internalism – it is effectively just the motivational requirement restated. It makes explicit the internal, necessary connection between reasons for action (of which moral reasons are simply a sub-set) and motivation. *All* reasons for action must be capable of motivating agents in order for them to be reasons at all.

It would seem to follow then, if (1) were true it would be possible for there to be a reason for action (a moral one) that might not be capable of motivating. And if (2) were true, it would necessitate that any moral reasons there were would have to be

ones that could motivate. The problem for Internalism then is to resolve this apparent conflict.

There are many different acceptable ways, at least to me, that this could be done. The internalist could, for example, show that there are certain morality-inclining motivational states so perineal to agents that there would never be a time when they couldn't be motivated to act morally. Another might be to say that moral action has some peripheral utility which vicariously serves an end that agents are invariably motivated by. It could be that the provision of moral reasons is guaranteed by universal features of practical reasoning itself and that as such agents always have moral reasons simply by virtue of being rational beings at all. We shall be looking at several of these options. I will repeat a point I made in Chapter One though; for an internalist theory to meet the categoricity requirement it is not necessary for an agent to have that reason independently of their desires, only that they have the reason *whatever* their desires are.

What of the criterion of weight? As I wrote in Chapter One, it must be possible to generate moral reasons in such a way that at least some of them won't be able to be outweighed by anything except the very strongest of non-moral reasons. Now the issue for the internalist then becomes providing an account for how moral reasons gain this weight and how and to what extent they are dependent on or limited by the motivational states in question. To use a specific example, in Chapter Four we'll be looking at Mark Schroeder's innovative rejection of a trend in Humean internalist theories that he refers to as *Proportionalism*. This is the idea that the weight of an agent's reason to act in some way is proportional to the strength of the desire they have, which is furthered by them acting in that way. If Proportionalism were true it would follow that where desires are weak for those ends furthered by acting morally, the corresponding moral reasons would also be weak.

More generally then the problem for the internalist in meeting the weight criterion is to show that the ends furthered by acting morally are invariably desired strongly (which will be David Gauthier's strategy – see Chapter Three), or that while moral reasons are grounded by motivational states, their weight is determined by some other factor (which shall be Schroeder's tactic).

That just leaves the question of the right grounding. Recall, one of the criteria I argue is essential to a reason for action being a moral reason is that it must be

grounded, at least in part, by something that is essential to the act in question being the type of act that it is. There is arguably something, or some combination of things, call it '*x*', that is essential for an act of theft to be accurately classified as an act of theft. It is this *x* that must play a crucial role in explaining why an agent has a moral reason not to steal.

Immediately, we can see the potential for a problem to arise for the internalist. I say there must be a necessary connection between an agent's moral reasons and this *x*-factor, and the internalist says there must be a necessary connection between an agent's reasons for action and some motivational state. Therefore, in order for some form of Internalism to provide moral reasons, to my standard, it must provide reasons that meet the motivational requirement, but also and without fail, reasons that are grounded by this *x*-factor. Since there is no *a priori* reason to believe that *x* and the appropriate motivational states are either identical or necessarily connected, there is potential for an unacceptably substantial mismatch from the outset of any internalist's project.

For example, an internalist moral theory that could somehow show that deep down we all have a reason to act morally because we all ultimately take the deepest satisfaction from doing what is morally right, would readily meet the motivational requirement – assuming that we can be motivated to seek what gives us deepest satisfaction. But this would not be the right grounding for moral reasons as it is grounded by the fact that doing what is morally right happens to invariably give satisfaction, not by any feature of the acts themselves. If lying, cheating or stealing happened to give greater satisfaction in the end, this would mean one would have the same reason, in terms of attainment of satisfaction, to do the opposite. Yet there is still, or should be, a moral reason not to lie, cheat and steal.

However, as I have already stated in 1.4, I want to reiterate that the right grounding stipulation does not beg-the-question against Internalism. It doesn't rule out *a priori* that there *is* a necessary overlap. If there were some reason that agent's invariable desire to do the morally right thing, *qua* the right thing, this would be an instance of an internalist theory meeting both the motivational requirement and the illusive *x*-factor. All that is required is that when the internalist is satisfied that they have shown that there is always a motivational state to ground a reason to behave morally, either the motivational state in question or the reason grounded by it be of a

kind that goes at least some way to satisfying our standard notion of a reason or motivation we would typically consider a moral one – as usually indicated by our willingness to confer upon it either praise or blame. I can see no foregoing reason to see why this particular kind of reason should not be possible to provide within an internalist moral theory. Hence, while I may be bringing the goalposts a little closer together from the outset, but I am by no means begging the question against Internalism.

2.5 What the Internalist Must Deliver

We have seen now, why the internalist has their work cut out for them. As it happens, I believe, for reasons that I hope will become evident throughout the rest of this thesis, the task of meeting my three criteria is unattainable by any moral theory that simultaneously necessitates that reasons meet the motivational requirement. However, I do not wish stack the deck against Internalism from the outset. As we examine the three specific internalist theories we will be looking at in Chapters Three, Four & Five, I will do my best to see in what ways each of them might be adapted or interpreted so-as they might be able to meet them. Yet, in the light of the issues discussed in this section, I think it is vital to keep in mind precisely what any internalist theory must achieve in order to provide moral reasons as I define them.

1) The theory must provide reasons to act morally that an agent could not fail to have by dint of their lacking a requisite motivational state that does or would perform the role of motivating them – i.e. for any given moral reason, it could not be the case that an agent might lack a motivational state that would motivate them to act in accordance with it.
2) At least some of the reasons generated to act morally must either invariably motivate an agent strongly enough that they can't be outweighed by anything other than the very strongest non-moral reasons; or alternately the weight of the reasons that meet the motivational requirement are not wholly determined by the strength of the motivational state in question.
3) There must be wide, reliable and consistent (and preferably necessary) overlap between reasons for action or inaction that are grounded on factors essential to what constitute those actions or inactions, and reasons for action or inaction that motivate or have the potential to motivate an agent to act.

Clearly the job before the internalist is not an easy one. Yet I believe this is an inevitable consequence of stipulating that all reasons for action must meet the motivational requirement. However, I hope the reader will agree that, while challenging, the labour is not inherently impossible to enact – it is not rigged from the start. There would be no logical inconsistency in a theory that achieved all of these things. In fact, if the internalist were to succeed, none would be better pleased than I!

In concluding this chapter, I think it important to re-affirm that the purpose of my thesis is not to show that the motivational requirement is somehow incorrect – that somehow, the conditions necessary for reasons for action to be moral reasons, as I understand them, proves that it is false. My thesis is entirely consistent with the out-and-out truth of the motivational requirement. My thesis is only that *if* the motivational requirement is correct there can be no truly moral reasons.

Chapter Three

Gauthier

3.1 Introduction

The first internalist moral theory I shall be assessing will be the Contractarianism of David Gauthier, as presented in his *Morals by Agreement* (1986). I have chosen to start with this because I believe in some respects Contractarianism is an attempt to reconcile morality with one of the simplest and most uncontentious normative models to exist - specifically, that an agent has reasons to act in their own self-interest. As we might put it, Contractarianism has a minimalist normative ontology. Therefore, in the spirit of Ockham's Razor, if self-interest alone can prove sufficient grounding for moral reasons, we would not need to argue for or establish a more contentious or dubious array of desires or motivational states. All things being equal, this is always an advantage for any theory.

Contractarianism is an attempt to show that moral rules, *qua* constraints an agent would accept on their own actions, are nothing over and above the set of principles an agent would accept as reasons for them to act or not act in certain ways, when they are reasoning soundly and purely motivated by their own self-interest. The reason I consider this an excellent starting point is because of all reasons for action, furtherance of one's own self-interest is often seen as the most primal and the least controversial. What I mean by this is, aside from the most ardent normative skeptic, it is very widely accepted that if agents have reason to do anything, then they surely have a reason to pursue their own desires, welfare and long-term goals. Thus, if a moral theory can give rise to moral reasons that are not only consistent with but are in fact grounded by agent's own interests, and require no additional normative grounding, it will be one that stands on a strong foundation.

3.2 From Basic Assumptions to The Archimedean Point

Gauthier begins, as his intellectual forebear Hobbes does, with a series of basic assumptions concerning agents and then, given what these assumptions are and what

they must entail, proceeds to extrapolate from them the kinds of reasons for action agents must ultimately have[46].

His first basic assumption is that an agent has the strongest reason to act in pursuit or furtherance of those things they value. For Gauthier's purposes, he takes it that for a thing to be valuable is for it to be valued.

Gauthier does not reject the possibility of there being other potential (external) sources of value – i.e. ones not constituted out of the subjective choices or yearnings of agents. He only observes that whereas the existence of subjective values can hardly be doubted, the existence of alternative sources of value that provide reasons for agent's actions can. Since, in Gauthier's estimation, they are not necessary to provide morality with its foundations he considers the burden of proof to be on those who maintain the primacy of such alternative sources of value. So for Gauthier, at least for the purposes of his theory, value is constituted out of individual subjective interest or desire. In turn, pursuing the objects of these desires or interests provides an agent with reasons for action.

For a second basic assumption, Gauthier takes it as axiomatic that any agent acting rationally will intend to take the course of action they believe will achieve the greatest *maximization* of whatever they value. In other words, agents when behaving rationally will perform utility calculi. They constantly weigh up the different gains they can reliably predict will come from the different courses of action open to them, and select the one that they believe will garner them the greatest relative gain, in any given situation. All other things being equal, if one course of action will garner a profit or gain the agent values twice as much as another, we are justified in having the rational expectation that they will opt for this course of action.

To summarize these two assumptions, and express them more crudely; people have reasons to pursue what they want and they have the greatest reason to choose the course of action that will bring them the greatest amount of what they want. Gauthier firmly believes that armed only with these two assumptions, extrapolating from them with sound deliberation and reasoning, it is possible to demonstrate that all agents have reason to behave morally. But before we can see this properly we need to clarify exactly what Gauthier takes morality to be.

---

[46] David Gauthier, *Morals by Agreement*, Clarenden Press (1986), the whole of Chapter 2.

What, Gauthier asks, can morality be most concisely characterized as other than 'a constraint on each person's pursuit of his own interest'[47]? It is surely in the very essence of morality to prohibit or illicit actions that agents would otherwise be inclined to do or not do, respectively, in the natural course of their rationally seeking the maximization of their own interests. Otherwise, as Hume rightly observed, there would be no need of morality in the first place. Yet without employing any pre-existing moral concepts and ideas, how could one provide a rational agent with reasons *not* to pursue actions that would appear to maximize their interests – even if these actions include those typically regarded as immoral, such as deception or stealing? This is the problem Gauthier sets himself; the problem of justifying *rational constraint*.

In order to demonstrate the rational basis of personal constraint, Gauthier distinguishes between what he terms *straightforward maximization*, on the one hand, verses *constrained maximization* on the other[48]. The straightforward maximizer is the kind of agent who, in every single situation, weighing up all of their alternatives, calculates the single best course of action for the maximization of their aims and acts in accordance with this conclusion. However, Gauthier points out that calculi of this nature will inevitably lead, for the most part, to repeated failures to attain what might have been achieved by *all* parties involved, had the agent(s) in question taken each other's perspectives into consideration. This is because straightforward maximization does not include the taking into account of the similar calculations that other, equally rational and egoistically-minded agents – with whom some degree of interaction is inevitable, and indeed, on occasion, desirable – into due consideration.

Modern economic theory is replete with examples of self-defeating strategies that fail due to this kind of lack of sophistication and nuance. For example, two high street competitors are vying for customers. One of them decides to slash their prices considerably in the hope of drawing custom away from the other. This would seem sound at first glance. However, it is a rational expectation that their competitor knows as much as them and can be reliably predicted to attempt to lower their own prices to match, in order to counteract the actions of the first competitor. The action of the one competitor *in response to* the other's alters (or in this case neutralizes) its outcome. The situation is now that they *both* remain equally appealing to their customer base, but

[47] David Gauthier, *Morals By Agreement*, Oxford University Press (1986), p.8-9.
[48] Ibid, p.167.

both are making less money than they were before. The paradox being that had neither sought the maximization of market share in such a straightforward manner, they would both be better off.

This is obviously a simplified example. However the point is that straightforward maximization will fail for the most part if it does not factor into its calculi the actions that other similarly-placed agents are likely to take to respond to and counteract the former's attempts to do so. In situations involving more than one agent, the outcome of an individual's action will depend on the complex dynamics or expected responses, and responses to responses, of all other rational agents participating.

Borrowing from an example used by Hume in *A Treatise of Human Nature*, to highlight the predicament further, Gauthier asks us to imagine two farmers running adjacent farms[49]. They are not hostile to each other, yet are not friends either. One's crop is due for harvesting now, the other's a month hence. Each can harvest their crop with greater efficiency and benefit to themselves if they have the help of the other. However, the one with the later harvest will only help the other if the latter will help them with theirs in a month's time. Each helping the other would therefore lead to them both being better off. However, the problem arises when you consider that once the farmer with the later harvest has helped the one with the harvest due now, there will be no reason for the one whose harvest has already been gathered to then help the other in a month, as this would only constitute a loss for them in time and energy.

In Gauthier's scenario, we are imagining that, for whatever reason, there would be no negative consequences to the one with the harvest already gathered, refusing their help to the other later. So, in that case, if the farmer with the earlier harvest is straightforwardly maximizing they will seek the help of their neighbor now and then deny their help later. However, the point is because the farmer with the later harvest knows it is in the interest of the other farmer to do precisely that as well, they will not agree to help with the earlier harvest. In this case if both farmers pursue maximization straightforwardly, both will end up losing out. As Gauthier puts it, '[I]ndividual benefit and mutual advantage frequently prove at odds.'50

These considerations, of course, echo the sentiment expressed by Hobbes in *Leviathan*, that Man outside of civilized codes of conduct, without well-established and

[49] David Gauthier, *Assure & Threaten*, presented in *Ethics*, Vol. 104, No. 4 (Jul., 1994), p.692

[50] David Gauthier, *Morals By Agreement*, Oxford University Press (1986), p.13.

ingrained institutions like promise-keeping and fidelity, is left very much isolated, unable to form mutually beneficial contracts with others. Ironically this would lead to a far lesser attainment of what he desires. Famously, Hobbes made the stronger point that taken to the extreme, totally unconstrained maximization would lead to a state of war of all against all, and a life that would be 'solitary, poor, nasty, brutish and short'[51]. Gauthier does not make the case so strongly as it is not crucial to the case for morality he is making. His point, on the other hand, is simply that the stance of straightforward maximization, in a significant number of situations, makes it *impossible* to form agreements that would lead to more advantageous interactions between *all* of the agents involved in the exchange. In such situations then, Gauthier maintains, it can be positively disadvantageous for *all* agents concerned to attempt to maximize straightforwardly.

Instead, he offers an alternative; *constrained* maximization. If an agent were to factor into their deliberations not only the rational expectation that all of the agents with which they are interacting will likewise be attempting to maximize their gain, *but also* that each agent will be perspicuously aware that every other agent (including themselves) will be doing likewise, we can begin to see the outline of a fully rational basis for an agent adopting constraints on their own behavior.

Let's go back to the two farmers for a moment. It is simply a matter of fact that had the farmer with the earlier harvest *sincerely* believed themselves to be committed to helping their neighbor, purely by virtue of the fact that the farmer with the earlier harvest had helped them – even though, after the fact, they had nothing to gain from it – and the other farmer been aware of this, they would have been able to co-operate to mutual advantage.

According the Gauthier, 'Morality arises from market failure'[52]. Straightforward maximization is often self-defeating in the bid for the greatest benefit. Whereas, in certain circumstances, by forgoing straightforward maximization and seeing the larger dynamic, which includes the deliberations of their fellow agents, opportunities for greater maximization become possible for all agents concerned.

When the straightforward maximizer stops to consider the rationale of their own actions in the broader context of the rational expectations they can have of every other

[51] Thomas Hobbes, Leviathan, Penguin Books (1985), p.186.
[52] David Gauthier, *Morals By Agreement*, Oxford University Press (1986), p.84.

agents' choices when doing likewise, their horizons for greater accumulation of personal benefit can't fail to broaden with it. However, it is important to note that no expansion in empathy or altruistic concern is necessary at this stage. The enrichment of one's plethora of reasons to act can still be seen as being entirely motivated by self-interest.

Take the scenario with the two farmers again. What if the farmer with the earlier harvest had genuinely been *the sort of person* who could enter into agreements with their neighbors and then fulfill them *for the sake of* honoring those agreements, rather than the actual utility doing so would garner at the time of fulfilling them? Well, in that case he would have been the kind of agent that the farmer with the later harvest would have been willing to enter into a contract with. Thus, the agent who honors agreements come what may, over always opting for maximizing their gains, will oftentimes do better at maximizing their gains!

In this way, when the strategy of straightforward maximization can be seen by enlightened agents to stunt their potential for more beneficial arrangements, the first kernel of moral sensibility is sown. So, by Gauthier's lights, 'morality' emerges when agents, for no other reason than their own self-interest, genuinely make themselves into desirable collaborators for joint enterprise, to their fellow agents, and vice versa – with the genuineness of this conversion as the most crucial element of Gauthier's case. All agents come to see the utility of partaking in both individual contracts, which might be one-off encounters, and of being a member of the larger, ongoing social contract of agents living in commune with each other for the purpose of mutual benefit[53].

We have been talking so far in terms of simplistic contracts. The kinds of 'morals' being envisaged at this early stage of the story are somewhat limited to contract fulfillment and oath-keeping. But this does not imply only that these would be the sum total of the types of moral rules in the final analysis. The point Gauthier is making is that the cultivation of an honest character is the most surefire means to establishing a reputation for trustworthiness, which increases ones opportunities to be involved in all manner of different forms of interpersonal involvement[54]. I believe that this is the common thread that Gauthier imagines runs through all the rules of morality. Loyalty, honesty, integrity, generosity, compassion, gentleness and all of the constraints on

[53] Ibid, p.336-337.
[54] Ibid, p.173.

behavior these classic virtues imply, all make agents more desirable to be associated with. When reliably embodied by an agent they make that agent the sort of person other agents either want, or feel they safely can, have all variety of different mutually beneficial relationships. 'Good character' then, conceived of as this kind of reliability, is in the long-term interests of the one who cultivates it in themselves, and would create suitable motivation for them to do so.

'Morality' then, on this account, is the set of rules that rational, self-interested agents do or would voluntarily assent to – when they're in possession of all of the salient facts regarding what actions and strategies actually do maximize their outcomes – for the sake of successful interactions with one another[55].

I should probably say a little bit here about the role of what we might call the 'sincerity' or 'fidelity' of the agents who adopt the kinds of moral principles Gauthier is trying to lay the groundwork for. To a degree, Gauthier sees adoption of moral principles as being almost a suspension of the agents' awareness of the ultimate foundations and utility of the rules, and follows them for that sake of the rule itself. For example, by Gauthier's lights, it is no good if an agent only fulfills contracts for the sake of what fulfilling any given number of contracts will bring them. For potential partners will always be wary of entering into contracts where there is potential for the agent to achieve a greater boon for themselves by reneging. If on the other hand the attitude internalized by the agent is that one fulfils contracts purely because *that's what is to be done with contracts*, the agent has what we might call 'integrity'. It is the rules themselves that motivate compliance by the agent rather than the gain the agent hopes to garner by following them that motivates the agent. Thus 'sincere' adoption of moral rules allows them to achieve the authority they require so they can consistently inspire the compliance they need to in order to fulfill their ultimate function, rather than mere awareness of their utility. We might say that morality for Gauthier is nothing over and above the rule of law. Both acceptance and compliance with such a set of constraints on an agent's action thus serve each adherent of them. It provides even the most utterly self-centered agent with a reason to observe them – with a reason to be moral.

It is important to note that, like Hobbes, Gauthier does not make any stipulation that this agreement be actual – i.e. it is not necessary to the validity of his case that any actual group of individuals in fact do sit down and hash-out precisely what laws and

[55] Ibid, p.125.

mores they are going to live by. The fact that a clear case can be made for the rationality of each or all of the morals in question can be made is sufficient. It is acceptable for the agreement or contract to be purely hypothetical[56].

This theory is developed further still by Gauthier's stipulation that the rules agreed upon are entirely 'non-tuistic'. That is, they take no account of any actual *individual* within the society/contract. By the rules being set-up in such a way that no advantage is garnered by individual participant within the system, the set of rules that bind all participants would be ones that all would rationally accept regardless of the position they hold within the system. For example, if a law were to allow certain individuals to be exempt from the complete fulfillment of contracts by reason of gender, ethnic group or even membership of a certain family – it would ultimately harm those exempted, as others would be less willing to engage in contracts with them. By striping away all idiosyncrasies from the individuals within the system and instead tailoring the rules so that they serve all potential agents equally, we approximate or approach what Gauthier refers to as the *Archimedean Point*[57].

The Archimedean Point is one where any rationally-minded and informed participant in a social contract or system of rules has sufficient vantage point to see how the rules serve all other agents within the system as well as themselves[58]. And from the Archimedean Point, any agent would conclude that, *ceteris parabus*, they would be nobody better served by the rules if they occupied a different position within the system. For example, ideally all citizens in an egalitarian state should have the confidence that if they are charged with a crime, the law and judicial system of that nation would serve them no better if they had a different, ethnicity, gender, profession, income or family.

The Archimedean Point is reached by striving for equality in rule selection. This is a set of rules that provide that in any given permutation – i.e. in any societal or contractual arrangement, no one could be better served by the rules without someone else being served worse. To put it another way, the Archimedean Point has been reached when any agent within the system, looking down at the positions occupied by

[56] Ibid, p.10.
[57] The reference here is to Archimedes famous claim that with a fulcrum and a lever long enough, he could move the world. The allusion being that from a sufficiently removed vantage point it is possible to have an authoritative hold on the whole system.
[58] Ibid, p.233.

every other agent, could not regard the maximization of their interests being improvable by switching places with anyone else in the system, all else being equal. By consistently renouncing the idiosyncratic wants of the individuals who must assent to them, they become equally serviceable, endorsable and acceptable to all potential participants in the system. This, for Gauthier, is the grounding of the moral authority of the system of rules – a system of rules that are equally rationalizable, and thus equally compelling, to any possible agent within the system. Furthermore, the acceptance and observance of these rules would be in the long-term interest of all parties[59].

Gauthier does not leave it here, however. For to do so would have laid the groundwork sufficiently to justify only what he refers to as 'Economic Man' – i.e. an individual who ultimately is motivated to observe morality purely for its utility. Though sincerely adopting moral principles and complying with them in accord with their commitments, it is done so without any true sense of honor, compassion or altruism. But Gauthier's philosophy leaves plenty of room for what we might call morality's nobler character. 'Morality', on Gauthier's account, is an emergent phenomenon with a rationale of its own, quite distinct from, and yet at the same time founded upon, the striving toward maximization that gives rise to it.

From here, there is a conceptual transition from the utility motivating economic man to what Gauthier terms the 'liberal individual'. The sentiments of kinship, decency, compassion and kindness become the subjective fuel-to-the-fire, so-to-speak, that provides individual agents with their motivation to morally correct action. For these feelings lead us to value the kind of co-operative, communal, social activities, which in turn provide individual benefit. In this way, the moral sentiments end-up engendering other-regarding, tuistic constraints on action, making what for economic man might at times be a burden – i.e. the occasions of having to give up a desired end in order to fulfill ones moral obligations – a source of selfless pleasure and joy in and of themselves for the fully transitioned liberal individual. Furthermore, the noblest facets of morality, such as forgiveness and self-sacrifice, do not need to lose any of their peculiar splendor. It is just the reverence that morality inspires that enables it to have its ultimate utility[60].

It is in no way essential to Gauthier's case that the truly moral agent thinks in terms of their principles being grounded or motivated by self-interest. In fact it is

[59] Ibid, p.238.
[60] Ibid, p.345.

beneficial if they don't think in such terms. It is only necessary to show that such principles, however they are conceived of when acted upon, *can* be provided with a rational foundation. Thus, the liberal individual is one who has fully internalized morality – observing it for its own ends, yet in the long run, individually reaping the benefits of doing so.

3.3 The Problem of Rational Compliance

Since its original publication the arguments put forward in *Morals by Agreement* have attracted a great deal of debate and inspired much writing. Questions and criticisms have ranged from whether or not it makes sense to base a moral theory on a very narrowly-framed conception of 'reason', which takes pursuit of self-interested goals as fundamentally constitutive of its character[61]; to the soundness of the cost-benefit analysis of constrained vs. unconstrained maximization. In the case of the latter, Gauthier asserts that this provides a purely self-interested reason for constrained (a.k.a. moral) behavior.

However, in this section I am going to focus exclusively on what I think is the single biggest problem for Contractarianism from a theoretical basis, the problem of *rational compliance*.

Let us return to the example of the two adjacent farmers, discussed in the previous section. The cold, hard reality in this scenario is that when the time comes around for the farmer who had the earlier harvest, to fulfill their end of the agreement there is no self-interested reason to comply with it, as it will only result in a cost to their time and energy. Recall, we can make no appeal to integrity or honor, as this would beg the question. They are the very kind of moral concepts Gauthier is trying to ground.

Gauthier's argument would seem to hold true in a very large number of interactions – and for the creation of a broader social contract generally. However, if we assume there would be literally *no* negative consequences to the farmer with the earlier harvest reneging on their agreement – e.g. as Gauthier suggests, they are planning to move away to Florida as soon as their crop is in, where they will never encounter anyone who know they are an agreement-breaker – what then? If this situation arises, if the farmer were to perspicuously and insightfully examine their situation, would it not

[61] See Holly Smith's, *Deriving Morality from Rationality* (1991).

be eminently rational of them to re-examine any moral precepts they have internalized and break their agreement.

The problem of rational compliance is this; whilst general compliance with the rules governing agreements or society does undoubtedly maximize an agent's interests for the most part, inevitably there will be at least *some* situations where it simply isn't the case that it is, *in any sense*, in their interests to comply. Yet, typically we would think that the demand of the rule – the demands of morality – no less applies in those situations. In the case of the farmers, whilst it is fairly uncontentious to say that agreement is in both parties' interest, for Contractarianism to work, an indisputable reason must be shown to exist for why, despite the cost, the farmer with the early harvest has reason to comply at the later date. I do not believe it any exaggeration to say that Gauthier's entire model hinges on being able to provide such a reason.

In essence the problem of rational compliance is the age-old problem of the Hobbes 'Foole'[62], Hume's 'Sensible Knave'[63], and the free-rider[64]; simply in a different guise. This boils down to the question of how, if one does not assume the foregoing authority of moral concepts (hence begging the question), it can be said that an agent behaves incorrectly or irrationally if they do something morally wrong that is unequivocally to their advantage.

Realistically and practically, surely no system of laws or mores can be expected to be so comprehensively and consistently enforced that occasions for advancement by transgression will by impossible. David Copp makes this point very deftly in his *Contractarianism and Moral Skepticism*, where his comments regarding the threat of moral skepticism that could be applied to this issue of the Foole[65]. What say the contractarian then, to those who find themselves in those positions where they have opportunities to get out of their commitments unscathed?

> '[T]o refute the Foole. I must defend not only the rationality of agreement, but also that of compliance.'[66]

---

[62] Thomas Hobbes, *Leviathan*, Penguin Books (1985), Chapter 15, p.203
[63] David Hume, *Concerning the Principles of Morals*, Open University Press (1999), Section 9, p.282.
[64] David Gauthier, *Morals By Agreement*, Oxford University Press (1986), p.96.
[65] David Copp, *Contractarianism & Moral Skepticism*, presented in '*Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*', Cambridge University Press (1991), p.205 & 220.
[66] David Gauthier, '*Why Contractarianism?*', presented in '*Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*', Cambridge University Press (1991), p.25.

I believe Gauthier would generally respond that sincere and wholehearted internalization of morality provides an agent in society with the optimum chance for success. Furthermore, this kind of sincere, integral and thoroughgoing internalization is the kind that *can't* simply be overridden or set aside when such rare opportunities for profitable transgression present themselves.

So to expand slightly, the Foole is an agent who acknowledges the benefits that come with the general practice of honesty, oath-keeping, etc. but at the same time sees that they should only comply with morality as long as it serves their purpose, but is free to break faith and behave immorally whenever they can get away with it – supposedly securing for themselves the best of both worlds. Gauthier explicitly sees it as his task to show that the 'Foole' is precisely that – a fool – and will ultimately undermine their own best interests by not sincerely constraining themselves to the same degree and in the same manner that they desire those others, with which they interact, be constrained[67].

Where I differ from Copp however, is that I believe Gauthier's refutation of the Foole is founded by what I have referred to earlier as 'sincere', though Gauthier uses the term 'real', acceptance of moral rules. Unless agents in a community sincerely undertake to follow the rules – i.e. they follow the rules for the sake of following them and not due to the expected utility – the expected utility is ultimately unobtainable. If agents are not sincerely honest, Gauthier seems to think that trust will inevitably be diminished, which is disadvantageous to all parties, including the Foole ultimately. Hence, Foole-ish conduct is self-defeating in the long-term as it undermines the system that they are trying to benefit from.

However, is this enough to rule out Foole-ish conduct in every, or even in the overwhelming majority of interactions? Can we really make the contractarian case so strongly that *no* instance of outright deception, contract-breaking, or any other form of unconstrained behavior will *ever* be in the agent's own best interest, overall?

Geoffrey Sayre-McCord (1991)[68] points out that Gauthier's argument seems to depend on every agent within a community being 'transparent' as regards their character and intent – i.e. on it being possible for each member of the community to reliably assess whether or not a potential collaborator will in fact comply with the agreements they make or breach them when it is in their immediate interests to do so.

---

[67] David Gauthier, *Morals By Agreement*, Oxford University Press (1986), p.161.
[68] Sayre-McCord, *Deception & Reasons to be Moral*, presented in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, Cambridge University Press (1991). p.187

The transparency of an agent is the tendency they have for their actions or intended actions being readily predictable to others. Gauthier seems to think that if an agent had the character of a straightforward maximizer, who would renege on an agreement when it appeared to benefit them, this would be somehow reliably accessible or knowable to others – i.e. other parties would know by some means or another that contracts would not be honored by this agent. In other words, an agent's character would be out in the open for all to see, or at least for the most part for all practical purposes. Furthermore, it is the fact that an agent's true character is transparent, or sufficiently transparent (what he terms 'translucency'), in this way that Gauthier seems to assume would be a key motivating factor in justifying their eventual sincere adoption of morality[69]. An agent who was a straight-forward maximizer could be reliably expected to be consistently identified as one and thus excluded from a sufficient number of co-operative endeavors to make their straightforward maximization an overall disadvantage to them. Yet the transparency of agents in Gauthier's account is a crucial element since the rational expectations of the choices of the other agents they interact with, depend on them being properly informed as to the character of those they enter contracts with.

However, translucency of character is by no means certain. In fact, using our own life-experiences as a model; total transparency or sufficient translucency to others and of others is incredibly rare. Sayre-McCord says that more often than not we are only 'translucent' and at times 'opaque' to other agents, and they to us. To be translucent is for others to have only a partial view of our character or a view on which they can only partially rely. An agent would be opaque if they are somehow totally unable to convince others of their good character under any circumstances. To be either will inevitably limit the frequency and extent to which would-be collaborators are willing to put faith in us. Even if the farmer with the earlier harvest is utterly sincere in their intent to provide assistant to their neighbor at a later date, this is not enough. Their neighbor must see or believe that they are sincere and to a sufficient degree that they are willing to risk providing their labor to them for the earlier harvesting.

He goes on to argue that a proper defense of morality, the kind Gauthier is trying to make, tacitly relies on the assumption that all agents involved in the agreement or

[69] Ibid, p.174.

social contract 'have full knowledge of their cohort's character'[70]. Your good character then, is only a guarantee of successful agreement making if one can reliably assume that those we interact with can identify you as a person of good character. Sayre-McCord does acknowledge, assuming transparency *would* 'rule out, by fiat, the possibility of deception'[71].

However, this is a fairly large assumption given the realities of everyday life, of real interactions and of actual human capacities to discern the minds of their fellows. Sayre-McCord argues that the spectrum of possibilities in regards to how translucent an agent is to their fellows and how gifted they are at assessing how authentically transparent their fellows are to them, very significantly alters the calculi of possible interactions – and hence the best possibility for long-term maximization. As such, he argues that an agent could find that they have a gift for being opaque, but also of making themselves *appear* to others as highly translucent or even transparent; whilst at the same time being exceptionally good at accurately reading the character and dispositions of others. In that case, if they found themselves in a community of individuals that were for the most part *actually* translucent or transparent, such an agent might be genuinely better off in not sincerely embracing morality.

Sayre-McCord writes that Gauthier's argument is overly simplistic[72]; that an agent's success at entering into beneficial agreements depends *only* on the facts about the agent's *actual* character. Sayre-McCord points out that the accessibility of the facts of an agent's character to other agent's – i.e. their perceived character – is just as significant a factor in agreement-forming success. In other words, it is naïve to assume that straightforward maximizers would automatically be ruled out of being able to enter into agreements with constrained maximizers, simply because they *are* straightforward maximizers. If they are adapt at concealing their true character, it is only necessary that others believe they are sincerely trust-worthy. Straightforward maximizers 'who are mistaken for being moral may take advantage both of cooperation and of promising exploitation strategies'[73]. This, Sayre-McCord believes, returns us to the original problem.

[70] Ibid, p.187.
[71] Ibid, p.187.
[72] Ibid, p.187.
[73] Ibid, p.187.

The point Sayre-McCord is trying to make is that Gauthier relies too hastily on the transparency assumption in his argument and doesn't take seriously enough the possibilities of being a successful deceiver. Gauthier doesn't need to assume pan-rationality across agents – since it is not essential to his argument that all agents actually accept his argument, only that he can convincingly argue that if they *were* consistently rational, they would do so. Sayre-McCord is also happy to concede that the principles a community would arrive at need make no accommodation for individual tastes and interests, i.e. that the principles are non-tuistic. However, within a community of real people, it is not sound to assume that some will not be better than others at not only concealing but in positively misrepresenting their true character. In addition to this, such a difference will have a significant effect on the relative merits of truly adopting moral dispositions verses merely maintaining the appearance of having done so.

It is true that if one is translucent, in a community of others who are likewise, being honest is a highly rational choice, as it will make one a suitable potential collaborator in a wider variety of ventures. On the other hand though, if one may make one's self appear strictly honest, without actually being so – for at least some of the time – it seems obvious that one will be able to enjoy all the benefits of the sincerely honest agent, but with the added potential for the, at least occasional, successful exploitation of our genuinely translucent fellows.

> 'Sir Desmond Glazebrook: *They've broken the basic rule of the City.*
> Sir Humphrey Appleby: *I didn't know there were any.*
> Sir Desmond Glazebrook: *Just the one. If you're incompetent you have to be honest, and if you're crooked you have to be clever.*'[74]

The simple case of an individual not being reliable to fulfill a single given contract, when they will be having no further interactions generalizes, if in place of forgoing all future interactions our free-rider just gets very good at concealing their transgressions. Given the seemingly obvious reality that such people can and do operate in society all too often, and have historically found great success and wealth from doing

[74] Anthony Jay & Jonathan Lynn, *Yes, Prime Minister*, 'A Conflict of Interest', BBC1 1987, IMDB.

so, it does at least seem like a real problem for Gauthier. Why be honest if you can – and some people *can* – be successfully dishonest?

The response to this problem, inspired by Gauthier's own writings on the subject[75], can be broken down into two major strands; the game-theoretic and the pragmatic. The former I hold to be unsuccessful, the latter, less so, but not enough.

Let's look at the game-theoretic response first then. Let's call an agent who is in reality a straightforward maximizer but is superlatively adept at appearing thoroughly honest 'trans-opaque'. Other agents *believe* the trans-opaque agent is transparent and that they can see their true character, but they are in fact opaque and able to conceal it. Now, can one make a sound case for trans-opacity as opposed to sincere honesty? If an agent is trans-opaque rather than sincerely honest, regardless of how good they are at maintaining their presentation of sincere honestly, there will always exist for that agent a lurking threat of discovery as a knave that can't ever be fully eliminated. The sincerely and consistently moral agent on the other hand, has no such fear of discovery, for there will be nothing *to* discover.

So potentially devastating to the reputation of an agent, and the inevitable curtailing of eligibility for future beneficial co-operations with others within a community of honest people, should even one major act of infamy be discovered, that a case could be made that the only rational long-term strategy is to eliminate this risk entirely by sincerely adopting morality. In other words, that which constitutes the key difference between the conduct (or potential conduct) of a trans-opaque agent and a sincerely moral one, is one that is *essentially* of greater risk/cost than anything that the violation of the agreed upon precepts could hope to garner for the transgressor. To put it another way, given the inherent risk of discovering, the expected utility is and always will be greater for the sincerely moral agent than the trans-opaque one. It is therefore in any agent's interest to be the former rather than the latter.

This however is not convincing. I turn now from the points made by Copp and Sayre-McCord, which have informed the bulk of my criticism up until this point to my own argument as to why Gauthier has failed to meet the challenge of rational compliance.

---

[75] Gauthier, David, *Rational Constraint: Some Last Words*, presented in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, Cambridge University Press (1991), p326.

It might be true if we assume that all such encounters take place within a framework of ongoing interaction, where even a single lapse might be enough to lead to expulsion from the community. But it can't apply to stand alone agreements, as with the two farmers. Suppose that the farmer with the earlier harvest is trans-opaque rather than sincerely honest. Their commitment to abiding by contracts has been guaranteed up until now because they were not moving away to Florida. Therefore, their undertaking of all previous agreements, either with their neighbor or those of which their neighbor would have been aware, would have been sincere. But this ultimate agreement takes place within rarefied conditions. Having played the long game, so-to-speak, the farmer finally feels the sweet release of being in a position to profit significantly from their lifetime of being trustworthy. Being trans-opaque would pay if one reserved ones transgression for one big score[76]!

It is simply a fact that life could very well throw at the trans-opaque agent an opportunity where the precise circumstances make the potential gains worth the risk of transgressing, when looked at on any valid cost-benefit analysis. Without any foregoing moral concepts to rely on, Gauthier has no convincing case to make as to why an agent would be acting irrationally to transgress the moral law.

What I refer to as Gauthier's pragmatic considerations in defense of sincerely adopting a moral disposition, could be seen as echoing, at least in part, Plato's city/soul analogy. The reader will recall that Socrates answered Glaucon's skeptical challenge by stating that the soul of the unjust would be always out of kilter, in a state of disharmony and divided against itself. In a similar vein, Gauthier questions how effective a strategy it is for long-term success to be in a state of effective cognitive dissonance the like of which would be necessitated to have a viable chance at pulling-off trans-opacity with any degree of worthwhile profit.

> 'The self that agrees and the self that complies must be one.'[77]

[76] Criminal history is replete with examples of this sort of crime. The bank teller of forty years, for example, implicitly trusted by all their colleagues, having spent their whole working life dotting every 'i' and crossing every 't', finally makes off with a small fortune out of the vault and escapes to the Costa del Sol. Or a 'Walter White' type who, on the realization they are dying, trades on their squeaky-clean, average-Joe public perception in their last few months of life, to become a meth dealer and so secure their family's financial future after they've gone.

[77] Gauthier (1991), *Why Contractarianism?*, p.30.

The trans-opaque agent would have the 'the self that agrees' be one that only acts out some semblance of agreement-making – a pretence of commitment only – rather than *actual* agreement making[78]. Gauthier does not couch this within an argument for what is ultimately beneficial to the agent, *a la* Plato. Rather he is making the point that unless the self is a unity, sincerely committed to adhering to its agreements, *it can't be said to be capable of making agreements at all*. The duplicity or cognitive dissonance of the kind necessary for being trans-opaque renders agreement making extremely difficult and to a degree that undermines its overall utility to the agent in question.

I believe Gauthier's point is sound here. But what of it? Whether the agreement is real or illusory is surely irrelevant. Only whether or not the intended dupe is fooled, matters. I am forced to return to the hard, empirical fact that it is frequently possible for the sufficiently gifted con-artist to get away with their crimes, either undetected or unhindered, long enough to make their profit. It is hard to see, given the gifts for deception which nature has given them, how they are guilty of irrationality by exploiting them to the full by exploiting the honest.

In my estimation, Gauthier has not been successful in solving the problem of rational compliance. This failure has vital ramifications for my assessment in the next section, of how well he has provided reason to be moral that meet the categoricity and right grounding requirements, for being moral reasons.

3.4 The Poverty of Egoism

Gauthier's theory is compelling. It provides reasons to behave morally, *qua* restraining from certain action-kinds that agents might otherwise consider they have reason to carry out, founded on the reasons of self-interest that all but the most ardent normative skeptic accepts as real. With its assertion that the Archimedean Point is the place from which moral authority reigns down, it seeks to provide the equal and universal applicability of moral reasons to all rational agents. Additionally by accommodating all of the things intuitively held to be morally imperative by people living and dealing with one another, the moral reasons provided are extensionally adequate.

[78] I am put in mind of Wittgenstein's remarks on the impossibility of a genuine private ostensive definition. To paraphrase §268 of *Philosophical Investigations*, 'the right hand may give the left hand money, but it may not be called a gift'.

We are now in a position to assess how well Gauthier's theory is able to meet my three criteria for reasons for action being genuinely moral. It is my position that he fails to meet two of them, but does fairly well at meeting the third. While other writers have covered some of the ideas discussed here, their specific application to Gauthier's model is, to my knowledge, original.

To summarize in advance the single biggest reason it fails is for the simple and fundamental reason that the reasons that can be provided by any thoroughgoing egoism are inadequate to the task of grounding moral reasons. This is because it is in the nature of morality for it to call upon agents, at times, to do things that could not by any reasonable measure be considered in that agent's interest.

The simplest possible example of this would be that moral reasons are the kinds of things that could call on a person to throw themselves onto a grenade to save their comrades, sacrificing their own life in the process. Such examples of extreme self-sacrifice and charity are paradigmatic of the moral, yet wholly incompatible with egoistic self-interest. To put it starkly, the reasons of morality and the reasons of self-interest are all to often diametrically opposed to one another. This perennial possibility of mismatch means that there is an inherent poverty to the caliber and range of reasons that egoism can provide, but are necessary to ground some moral reasons.

3.4.i The Morality of the Last Man

Recall that for a reason to be categorical it must apply to an agent whatever their individual circumstances or psychology are, and whatever their motivational states happen to be. Either a moral reason exists to φ or it doesn't. If it does in cannot cease to apply merely because in some particular situation an agent finds themselves in, φ-ing is not actually in their interest. That is not to say that a moral reason can't be outweighed by countervailing reasons or some other considerations, at times. But the reason not being the strongest or best reason for an agent to act is not the same as it not apply altogether.

So, for Gauthier to have met the categoricity requirement he would have to have shown that acting morally is *invariably* in an agent's interest. This however, he has not done – not by a long way. If we run through the basic rationale,

1) All agents have the best reason to act in pursuit of their greatest long-term advantage.

2) Being a secure member of a fair and just society is invariably to the greatest long-term advantage of any agent.
3) For any agent to be secure in their membership of a fair and just society, they must act morally consistently and without any major transgressions of the laws or established moral rules of that society.
4) The only way an agent can ensure that they act morally consistently and without any major transgressions of the laws or established moral rules of that society is by cultivating sincere moral virtues and becoming genuinely moral.
5) Therefore, all agents have the best reason to cultivate sincere moral virtues and become genuinely moral.

For the sake of discussion, we'll grant the truth of (1)-(4). The argument is valid. However, it is insufficient to establish the categoricity of moral reasons as it is entirely contingent on an agent's secure membership of a fair and just society. Most of the literature, likewise, makes this assumption. My argument, however, is that is goes to the essence of the question of categoricity of the reason to behave morally is, what happens if an agent were outside of anything we would typically regard as a functioning society? Perhaps they are in some very remote part of the world or have survived a global holocaust where little of humanity survives and all governments have been annihilated? Is it still going to be to the agent's greatest long-term advantage to behave morally? It seems obvious to me that under such circumstances, they wouldn't. However, that does not reduce my conviction that moral reasons, as I define them and if they exist, still apply to agents even in these conditions.

A defender of Gauthier might argue, that it would be in the best interest of any agent in these circumstances to re-establish some semblance of society or community with whoever's left, and that would still be best achieved by establishing rules and laws. However, it really wouldn't be in their self-interest to enter into a social contract or any agreement with a group of say, mentally handicapped, infirmed or disabled people they might come across, if these people had nothing to offer him by means of co-operation but only as possible sources of exploitation. In such case, I don't see how the agent would still have a reason to act morally toward them – with 'morally' here being understood on Gauthier's terms.

Yet murder is still murder and exploitation is still exploitation, regardless of whether they take place within society or outside of it. It is still wrong to wantonly

torture and kill a person, even if you are the last two people alive and by every metric, you have significantly more to gain by doing it than not doing it.

The key weakness of Gauthier's theory is its absolute dependence on moral actions being ultimately to an agent's interests. But that is not the limit of the kinds of action we typically think an agent can have moral reasons to do. I mentioned earlier the example of the soldier laying down their life to save others. Moral conviction is often needed most when the stakes are highest. The higher the stakes for the agent however, the more in conflict the demands/duties of morality would seem to be in conflict with those of ultimate self-interest. To put it another way, the scope of what we would intuitively want to call morality, could legitimately call upon an agent to sacrifice, is *prima facie* limitless. But the scope of what Gauthierian 'moral' action could legitimately call on an agent to do, is inextricably limited to what in the final analysis serves self-interest.

It might be responded that the kinds of scenario's I am describing are extreme or far-fetched. The vast majority of humanity will have the opportunity to be part of some society and that requires adopting morality. Well, even if I grant this, it is not the point. The fact that an agent *could* be in a situation where the purported moral reasons provided by some theory, genuinely do not apply to them is sufficient to show that those reasons are not categorical. This aspect of my argument is where I digress from many critics of Gauthier, who seem to tacitly by into this limited view of what morality can theoretically call upon an agent to do and its utilitarian aspects – and then argue against it from within this assumption. By rejecting this limited view, I add to the criticism by showing that the inherent limits of egoism leave it constitutionally incapable of meeting the extensional requirements of any satisfactorily categorical moral theory.

3.4.ii '... forfeiture of all future trust and confidence with mankind'[79]

As I wrote above, Gauthier's theory doesn't get everything wrong. The reasons to behave morally that he provides do appear to me to be perfectly weighty, as and when they do apply to agents (not withstanding the lack of true categoricity I discussed in the previous subsection).

[79] David Hume, *Concerning the Principles of Morals*, Oxford University Press (1999), p.283.

In Gauthier's theory merely behaving morally allows agents to be members of society, which is greatly to their advantage. However, *genuinely* becoming someone of good moral character, secures an agent's place even more firmly in society. Remember, for Gauthier the genuinely moral person is one who observes morality not for the gain it gives them, but because they have internalized morality to such an extent that they now act purely for the sake of the values exulted in that moral system.

Furthermore, not only does being outside of society mean greatly reduced opportunities for advantage, but to commit a palpably heinous deed, such as mass murder, would actually put an individual in direct state of conflict with the rest of society. No agent would be able to resist the whole of their society were it to align against them. It in all cases it would either result in the execution or incarceration, or else they would have to be an outcast or fugitive for the rest of their lives.

Although Gauthier does not spend any great amount of time discussing the weight of the moral reasons he provides, I believe a very good one can readily be extrapolated from it. This is also, as far as I know, and original point regarding Gauthier's theory. Where the literature does discuss weight in this context, it is typically in reference to the strength of reason one has to comply with the mores of a given community as opposed to the possible benefits of transgressing them, taken as a whole system – i.e. the weight of an agent's reason to be 'in' or 'out' of the game generally. My argument however focuses on how Gauther's theory can be extrapolated fairly readily to account for the difference in weight of the reasons to adhere or abstain from individual morals.

Within almost every society, certainly in Western egalitarian ones, there is a rank ordering of transgressions of the law and the societal response for the violations of certain cultural mores. This rank ordering more often than not reflects the moral values of the society. According to the rank ordering, different crimes will come with different punitive measures. Murder almost always carries a stronger sentence than bank robbery. Bank robbery carries a greater sentence than shoplifting. And shoplifting carries a stronger sentence than jaywalking, and so on.

Equally in non-criminal matters; cheating on one's partner is more frowned upon than using extremely vulgar language in front of small children. Using extremely vulgar language in front of small children is, arguably, more frowned upon than telling a casual acquaintance a white lie to get out of a social engagement. Each one would be

expected to garner a different degree of societal condemnation or ostracization for the agent in question, by their family or community.

So Gauthier's theory can be seen to accommodate the relative weight of moral reasons in a twofold manner. Firstly because, each agent has the best reason to be of genuinely good moral character, they will to a large extent have internalized and accepted the rank-ordering of transgressions prevalent in their society. This will, by its very nature, lead them to have moral (in Gauthier's sense) reasons of different weights to do and not do different things within their society. Secondly, because of the legally or socially punitive consequences associated with the different kinds of transgressions, each agent will have a straightforwardly egoist reason to refrain from them, for fear of being caught.

To summarize then, almost all systems of morals provide an account of the weight of the moral reasons have to do and not do things, which is in accord with the values of that system. Since compliance with this system is vital for the wellbeing of any agent, these reasons will have the associated weight for the agent in that society. These will be quite in keeping with our normal intuitions regarding the weights of moral reasons, as they will be modeled on the values already prevalent in our societies.

Assessed purely when it comes to accounting for the weight of the kinds of moral reasons Gauthier believes he provides, he has been successful. Or, to put it more accurately, in no sense is his theory antithetical to providing an adequate account of the relative weights of moral reasons. Additionally, as I have gone a small way towards doing here, a more comprehensive account can be readily provided using Gauthier's basic theory.

3.4.iii Inherently Selfish Grounds

The single biggest shortcoming of Gauthier's theory, in terms of meeting my three criteria for the kinds of 'moral' reasons it provides, is its abject failure to give anything even approaching the right grounding.

So what is the right grounding? As I've already discussed, the right grounding must include certain features that go at least some way to accounting for the differences in reactive attitudes we typically have toward them in terms of whether we deem them of either praise or censure. In practice what we're usually talking about what reasons we have to do things that are morally right or wrong, has to be in the interest/welfare of

some individual other than the agent who performs it, or else the goodness of the end intended or achieved.

Now a defender of Gauthier might object that he does incorporate this intuitively necessary characteristic of the moral into his account. To be a superior prospect for collaboration by others, the agent has to have *sincerely* adopted a moral mindset – i.e. it is the fact that they will often act for the sake of others without regard for themselves or in a self-sacrificial way (in our sense 'moral'), which ultimately, and paradoxically, serves their long-term best interest. But this is not satisfactory. Again, I consider this to be an original departure from much of the existing literature, most of which focuses on the pragmatic reasons an agent does or doesn't have to live within the social contract. The point that I raise here specifically, on the other hand, is that by making the ultimate reason for any rational agent to comply with morality, self-interest you make the grounding one of advantage to the agent and not the goodness of the action.

Even though in this case we still might be willing to praise the civilized agent who has fully and sincerely internalized morality, and conforms with it consistently, we could still acknowledge that if it were not for self-interest no agent would have reason to act in a morally praiseworthy manner. This lack of justification for truly selfless but still praiseworthy action is indicative of a mismatch between the kinds of reasons Gauthier is providing and those we should consider moral reasons, in my sense.

Whilst on this account, we could say that the Gauthierian agent has a reason to act morally, they do not yet have a moral reason to act that way. The justification for moral action on Gauthier's account ultimately rests on the (for the purpose of this subsection, granted) utility of moral behavior in securing long-term self-interest. Take away this ultimate utility and the agent would no longer have any reason by Gauthier's lights to act morally. Yet one *always* has a moral reason to act morally.

If we grant for a moment that there are moral reasons for action; if an innocent person is on the floor, writhing in agony, then all things considered we regard that you have a moral reason to render what help to them that you can. That is to say, if we have a moral reason to do anything we have a moral reason to help innocent people writhing in agony. Yet as I have argued previously, surely our sense of there being a moral reason to help such a person is grounded, at least in part, by our sense of the goodness of the intentions of the agent or the state of affairs that results in alleviating unnecessary or unjust suffering. This is indeed linked to why we typically think that moral reasons

apply to agents regardless of what their motivational states are – what I have called the categoricity requirement. In the other words, the moral reason tracks the intended or actual outcome of the action, not the nature of the motivational state that provides for it.

If it were not the case that moral reasons have to be of a certain kind, how else would we account for our strong intuitions concerning the attribution of praise or blame when it comes to assessing peoples actions? If someone does save a person's life, from drowning, say, our intuition to morally praise that person would be severely stunted, if not eradicated entirely if we were to discover that this person is a sociopath who only saved the person for the acclaim they would receive and the possibility of a reward. I am not suggesting of course that our intuitions are indubitable in such matters. However, just as I think the failure of a fundamental philosophy of mathematics to align with our most basic arithmetic calculations – e.g. 1 + 1 = 2, etc. – is indicative of a major problem with it, so is the failure of any moral theory to account for the differences in the different kinds of reasons we clearly intuit and react to when considering reasons to act morally, highly indicative of that moral theory's overall validity. I put it no more strongly in this thesis.

Now Gauthier might respond by saying that these sentiments all play an important role of encouraging the *sincere* adoption of morally good character, as this serves every agent's best long-term interests. However, this is not satisfactory either. No amount of establishing the rational foundations of moral behavior undermine our sense that there is some end to moral action itself which is worth attaining, regardless of any secondary, amoral utility it may have.

On Gauthier's account, justification for moral reasons, *qua* pragmatic reasons, is contingent upon it doing something besides promoting morally desirable ends. In this way, moral reasons are parasitic on pragmatic reasons. Even if we could eliminate the possibility of mismatch between moral and pragmatic reasons – i.e. Gauthier or someone like him, could demonstrate that moral reasons and self-interested reasons were necessarily co-extensive, thus putting an end to my categoricity objection – it would remain the case that it is not the moral dimension of the reason that is doing the work. To be the right grounding, we expect, *nee* demand, that moral reasons do their own 'heavy lifting', so-to-speak. A moral reason carries its own force – not one derived from an extrapolated purely pragmatic bedrock. Without this unmediated connection

between the constituents of an action/situation and its wrongness, the moral dimension to any theory like Gauthier's will be mere epiphenomena.

Gauthier has provided us with an excellent theory of pragmatic reasons as to why we should behave morally. What he has not provided are moral reasons.

Chapter Four

Schroeder

4.1 Introduction

As discussed in Chapter Two, a crucial problem with all forms of Internalism is the motivational limitation it places on the kinds of normative reasons it allows to exist. To be a normative reason for action the reason must be capable of potentially playing some role in motivating the actions of an agent. At first glance then, this presents a challenge for the provision of truly categorical reasons for action. Intuitively, we think what motivates an agent must be dependent in some way on the content of their psychology – and hence, not categorical.

However, if an internalist could demonstrate that there are reasons for action that are capable of motivating that *every* agent has whatever the individual content of their psychologies are – i.e. motivating reasons that are genuinely 'agent-neutral' – then they will have met the categoricity requirement. Enter Mark Schroeder.

4.2 Hypotheticalism

Schroeder advocates for a theory of reasons that he dubs *Hypotheticalism*, which is in the tradition of a broadly Humean picture of reasons and of motivation. The Humean Theory of Reasons (HTR)[80], you'll recall, states that for something to be a reason for an agent to act it must be capable of motivating that agent and the only thing capable of motivating an agent is one of their desires. Hence, for there to be a reason for *A* to φ, *A* must desire to φ.

The Humean Theory of Reasons is highly intuitive and has been strongly adhered to by many philosophers of action during its history. A major criticism however has often been its inability to provide moral reasons in all the cases we would like them to be provided. Oftentimes, agents do appear to have genuinely no desire to do the morally right thing. Yet intuitively and typically we do not hold that lack of desire means they do not have a moral reason, all the same.

If a theory is to provide the kinds of normative reasons we require it to, it will be necessary for it be able to generate truly agent-neutral reasons. These are reasons which do not apply to any specific agent, but are reasons for any agent. This, in turn, will

[80] Mark Schroeder, *Slaves of the Passions*, Oxford University Press (2013), p.5.

require a theory where an agent will have at least some reasons to do things whatever their actual desires are. A problematic desiderata for any Humean theory. However, Schroeder believes that Hypotheticalism can do just that.

In a nutshell, he believes it does this by establishing that for *any* given action there will be *some* desire that an agent has, that's object will be promoted by carrying it out – and hence, a reason to do it[81]. His key to arguing for this is firstly to show that reasons for action are far more abundant than we might usually think; and secondly that the strength of our reasons for doing things is not dependent on the strength of our desire to do that thing. Respectively, he goes about this by rejecting *No Background Conditions* and *Proportionalism*[82].

Schroeder asks us to consider how a reason is constituted. A frequent assumption associated with HTR is that if desires are a necessary requirement for their being a reason for an agent to act, then the desire is a constituent part of that reason. Schroeder however says that this assumption is ill-founded. In one analogy of his, he points out that having grown on corn plant is a necessary part of what makes something a serving of corn-on-the-cob. On the other hand though, the fact that something has grown on a corn plant is not *part* of any individual serving of corn-on-the-cob. Having grown on a corn plant is not a constituent part of the corn-on-the-cob; it's merely a background condition of what makes it corn-on-the-cob[83].

This is relevant to an analysis of reasons for action when we ask ourselves what the nature of our everyday reasons for action are. If Susan desires a coffee, that might be a reason for her to go to the staff room, since there is coffee in there. On the No Background Conditions view her desire for coffee combined with the fact that there is coffee in the staff room *is* her reason to go to the staff room. But for Schroeder, it is *only* the fact that there is coffee in the staff room that makes it a reason. The fact that Susan desires a coffee is merely what makes it a reason *for her*. After all, Susan's desire for coffee might give her a reason to do any number of different things[84]; get up and walk down the road to a café, or ask her office-buddy to go and make coffee for her. Conversely, the fact that there is coffee in the staff room may give her reason to go in there, even if she has no desire for coffee at all – she might even hate coffee. It remains a

[81] Ibid, p.121.
[82] Ibid, p.23-27.
[83] Ibid, p.24.
[84] Ibid, p.30-31.

reason for her though, as there is a colleague she needs to talk urgently and her colleague's passion for coffee means there is the greatest likelihood that they will be in the staff room.

For Schroeder a reason is, rather straightforwardly, a fact about the world. What makes it a reason *for* a given agent is a desire[85]. Crucially though, the specific desire that's object is promoted by an action, is not necessary for that fact to be a reason. So the desires that make certain facts reasons may be seen as variables in the constitution of those reasons. It also follows from this that every reason is at least partially overdetermined by desires – i.e. every reason for an agent to act can be explained by the having of more than one desire[86]. We will come back to the importance of this overdetermination shortly.

Schroeder also identifies two problems we've seen already in Chapter Two. These are the problems pertaining to the extension of moral reasons, which I consider perennial to Internalism – the *Too Many Reasons* problem and the *Too Few Reasons* problem. Essentially, the general issue with these is that if an agent's reasons somehow depend on the provision of agent's desires, desires do not seem to track well with the kinds of things we want to end up being normative reasons to act and not act.

With *Too Many Reasons*, the problem is explaining why someone with an extremely eccentric or even nefarious desire does not in fact have a reason, or at least, a good reason to carry it out. To use an example of Sharon Street, there may be someone whose dearest wish is to spend the rest of their life counting blades of grass. Or there may be someone like Caligula, who delights in the torturing of innocent people for their own amusement. If desires give reasons, the Humean might be committed to saying that these people really do have reasons to undertake these activities.

With *Too Few Reasons*, the problem is explaining how an agent can have a reason to do the sorts of things we intuitively think they must have reason to do even when, seemingly, the requisite desire that would give them a reason to do it is lacking. If Katie is in desperate need of help then, *ceteris parabus*, a reason exists for Ryan to provide that help. We think this is so regardless of *any* desires Ryan may or may not have.

Let's attend to Schroeder's handling of *Too Few Reasons* first. Schroeder sees this as being basically the problem of how one can square the existence of agent-neutral

[85] Ibid, p.29.
[86] Ibid, p.109.

reasons with the HTR. There are some reasons that we want every agent to have, whatever their desires happen to be. Here we return to the overdetermination of reasons by desires. If every reason is at least partially overdetermined by more than one desire, Schroeder holds it as at least conceivable that they may be some reasons that are '*massively* overdetermined'[87]. There may be reasons that are such because they promote the object of almost any given desire.

His solution to *Too Few Reasons* then, is to set the existential bar for something to a be reason far lower than it is typically and intuitively thought to be – i.e. by making reasons easy to find. The classic example Schroeder provides is that an agent has a reason to eat their car. They have this reason in virtue of the car containing the agent's recommended daily amount of Iron. It is of course counter intuitive and atypical to say that such a reason is a reason at all. Yet Schroeder maintains it remains a reason of sorts, none-the-less.

So, if there are reasons that are massively overdetermind by desires, there may be reasons that any agent has simply by virtue of having any desire whatsoever. For Schroeder, these would be genuinely agent-neutral reasons for action as presumably, any agent must have at least some desires and as such, they can't help but to have these reasons[88].

Agent-neutral reasons have more often than not been assumed to require a dyadic formulation – i.e. 'There is a reason *R* to carry out action *A*'. In such formulations, no reference to a specific agent or group of agents is made. However, Schroeder's view is that it should be regarded as a triadic relation more typical of agent-relative reasons like, 'There is a reason *R* for agent *x* to carry out action *A*'. The point being that since *R* is massively overdetermind, '*x*' shall include everybody.

Ryan may not have a specific desire to help Katie, but, Schroeder argues, he will have some desire that's object will be furthered by helping her. This is because the reason to help those in need, or something that is entailed by it, is the kind of reason that is massively overdetermined and hence agent-neutral. Schroeder is optimistic that the same will apply to most universal normative reasons we believe do or should apply to everyone.

[87] Ibid, p.109.
[88] Ibid, p.18.

As for *Two Many Reasons*; Schroeder is prepared to bite the bullet on what may seem at first glance like a profoundly counter-intuitive conclusion – but that actually turns out to be far more benign in the context of his theory as a whole[89]. He accepts that our would-be grass blade counter does have *some* reason to fritter their life away counting blades of grass; and that our Caligula wannabe does have *some* reason to torture innocent people for their own amusement. However, the extremity of this stance, he argues, is heavily mitigated by the fact that despite the intensity of their respective desires, they are reasons that are about as weak as it is possible for an agent to have. This is because he rejects the premise that the weight of a reason derives from the force of the desire that accounts for it – a view Schroeder refers to *Proportionalism*.

So, as we've already said, we take it that an agent's reasons for action vary in their 'strength' or 'weight'. We believe that while it may be true to say that we have valid reasons both to φ and to not φ, simultaneously, the reason to do one over the other may be far greater. Deliberating between the weight of our reasons to do mutually exclusive things may reveal what we have *most reason* to do (or as some theorists regard it – what we *ought* to do). A feature of HTR, which Schroeder argues has simply been taken for granted, is that since reasons are dependent on desires – i.e. desires are a necessary part of what makes a reason a reason – it is the strength of the agent's desires that (at least partially) determines the weight of the reason the agent has to carry out that action. He regards *Proportionalism* as being a tacit assumption of many versions of HTR, but is not actually integral to it. Liberated, in part, by his rejection of *No Background Conditions*, he is able forge a conceptual wedge between the weight a reason has and the strength of the desire for some object that makes it a reason in the first place[90].

Schroeder makes little specific argument against *Proportionalism*, *per se*. Instead, in keeping with Hitchen's Razor[91], he appears to hold that in keeping with most ubiquitous assumptions, little argument or evidence has ever been presented in its favor of it, thus little is required to unseat it. He speculates that the prevalence of *Proportionalism* has been chiefly because it is highly intuitive and because there has,

[89] Ibid, p.100-101.
[90] Ibid, p.97-102.
[91] "What can be asserted without evidence can also be dismissed without evidence."

until now, been few viable contenders for an alternative account of how reasons obtain their weight within the Humean tradition. Schroeder outlines one[92].

The weight a reason has, he argues, is not a function of the strength of the desire that explains it, but rather a measure of the appropriateness[93] of an agent leaning on that reason in their deliberations of what to do. If a reason's strength were variant with the intensity of the desire felt for its object, then an agent might take a course of action in pursuit of a more greatly desired object over a slightly less desired object, even if the probability of obtaining the more desired object were significantly lower than obtaining the only slightly less desired object. Surely, Schroeder opines, we would consider an agent who did this to be acting practically irrationally[94]. Yet if we are correct to always pursue the objects we most desire and simply did just that, deliberation would be in a sense incorrigible. Schroeder insists however, that there must be some standard of correctness in deliberation. An agent can get it wrong by placing to much or two little weight on a reason, independently of the strengths of their respective desires.

Schroeder argues that what provides the standard of correctness for appropriateness of deliberation is context - in the kind of activity one is engaged in. In playing chess, a player has a very strong reason not to castle out of check since this is a violation of the rules of chess. One might have a stronger reason to do it however – if say one was being offered a large sum of money to do so or had a gun pointed at their head. But regardless of the agent's desire/motivation to break the rules of chess, when deliberating the next move in your capacity *as a chess-player*, the weight of not doing so is the stronger[95].

Here Schroeder begins to sketch a case for why certain types of normative reasons, particularly the types of agent-neutral ones we deem to have the strongest kind of reasons for following. The idea seems to be that there are certain things an agent has reason to do simply by virtue of being involved in the process of deliberating what to do in the first place – i.e. the process of deliberation itself provides the context. For example, in deliberating what course of action to take an agent has reason to employ sound principles of reason, such as how best to achieve ones aims.

---

[92] Ibid, p.129-136.
[93] Ibid, p.129.
[94] Ibid, p.130-131.
[95] Ibid, p.135.

However, unlike the chess example, deliberation is arguably a universal and unavoidable context. So long as an agent is deliberating, they have reason to put weight on sound principles of reasoning[96]. (In my opinion, this is a form of constitutivist or at least pseudo-constitutivist argument that shall be discussed in greater depth in Chapter Five).

If Schroeder is correct, that there are objective standards of correctness or appropriateness when it comes to deliberation, then two things follow from this. Firstly, these reasons will be weighty, as they are by their nature the reasons agents have most justification to lean on when deciding what to do. Secondly, they will be reasons that *any* agent would have equally, in the same circumstances. They would be agent relational reasons for *all* of us, regardless of the individual agent's psychological make-up – in other words, whatever their desires are. Hence, according to Schroeder, there would be universal agent-neutral reasons, not only that all agents had but also that were *weighty* for all agents. It is the crux of Schroeder's endeavor, that reasons to behave morally, say helping those in life-threatening danger for example, might turn out to be just this kind of universal, agent-neutral reason[97].

But Schroeder's account isn't quite finished yet. He has asked us to entertain the possibility that our core moral reasons may well be these universal, agent-neutral reasons he has argued are possible. Even if we grant that they are, however, he must give us some additional argument for why he thinks there is good reason to think that there actually *are* such reasons. To this end he offers an account of moral virtue, the interconnectedness of reasons and the unity of actions and deliberation, largely inspired with one given by Aristotle. To lay the groundwork for how this would work with reasons for acting morally, I will use his epistemological analogy.

Mary wishes to by some shoes. Whilst it is not inconceivable that she would still be able to do so successfully, even if she was not deliberating soundly or her sound deliberations were based on false beliefs, unquestionably the fewer false beliefs she is in possession of, the greater her chance of managing to buy a pair of shoes. Schroeder argues that there are some beliefs – e.g. where a shoe shop is actually located, when it is actually open – that Mary having false beliefs about would substantially reduce her

[96] Ibid, p.131-132.
[97] Ibid, p.141-143.

chance of successfully buying shoes[98]. Therefore, her desire to buy shoes gives her a reason to only believe statements concerning these things if they are true. Seems reasonable. However, he makes the stronger argument that a false belief *anywhere* in her web of belief will have a marginal disutility in allowing her to buy shoes. If not because the belief directly bears on the task in hand, any false belief may undermine other beliefs in her web, which may undermine others, and so on until a belief that *does* directly bear on her buying shoes is compromised. From this Schroeder believes that Mary has a weighty reason to only believe true statements, on any subject, because they affect her ability to deliberate and promote her desire to buy shoes. The same would be true, according to Schroeder, of any agent and whatever they are trying to achieve. Hence, there is a weighty, agent-neutral reason to only believe statements if they are true.

Although he provides no clear or thoroughgoing examples, Schroeder suggests that certain reasons to act morally may turn out to operate in the same way. There may be certain actions – helping those in life-threatening danger, say – which carrying out will always go someway to promoting your desires; and failing to do will always impede. As such, this will mean that there will always be a weighty, agent-neutral reason to do so. Remember, with the rejection of *Proportionalism*, it is not necessary that the desires for the objects promoted be strong ones for the agent to have a reason, just that they *are* promoted by the action.

However, though Schroeder consistently maintains that strength of the desire an agent has for some object does not equate to the strength of the reason for that agent, he does acknowledge there is a connection between the desire felt for some object and an agent's motivation toward that object. There is therefore, a weighty reason for any agent to desire to be motivated by those things they have best reason to do. In other words, since they have weightiest reasons to behave in accordance with their agent-neutral reasons, the agent who's desires most readily incline them to act in accordance with what they have best reason to do are, *ipso facto*, the most practically rational. For Schroeder then, the 'virtuous' person is the one whose sincerest and most heartfelt desires best align with their strongest reasons to act[99].

[98] Ibid, 113-115.
[99] Ibid, p.168-170.

So, to summarize; Schroeder believes that every reason to act is a reason *for* an agent in so far as they have a desire that's object is promoted by carrying out that action. In this way he is both an internalist and a Humean. The desires that make reasons reasons are not constituent parts of those reasons, only background conditions of them being reasons. This means reasons can be overdetermined by desires. This being the case, Schroeder postulates, some reasons may be so massively overdetermined that any desire whatsoever an agent has will mean that they have that reason. If that were the case, there would be reasons that all agent's had in as far as they had any desires at all. Such desires would be agent-neutral by his definition of the term. The weight of a reason is not proportional to the strength of the desire that makes it a reason, but rather by the appropriateness of utilizing that reason in our deliberations. Since there are objective standards of appropriateness or inappropriateness, correctness or incorrectness provided by the nature of sound reasoning, any agent-neutral reasons there turn out to be will be equally weighty for all agents. Reasons to behave morally may well turn out to be agent-neutral reasons of this type and as such we will all have strong reasons to adhere to them regardless of our actual psychological make-up – whatever our actual desires are.

4.3 First Impressions

I'll begin by reiterating something already mentioned in Chapter Two. The version of Humeanism that Schroeder is advocating in *Slaves of the Passions* is called Hypotheticalism. This is in part meant to highlight the indispensible place that desires (or some other motivating psychological state) plays in his thesis. The hypotheticality of reasons in Schroeder's account, then, is baked-in from the start. It might be thought then that it is constitutionally incapable of meeting my insistence that truly moral reasons be categorical.

However, while Hypotheticalism necessarily requires the presence of *a* desire in order for reasons to exist for an agent, it does not necessitate that a specific desire account for that reason. If moral reasons turn out to be agent-neutral, by Schroeder's lights, then any desire will explain them, which means that no agent could fail to have them[100]. Furthermore, they are *always* weighty to an agent – i.e. an agent can be deemed

[100] Like Schroeder, I cannot conceive anything I would be willing to consider 'an agent', that did not have at least one desire.

in error for not abiding by them – regardless of what their psychological make-up is. If successful, these inescapable, albeit ultimately hypothetical, reasons would be sufficiently categorical-like (quasi-categorical) and I would consider the categoricity requirement adequately met. But is it successful?

My criticism of Hypotheticalism breaks down into four main points. Firstly, there is the question as to whether or not there really could be any reason *so* massively overdetermined that it could meet Schroeder's definition of an agent-neutral reason. Secondly, even assuming that there were agent neutral reasons, what reason do we have to believe that moral reasons, or indeed, any moral reason at all will turn out to be agent-neutral ones. Thirdly, there are significant problems presented by Schroeder's own positive account of reason weight. And fourthly, even if there are moral reasons that are agent-neutral, quasi-categorical and sufficiently weighty, they don't come close to being the right grounding. I will deal with each of these in turn.

The first, second and fourth points are based on my own individual reading of Schroeder. However, as to the third point regarding Schroeder's account of reason weight, I draw on the work James Lenman & David Enoch. I then demonstrate how their more specific criticisms of Schroeder can be used to illustrate a more general problem deeming Schroeder's account of weight as being an internalist one at all.

4.3.i Reasons for *all* of us.

When it comes to this I'd like start with the obvious. Despite being an incredibly rich, wide-reaching and sophisticated body of thought, in the two-hundred and twelve odd pages of *Slaves of the Passions*, I can't find a single actual example of a fully worked out or coherently formulated agent-neutral practical reason. They are highly conspicuous by their absence!

The closest Schroeder comes to providing an example of an agent-neutral reason is the example I've already mentioned of Mary, and her attempt to buy shoes. Here Schroeder is attempting what is sometimes referred to as a 'companions in guilt' strategy. This is where an analogy is drawn between principles of abstract reasoning pertaining to valid inference and deduction concerning facts and data – e.g. modus ponens, law of the excluded middle, etc. – that we consider essential to sound reasoning, and practical reasoning. It is thought that since practical reasoning also relies on objective standards of sound reasoning, what goes for one will hold true for the other.

Much work has been done in recent years calling this strategy into question[101]. For the purposes of the current discussion, though, we'll assume that it's sound.

Now as Schroeder has not provided us with an example of an agent-neutral practical reason, we'll look at the example he does give to see if that works out. Mary, recall, wishes to buy shoes. The more correct information she has salient to buying shoes, the greater her chances of successfully buying shoes. So, Mary's desire to buy shoes gives her a reason to only believe a statement salient to her buying shoes if it is true. But Schroeder goes further. Mary's beliefs form a network. A false belief in some statement totally removed from the purchasing of shoes may cause her to fail in some other part of her life that will impede her success at a later time, to buy shoes. For example, she believes a job interview is on a different day to when it actually is. This leads to her having less funds to buy shoes when the desire for them comes upon her.

Schroeder's position is that the holding by Mary of *any* false belief, then, has the potential to marginally undermine her success at buying shoes and thus, she has a reason not to believe it. This means Mary's desire for shoes means she has a reason to only believe things when they're true. However, this is surely too much of a leap. I'll acknowledge, by and large correct information does lead to a higher success rate at achieving our goals, but can we rule out the possibility that a false or, at the very least, a distorted view of reality might occasionally serve the buying of shoes? What if Mary perennially believed she had slightly longer left on her lunch break than she actually did? This might mean she spends slightly longer each day perusing shoes. What if she greatly over estimated how much people notice the shoes that she's wearing, or how good they will make her feel when she buys them. Is Schroeder justified in asserting *a priori* that no false belief can consistently, regularly or ever serve to increase the success rate of attaining something an agent desires?

I must confess, I can't think of any reason that even comes close to being served by *any* possible desire. Even acknowledging the possibility of the kind of agent-neutral reason Schroeder envisions, such a reason would be a little queerer that the kind we're usually willing to acknowledge and would take considerable argument to establish. I would say under the circumstances that the burden of proof for such reasons rested

[101] Christopher Cowie, *Why Companions in Guilt Arguments Won't Work*, The Philosophical Quarterly Vol. 64, No. 256 (2014).

squarely on Schroeder. Until he provides one, I feel justified in being highly skeptical concerning their existence.

Furthermore, granting that such reasons do exist – and may even turn out to be plentiful – what argument does Schroeder offer, or even hint at, that agent-neutral reasons might not directly contradict one another? The same agent after all can have contradicting agent-relative reasons. Why couldn't we all end up having contradicting agent-neutral reasons? In the context of playing chess there are oftentimes conflicts when no move is necessarily superior to another. The context alone can't resolve such conflicts by itself. Why should we believe this is not a possibility for agent-neutral reasons alike?

In Schroeder's defense, I could imagine him responding that possible contradictions, where deliberation leads unavoidably to an impasse, are a possibility in a large number of ethical theories, and may be resolved in other ways. Perhaps there could be meta-agent-neutral reasons that govern what reason an agent has to decide between conflicting first-order agent-neutral reasons.

My point is, I believe it should at least give us pause that the way Schroeder's agent-neutral reasons, as a means for deliberating action, are grounded does not preclude from the outset that there could end up being weighty agent-neutral reasons simultaneously both to φ and to not-φ.

4.3.ii Why Should Moral Reasons be Agent-Neutral?

Slaves of the Passions is first and foremost a book about reasons for action. It is not primarily a work on moral theory. However, it is clear from the tone of certain passages and indeed, whole chapters within the book that Schroeder's insights on reasons for action are meant to have applicability to our thinking on moral reasons – especially the ninth chapter, *Motivation, Knowledge & Virtue*. Specifically, that there is room enough within a Humean theory of reasons to accommodate many of the characteristics intuitively thought to be integral to moral reasons.

In the previous subsection I said that the book does not contain a single thoroughgoing example of an agent-neutral practical reason for action. However, let us assume that this is not the case and we can be confident that agent-neutral practical reasons shall be forthcoming and copious. What reason is there to believe that most, or *any* moral reasons will turn out to be agent-neutral in Schroeder's sense?

As has been re-stated numerous times before in this thesis; there are certain types of actions that any moral theory is going to have to accommodate. Any moral theory that does not decry wanton murder, rape, torture and theft at least in most circumstances will not meet with wide acceptance – nor, in my opinion, should it. However, there seems to be no semblance of an argument in Schroeder as to why this accommodation should be expected in his theory. He seems only to suggest that it is left open as a possibility.

Well, it may well be – I'm not ruling it out. On the other hand however, it leaves at least two other possibilities wide open also. Firstly, that at least one of the reasons not to murder, torture, etc. will not turn out to be agent-neutral. In which case, what shall we conclude? That there is in fact no weighty reason for all agents to refrain from this activity or that there is something profoundly wrong with Schroeder's account of reasons?

Secondly, and I think more seriously, what reason is there to believe that reasons for wanton immorality will not turn out to serve any desire an agent might have. As Nicholas Shakel points out in his Still Weighting for a Plausible Humean Theory of Reasons,

> 'It seems that bad reasons are not all the same, some are worse than others, which is to say that even bad reasons have weight.'[102]

Might we not see a path clear to saying that any desire gives any agent a reason to preserve their own existence at any cost – and a weighty one at that? The content of any desire is after all entirely moot if the agent ceases to exist before their desire is realized. Self-preservation at *any* cost could be the justification for any number of heinous acts.

Again, *no* argument is offered as to why we should even be hopeful that the road to establishing agent-neutral reasons and the road to establishing moral reasons will lead to the same destination, intersect or even run parallel.

---

[102] Nicholas Shakel, *Still Weighting for a Plausible Humean Theory of Reasons*, Philosophical Studies, Vol. 167, No. 3 (February 2014), p.13.

4.3.iii Worth its Weight?

Schroeder's alternative account of reason weight is perhaps the aspect of his writing that has attracted the most attention and criticism. By rejecting *Proportionalism* and *No Background Conditions* he has essentially confined the role of desires to being a simple straightforward, binary reason-maker. An agent either has a desire or they don't. If they have a desire, they have a reason to do anything that furthers the attainment of the object of that desire.

This is an incredibly clever move as it liberates desires from ever having the burden of explaining why agents have weighty reasons to do the morally right thing when they have little or no desire to do so, and have weighty reasons not to do the morally wrong thing even when they have a strong desire to do that. However, in order to successfully coax us away from the prevalent and highly intuitive contrary view, Schroeder's positive account of where the weight of reason derives has to be a compelling one.

Recall, for Schroeder the weight a reason derives entirely from the appropriateness or correctness of our leaning on it, and the standard of correctness is fixed by the context. A universal/inescapable standard of correctness is provided by the nature of sound deliberation itself – which means that they are equally weighty for all agents.

So Schroeder clearly has in mind an idea of 'pure' deliberation, one divorced from all particular contexts. Once again, an account of such a pure framework of deliberation is conspicuous by its absence. However, let's assume that such a standard exists. My first qualm is that it is not immediately apparent how it is possible to distinguish between acting in contravention of a moral agent-neutral reason and any other type of agent-neutral reason. For example, by Schroeder's lights I have as weighty a reason to observe the law of the excluded middle as I do not to engage in acts of genocide. Surely there is a quality of some kind in the censure, call it blame-worthiness, we attribute to those who do not act in accordance with their moral reasons that does not seem to fit with such a failure to adhere to our non-moral agent-neutral reasons. At the very least Schroeder owes us some explanation as to how his model accommodates or explains away this intuition.

Leaving this aside though, in his review of *Slaves of the Passions*, Jimmy Lenman writes,

> 'If Ryan's dislike for Katie is to be consigned to weightless oblivion for being too idiosyncratic, what is to save Ron's no less idiosyncratic fondness for dancing (or facts that speak to it) from a similar fate?'[103]

What Lenman is trying to highlight is the apparent duplicity of the role desires take in Hypotheticalism. On the one hand, Schroeder sets out to provide an explanation of how the self-same fact – that there will be dancing at a party – can be a reason for one agent to go to the party and for a different agent, a reason not to go to the party. His explanation is that what explains the difference in reason is that the former desires to dance, which will be furthered by going to the party, and the latter has the desire not to dance.

As Lenman points out though, Schroeder's account of weight seems to endanger, or at least threatens to infringe on, the legitimacy of agent-relative reasons of this kind. The supposed agent-neutral reasons have the power to outweigh or 'trump' *any* agent-relative reason – as they do in the case of Ryan and Katie. But surely in that case, there is the distinct possibility that *all* agent-relative reasons will be perpetually being outweighed by some agent-neutral reason. Hypotheticalism does not leave enough room for the role of agent-relative reasons in deliberation, he set out to establish for them from the first paragraph of the book. Schroeder does seem to be trying to have his cake and eat it too!

However, this relegation of desires in Schroeder's theory is part of a larger problem. My main criticism is with the spectacularly reduced role that desires play in Schroeder's account of reasons generally, when compared to its Humean counterparts. I think David Enoch sums this up nicely in his 'Critical Notice' of *Slave of the Passions*.

> '[T]he role of desires on Hypotheticalism is so unbelievably restricted that it becomes hard to see Hypotheticalism as an heir to the Humeans throne.'[104]

Reading Schroeder, there seems to be two distinct layers to his view. On the one hand, there is the question of how agents come to have reasons in the first place; and on

---

[103] Jimmy Lenman, review of *The Slaves of the Passions*, Notre Dame Philosophical Review.
[104] David Enoch, *On Mark Schroeder's Hypotheticalism: A Critical Notice of Slaves of the Passions*, Philosophical Review, Vol. 120, No. 3, 2011.

the other, there is the primacy of a reason, taken in-and-of-itself, and considered independently of any specific agent, in the nexus of valid deliberation. The former is what makes Hypotheticalism Humean – since it is the existence of desires that make reason attributions true or false. The latter on the other hand is more in line with some constitutivist claims about the objective standards of practical reason.

This points I have taken from Lenman and Enoch lead me to draw my own conclusion that if Schroeder were to come to believe that some psychological state other than desires could do the job of motivating an agent, and thus explaining reasons, then his account of weight could well stand unaffected. Arguably, even if Schroeder were to cease to be an internalist altogether, and was happy to accept that *sui generis* normative reasons were simply a part of the make-up of the world, his account of how these reasons have weight might, *mutatis mutandis*, work just as well.

I find it highly indicative that this vitally important characteristic of moral reasons – i.e. their non-trivial weightiness or strength – should be so wholly and *readily* divorceable from that aspect of the theory that makes it internalist in the first place. To put it another way, Schroeder's need to provide sufficiently agent-neutral reasons leads him to stretch the definition of Humeanism so far that it loses its founding justification for existing in the first place.[105]

4.3.iv Right Grounding, Wrong Place

Though not perfectly analogous, I believe Schroeder can be seen as explicitly addressing the issues closely related to what I refer to as the right grounding stipulation, with his response to what he calls *The Wrong Place Objection*.

Recall the example Schroeder uses of Ryan and Katie. Katie is in need of help. Intuitively we believe this gives Ryan a reason to help her, *simpliciter*. Now even when Ryan happens to want to help Katie, and so he has a desire-based reason to help her, it is objected by the critic of HTR that it makes this reason contingent on Ryan's desire. Yet we tend to believe that Ryan would still have just as much a reason to help Katie even if he had no such desire. HTR then supposedly puts the reason in 'the wrong place'[106].

[105] Perhaps a better name for Schroeder's version of Humeanism might be 'Homeopotheticalism'! For in the process of formulation the desires get watered-down so much as to be practically non-existent by the time we reach the ultimate solution!
[106] Mark Schroeder, *Slaves of the Passions*, Oxford University Press (2013), 137-140.

Schroeder believes that Hypotheticalism, by rejecting *No Background Conditions*, avoids this problem. According to him, the reason to help Katie is that she needs help – plain and simple[107]. What makes it a reason *for* Ryan is that he desires to help her. Ryan's desire is not part of the reason. His desire to help Katie plays an exclusively explanative role here, *not* a justificatory one. What makes helping Katie a reason for Ryan is something that could just as easily be done by any other of Ryan's desires, whose objects would be promoted by helping Katie. Schroeder believes that because desires are not part of reasons themselves but merely provide background conditions for them, the danger of placing desires in the wrong place in the constitution of a reason is mooted.

This is where the disanalogy between right grounding and *The Wrong Place Objection* becomes important. For the sake of argument, let's say that Schroeder has answered *The Wrong Place Objection*. There is still a very important sense in which he has misunderstood the concern of the critic of Humeanism, which gives rise to it.

Whether or not desires actually constitute reasons is not the issue. It is a question of the overall account as to how a reason gains its normative authority. For Schroeder a reason gets its normative force *for* an agent by being a means by which that agent can promote one of the objects of their desires. However, as I have stated earlier in this chapter, this does really leave it to luck or good fortune if it does actually turn out, after the extrapolation of any agent-neutral reasons (if any) there really are, that there are agent-neutral reasons that coincide with morality. What we need is a moral theory where moral reasons can't help but possess normative authority by virtue of some facet of their own constitution.

Hypotheticalism gets the direction of fit wrong. In the Ryan-Katie example, it is something about Katie's need for help, in-and-of-itself, which should account for why an agent has normative reasons to promote her welfare. Now it might be countered here that this isn't necessarily a strike against Hypotheticalism. It could be said that the fact that it is in Katie's interest is why any set of desires would make this a reason to help her. Yet the problem is that there is no provision that it is anything to do with Katie's welfare that makes it a reason. A desire for something much worse to happen to Katie in the future (i.e. setting her up for a worse fate) could serve as a reason for some agent to help her in the present. By Schroeder's lights, even the most ignoble desire could

[107] Ibid, p.103.

ground a reason to help Katie – and they would all count equally well as moral reasons under Hypotheticalism. It should not be that a massive overdetermination of desires happens to intersect in such a way as to make a certain reason have normative authority for any given agent. With the direction of fit the wrong way round, it is always a conceptual possibility that an agent's desires may not align with what we intuitively hold they have moral reasons to do. However, if a reason somehow plays a constitutive role in the constitution of its own normative authority this can't happen even in principle. It would also guarantee the distinctive *moral* character of the reason, rather than simply being a reason to behave morally.

Again his harkens back to the point Prichard made, that I discussed in 1.4. Though Schroeder is not insisting that Ryan helping Katie be to his advantage, he is allowing that the normative authority of his reason (a reason he is happy to count as 'moral') to help Katie, might be vested entirely on *any* desire Ryan might have to do so. Whilst I've no objection to any reason Ryan has to help Katie by virtue of any of Ryan's desires being said to have normative authority, for a reason to have moral normative authority it must be at least constituted by a desire to help Katie for the sake of the goodness of the action or the state of affairs it creates. To reiterate, even though the failure of Hypotheticalism to align with our intuitions regarding what reasons count as genuinely moral or not doesn't undermine it as a theory of practical reason, the fact that even the most base desires stand equal to Schroeder in providing reasons *for* agents to act morally makes it an unlikely prospect for accounting for the differences in our reactive attitudes toward the motivations agents have to act in accordance with their reasons to behave morally.

It has been suggested to me that this line of argument implies that no non-Rossian (or anti-theory) moral theory could be correct. Not so! The above observation only reflects that the role of any good moral theory is ultimately to tell us what reasons there (if any) for agents to act morally, *qua* morally. It is not, on the other hand, the role of a good moral theory to explain what reasons we have to behave morally that make no specific accommodation of the fact that there are things we are desirous that our theory show we have reason to do. To put it more crudely, an acceptable moral theory needs to show that there are reasons to behave morally; not there are reasons to act and by good fortune, some of them happen conform to the things we already consider moral. This is how Schroeder's model gets the direction of fit wrong.

For Schroeder, what makes helping Katie a reason for Ryan is some desire of Ryan's[108]. The fact that it is not part of the reason itself does not remove the problem. Even if we grant that Schroeder has succeeded in establishing that Ryan does have a weighty, agent-neutral, quasi-categorical reason to help Katie; I still maintain that it would not be a moral reason. This is because he is trying to shift the burden of grounding normativity away from reasons *simpliciter* and onto facts about how reasons and desires interact – i.e. onto reasons-for. However, I say that when we pre-theoretically pick-out things like helping Katie as things all agents have reason to do, it is exclusively facts about Katie and her needs that enable us to do this. Facts about the Ryans of the world and their desires do not figure into this identification.

It will always get the direction of fit wrong then, to ground the normative account of our moral reason to help Katie on desire-reason pairs. This is true whether the desire is a part of what makes something a reason or simply explains what makes something a reason. Either way, it makes desire an indispensible part of creating the normative authority of a moral reason.

With this in mind, properly viewed, the *Wrong Place Objection* should be rather that Hypotheticalism *fails to place* the reason (or some constituent part of the reason) properly within the account of what makes something a reason for an agent. This is the essence of the right grounding of moral reasons. There must be something about the grounding of a moral reason that means it makes both a necessary and sufficient contribution towards its own normative authority. The desire of any given agent is not fit to fulfill this function for the reason just iterated.

This is a general problem for Humeanism, which Schroeder erroneously believes his version sidesteps.

4.4 'In Closing'

So how well has does Hypotheticalism fare at giving us moral reasons?

First, categoricity. When it comes to this, Schroeder's version of Humeanism definitely has the potential to satisfy, at least to my standards, the categoricity requirement. The problem is that it depends on a tenuous and inadequately supported chain of 'ifs'. The first 'if' is whether there really are any reasons that are so massively overdetermined that *any* desire of *any* agent would promote them. The second 'if' is,

[108] Ibid, p.103.

even if there were such reasons, they would coincide for the most part with those things we intuitively believe we have strongest moral reason to do, and exclude those things we believe we have strongest moral reason not to do. Unfortunately Schroeder offers little in the way of argument as to why we should even be hopeful that it will provide these universal, agent-neutral reasons and so I see no reason to be impressed by this.

Secondly, there is the question of weight. Again, I am impressed by Schroeder's account of weight and agree that the weight of moral reasons should be sought, not in how much agents desire to do things but how correct they would be by an objective standard to carry them out. However, in line with Enoch's basic insight, the fact that the role of desire has had to be reduced so drastically in Schroeder's version of Humeanism, paradoxically inclines me to doubt even more strongly whether a Humean approach is one worth pursuing. Does the need to re-imagine how weight works in this way not rather imply that it would be better to abandon the Humean approach altogether, once and for all? To my knowledge, this is an original point of my own.

Thirdly, could Schroeder's agent-neutral reasons, even in principle be moral reasons? Here, I can only re-iterate what I have already written. If you make desires a necessary feature of an account of normativity, the direction of fit will always be wrong. It is the desire that will somehow make something a reason. To be the right grounding, it will always somehow have to be a conceptual possibility that a person's desires do not fit the moral reasons they actually have.

To summarize then; on categoricity and weight, Schroeder has failed to argue convincingly that he can meet the requirements. Although he has not ruled it out. On the right grounding stipulation however, he has positively failed.

Chapter Five

Korsgaard

5.1 Introduction

I now turn to the last individual that I will focus on in this thesis – Christine Korsgaard. Working along strongly Kantian lines, she attempts to elucidate how both morality and normativity itself are grounded in the constituent features of human agency.

There are certain facts that are constitutive of agency. In other words, there are certain things that must be true of something for it to count as an agent in the first place. Thus, in as far as an agent is an agent there are certain constituent facets of their own constitution that they can't avoid being. These facets are not only responsible for the existence of normativity, but furthermore, they necessarily imply that there are certain reasons for action that we cannot fail but to have. Among these reasons are reasons to be moral.

As I have written earlier, if successful Korsgaard's constitutivist approach seems to me to be the best place from the outset – i.e. it is the internalist philosophy with the best chance of success, going in. The reasons generated will be sufficiently categorical, as they do not depend on the idiosyncrasies of any individual's psychological makeup and can't be avoided. Due to their foundational role in the hierarchy of values, they will have the right order of weightiness. And given her rejection of the 'privacy' of reasons it will open up the potential to for genuinely selfless reasons – which would serve very nicely as the right grounding of moral reasons.

It looks too good to be true… but is it?

5.2 Obligation, Reflection & Practical Identity

In *Skepticism About Practical Reason* (1986), Korsgaard concurs fundamentally with Williams' basic point that in order for something to be a reason for an agent to act it *must* by necessity be capable of motivating the agent to act. She refers to this as 'the Internalism requirement'[109]. Where she differs from Williams is what limits this actually places on pure practical reason to furnish agents with novel reasons to act – i.e. its

[109] Christine Korsgaard, *Skepticism About Practical Reason,* reprinted in *Foundations of Ethics*, Edited by Russ Shafer-Landau & Terence Cuneo (2007), p303.

capacity to amend or add to the elements of an agent's subjective motivational set – which Williams himself had conceded as a possibility.

If we begin in the Humean vein, with the premise that practical reason is strictly instrumental, an agent may hate going to the gym on a regular basis – attending the gym is not part of their SMS. However, what is part of it is a strong desire for long-life, greater energy levels and the avoidance of heart disease. Reasoning soundly that the most viable course of action to secure these things is regular gym exercise, attending a gym could become an element of their SMS after all, and hence that they have reason for doing it. So, by Williams' lights what might end up in an agent's SMS is limited by what reasons can be extrapolated from the original elements of the agent's idiosyncratic SMS, by means of sound deliberation and no false beliefs. This leaves us with the perennial problem of how an agent might be said in any sense to be 'wrong' when, even with sound deliberation and no false beliefs, a reason to behave morally does not manifest itself in their SMS.

Korsgaard however, disagrees with this skepticism concerning the limits of practical reason to give rise to reasons[110]. Instead she asserts that the scope of practical reason to generate reasons is far greater than conceived of on a purely Humean terms. According to Korsgaard practical reason itself give rise to, and in fact *requires* that certain reasons be part of any soundly deliberating rational agent's SMS – which includes reasons to be moral[111].

Her starting point is an assessment of the supposed occurrence of obligation in life. Korsgaard asks under what conditions an agent might legitimately be said to be obligated to do anything. This is the question of the root of normative authority. Her answer is, when an authority commands it of us[112]. This she refers to as 'Voluntarism'. Such authority can have many different potential sources; whether it be a manager, superior officer, sovereign or deity. However, for Korsgaard, none of the commands of any of these external sources of authority could provide genuinely normative reasons. This is because, for any of them an agent could reasonably enquire why they are obligated to comply with the command of the given authority. There wouldn't seem to be any irrationality or internal contradiction involved in failing to comply with the command of an outside authority.

---

[110] Ibid, p.305.
[111] Ibid, p.306.
[112] Christine Korsgaard, *The Sources of Normativity*, p.7-10 & 21-28.

Two small digressions are essential before we go further. Firstly, what I mean by irrationality. If an agent sincerely wished to attain *x* above all other things, yet consistently and *knowingly* acted in a way that would prevent them from ever attaining *x*, we could say that this person was acting irrationally. Likewise, if a person wished to make accurate calculations but refused to adhere to the basic principles of mathematics, then they are not behaving rationally. Irrationality, in the broad sense I intend to use it, means the willful failure to employ principles or methods, either intellectual or practical, known to be necessary to successfully undertake an activity or achieve some goal.

Secondly, let us define what is meant by an agent. Korsgaard states that agency is what makes the difference between the *act* of moving your arm and an otherwise identical random spasm of the arm. The difference is that one was chosen and the other wasn't. Agency is the process of deliberating between the collection of different actions available to that agent (which includes inaction too), selecting one and willfully initiating the action. The actions of agents are understood to be the result of free will[113]. They are not determined but unconscious or mechanistic processes, but voluntarily undertaken. The agent is the first cause or bedrock of action. Without the agent's conscious choice constituting the action, there would be no action at all, but only behavior[114]. Agents act rationally in as far as they undertake courses of action they believe will attain the goals they wish to achieve. The goals they wish to achieve are the result of the things they value and the principles they adopt. What these values and principles are may have different sources, which will be discussed in greater detail as we proceed.

Failing to follow the command of any external authority, in-and-of-itself, does not incur any irrationality. For this reason, Korsgaard does not believe it can have true normative authority[115]. For her, to be have genuine normative authority, and hence a source of normative reasons, an authority must be internal to the agent. They must be commands or principles the agent confers on themselves. The thought being that there is something inherently irrational in an agent violating the commands that they have

[113] Ibid, p.100-102.
[114] Of course, I am not suggesting that all those who believe in free will accept this kind of requirement. That would be far too stringent. For Korsgaard it does appear to be a sufficient requirement for agent free choice that they be the bedrock of action rather than a necessary one.
[115] Ibid, p.9.

knowingly and rationally issued themselves. For this reason, the only possibility of true normative authority is for agents to be lawmakers unto themselves[116].

However, as Korsgaard is quick to point out, this does not give an agent *carte blanche* to adopt any principles whatever to utilize in their deliberations. There will be an independent standard of correctness that an agent must comply with if they can be said to be behaving rationally. To return to the previous example, whether an agent chooses to do mathematics or not is their choice; but if they do there are standards of correct calculation provided by the laws of mathematics that they must follow if they are to do mathematics at all. Korsgaard seems to believe that just as there must be a standard of correct practice for being a mathematician, there is likewise a standard of correctness to being an agent – i.e. a correct way, intrinsic to agency itself, of deliberating between and authoring actions in accordance with their principles. When it comes to the agent selecting these principles such a standard of correctness will be provided by reflective endorsement[117].

Rational agents are free – ironically, they are compelled to be so! They are not compelled to act in pursuance of any impulse that happens to seize hold of them. To be an agent at all is to have the freedom to choose which desires to pursue, if any. Agents may abdicate this freedom from time to time, or on a regular basis through use of drugs, self-deception, or simply by having a weak will. If they make this choice, as is often the case with intoxication, then the choice to become intoxicated or unconscious is a free one and hence, the act of an agent. However, for any period that they are no longer in control of their actions, they are not free and hence are not agents. The key point Korsgaard makes is that, *as far as an agent is an agent* at all they are compelled to submit their decisions for reflective endorsement or rejection[118].

To decide whether or not to accede to a given desire an agent can't invoke the relative strength of that desire alone as grounds for selecting it. In a sense, if that were to happen it would take the agency out of the deliberation. The free will of the agent would be replaced by a simple assenting to the course of action they believe would bring them what they desire the most. They would no longer be making a choice; the strength of the desire would be deciding the matter for them. Hence, the agent couldn't make an error about whether they should or shouldn't act in furtherance of what they

[116] Ibid, p.100-113.
[117] Ibid, p.49-51.
[118] Ibid, p.50.

most desire. They could only be wrong about whether the course of action they undertake will actually bring them what they most desire. If this were the case, practical *ir*rationality would be impossible. Everyone would simply do, always, what they most desired[119].

The fundamental choices an agent makes between which desires they choose to seek to fulfill can't be determined by the strength of the desire, for this eschews agency from the picture. Yet it can't be arbitrary, for this also lacks agency – i.e. it would just happen rather than be chosen for a reason. Agents choose there actions in virtue of their compliance with their principles.

The process of arriving at the principles by which agents make such choices – again, as far as they are practically rational – Korsgaard refers to as 'reflective endorsement'[120]. The rational, reflective, autonomous agent steps back to scrutinize a prospective principle and whether or not it is acceptable as a principle. It is acceptable as a principle if it complies with the standard of correctness integral to sound practical reasoning, and if it does not conflict with any other principles the agent has already accepted.

For example, an agent will not adopt a principle that would necessitate ignoring principles arbitrarily. Such a principle would be self-undermining. It would mean utilizing ones agency to endorse the negation of ones agency. Any principle adopted by an agent must allow the agent to continue being an agent, by not contradicting agency itself, or other principles the agent has already endorsed and chooses to retain. Choosing mutually contradictory principles would likewise be a violation of the principles of agency – the consistency and integrity of principles being a core principle constituent of agency.

If the agent finds the principle acceptable and no contradiction or conflict, they may endorse and apply it. The final choice to do so will depend on whether the principle serves a purpose that is likewise dictated by other principles that have been reflectively endorsed. For example, the principle of going to the gym regularly complies with a more fundamental principle of staying healthy, but it might also be in conflict with the principle of staying financially solvent if gym membership is outside the agent's means. This is part of what it is to be practically rational – to apply principles in a way that is

[119] Ibid, p.94-97.
[120] Ibid, p.72.

also compliant with principles of practical rationality. In a sense, the principle of having and complying with principles is self-endorsing. There is something fundamental to 'having a principle'. It is to govern ones actions in accordance with it consistently and with integrity. After all, one can't have a principle and at the same time adopt the principle of not sticking principles! There is a basic standard of consistency involved in the interrelation of principles that being principled itself necessarily entails. It is this consistent interrelation of principles that is constitutive of agency[121].

But what objective standard is there, by which an agent can scrutinize and assess would-be principles in this way? Korsgaard's answer; the standard is provided by our 'practical identities'. Throughout our lives we act in many different capacities – we have many roles. We can have the roles of friends, employers, sports-players, team-members, professionals, etc. Korsgaard calls these roles our practical identities. Each practical identity comes with its own set of requirements[122]. For example, one can't be a Latin teacher if one can't understand a word of Latin. One can't be a member of a football team if one never turns up for games, etc. Some of these requirements might just be called rules or expectations, without which you couldn't really be said to be those things. Sometimes these requirements are described in more morally loaded terms such as duties and obligations. Part of being a judge, for example, is treating each defendant equally and applying the law fairly. Being a parent means prioritizing the needs and wellbeing of your children, oftentimes over you own interests, and nurturing their healthy development. Practical identities furnish agents with requirements that provide a standard of correctness constitutive of having that identity, and as far as they are committed to that identity, this gives them reasons to behave in certain ways and not in others. When asked, 'What reason do you have to show up for the game on time, in proper kit and on good form?', an agent might respond, 'because I'm a member of the team'.

Playing a game of chess furnishes a player with a reason, at least an institutional one, to follow the rules of chess. Thus if one is truly committed to playing chess, one must accept that one has those reasons. That's what it is to be playing chess. In the same way, having a certain practical identity brings with it a set of reasons to act a certain way, for that's what it is to be those things. In as much as one is committed to our

[121] Ibid, p.101.
[122] Ibid, p.102-103.

practical identities, one must accept one has the reasons for acting a certain way that goes with them. Any agent may take up or develop a new practical identity at any time, or set aside one they already have. All accept one practical identity, that is! Korsgaard maintains that there is a foundational practical identity that an agent has by necessity of *being* an agent. It is fundamental and inescapable precisely because it is constitutive of agency itself. That is our identity as a rational, reflective, autonomous human being – our humanity[123].

By means of a transcendental argument, Korsgaard is attempting to provide a source of normativity, that can't itself be called into question. It is beyond question because, by grounding normativity in facts concerning agency itself it yields ultimate normative authority. It makes no more sense why you should act in accordance with the principles constituent of agency itself than it does to try and jump on your own shadow! To seek justificatory reasons for agency *is* to invoke agency – agency to Korsgaard, is the process of deliberating between actions based on adopted principles, remember. To question agency is like asking what reason do I have to do what I have reason to do. There is a limit to how far the questioning of ones grounds for the justification of action can regress. Korsgaard thinks this bedrock is provided by the facts constitutive of agency itself and is therefore ultimate and incorrigible.

According to Korsgaard, the valuing of things takes place within the context of some practical identity or other. The things that we value provide us with normative reasons to act within the context of the practical identity. All other practical identities are assumed contingently and might have been taken up. However, our practical identity as a human being – specifically, as a rational being that can question and scrutinize the validity of their own reasons for acting – may not be called into question[124].

A little terminological clarification might be good here. Korsgaard uses the term 'humanity', but for humanity I think we can read 'agency' – for I don't believe Korsgaard's arguments were ever meant to limited to *Homo sapiens*. Human beings are capable of acting rationally in the sense outlined earlier in this chapter. For her, agency *is* the rational application of principles. Therefore, a human being is an agent in as much

[123] Ibid, p.121.
[124] Ibid, p.121.

as they act in accordance with their rational faculties[125]. Submitting our principles to reflective scrutiny is for a human being to invoke their rational nature and thus, to act as an agent.

For Korsgaard then, our humanity is the bedrock. It ends the regress of seeking a foundation for value, not by fiat as she claims realists must ultimately do, but by providing something of unconditional value – i.e. that which makes valuing possible at all. Our fundamental nature as rational human agents is one of value-bestowers. We confer value onto things in the world, which implies that we ourselves have value as the source of it. This, she argues is the only way one might answer the moral skeptic.

To put it in Williams' terms; so long as there is anything at all within our subjective motivational set, the value of the set itself is entailed – or perhaps, to state it better, the possessor of the set. The existence of the set itself is a necessary prerequisite for any of the items within it being valuable. Therefore, according to Korsgaard, the subjective motivational set must be valuable too, as it is what makes valuing possible in the first place.

However, Korsgaard insists that this is not the end of the story. We have not quite reached morality. The process of rational deliberation leading from our individual acts of valuing does not lead merely to valuing our own humanity, but to humanity itself – to rational nature *per se*. It is not so much that the value of *my* humanity implies the value of humanity itself. Rather it is the value of humanity itself that makes my individual humanity valuable, as it is at once part of and one with humanity itself. To value my own humanity is to value humanity, and thus, humanity wherever it may occur – as it is instantiated in all rational beings.

To summarize Korsgaard's position then,

1) I am able to make free choices between possible courses of action and in accordance with principles.
2) As far as I am rational, I make choices based on principles I give myself in accordance with the things I deem valuable, once they have survived a process of reflective scrutiny and been endorsed.
3) My principles and what I deem valuable is determined by my practical identities, which are contingent and idiosyncratic to individuals.

[125] Ibid, p.122.

4) All practical identities are grounded by the practical identity of humanity, as the one inescapable identity that is constitutive of agency itself.
5) My valuing of anything implies the value of my own humanity as that which makes value possible.
6) The value of my own humanity implies the value of humanity *per se*, which commits me to value all human beings or rational agents.
7) The value of humanity and the reasons it must by necessity furnish all rational agents, is the foundation of morality and our reasons to behave morally.

If Korsgaard's argument works, as far as I can see it would tick all boxes in terms of providing genuinely moral reasons. The reasons to behave morally would be sufficiently categorical as they are inescapable; apply independently of the individual psychological make-up of any given agent; and indeed, are guaranteed by our very capacity to question whether we have categorical reasons in the first place.

They would very neatly meet the weightiness requirement since our reason to do *anything* is grounded by the value of humanity itself. Hence, the law of humanity, the moral law is the first, highest and weightiest of all reasons.

Thirdly, and most impressively of all, they would meet the right grounding requirement perfectly. Our reasons to behave morally would be neither derivative nor instrumental. They are grounded directly by the value of other peoples' humanity. Our reasons to behave with regard to other peoples' needs and value are *one and the same* as our reasons to act out of respect for our own value. This is the essence of what the right grounding requirement strives for.

5.3 Skepticism About Korsgaard

To a large degree the medals have already been awarded, so-to-speak, when it comes to criticism of Korsgaard's constitutivism – particularly that she presents in *Sources of Normativity*. For this reason, this chapter does rely more heavily on the arguments of others that have gone before, as I find that I am to alarge extent unable to improve on them, and indeed, see little reason to as they are adequate to establish the inadequacy of Korsgaard's theory. However, how these foregoing arguments may be applied specifically to establishing whether the reasons provided by Korsgaard could meet my own three criteria is original work.

My own analysis of her theory will take two parts. Firstly, I will go over the reasons why we have so many grounds to call her conclusions into question. Secondly, I will also argue, chiefly utilizing the arguments of David Enoch, that even if we were to grant that her arguments had been successful there is still reason to hold our reasons for valuing either our own humanity or the humanity of others is not sufficiently categorical to meet our requirements as stipulated.

5.3.i '… A Chain of Non Sequiturs'

I'm not sure I can improve on Rae Langton's neat summation of Korsgaard's presentation of the Kantian position; 'An unsympathetic reader may be tempted to view it as a chain of non sequiturs'[126].

I'll begin with the simple fact that Korsgaard's account of the nature of agential valuing, certainly to my mind, takes a form that blatantly flies in the face of the phenomenology of valuing. As Langton points out[127], she blanket-states, with little to no justification, that simple reflection reveals to each of us that when we consider the things that we value we will see that they are not valuable in themselves, but valuable merely because we choose them. We are the soul source of value, according to Korsgaard. Yet this is a cavalier generalization of what it is like to 'encounter' value in the world.

As Bukoski observes,

> 'The order of explanation could go the other way: our inclinations and choices could track what we regard as good.'[128]

I acknowledge that there are many things we consider valuable only in as far as they are things we like and select for how their properties appeal to us, like certain foods or films. However, there is no shortage of other things where we experience no such conference of value, even on reflection – Beauty, art or scientific knowledge for example. Here we feel that the qualities of the things themselves call on us to value them and that they would be just as valuable even if we or others did not value them.

---

[126] Rae Langton, *Objective and Unconditioned Value*, printed in Philosophical Review 116, 2007, p.169.
[127] Ibid, p.169.
[128] Michael Bukoski, *Korsgaard's Arguments for the Value of Humanity*, printed in Philosophical Review, Vol. 127, No. 2, 2018, p.206.

Perhaps this sense is illusory, but since Korsgaard has offered no compelling argument to back-up her account of valuing, I see no reason to consider it adequate in the face of the phenomenological counter-evidence.

Furthermore, I would argue that even if we granted that all things have value only in as far as we confer value upon them; this by no means secures the value of the things we choose, albeit on a new foundation of our own humanity. Rather it is just as likely to imply the invalidity of the process of valuing itself. Speaking from my own angsty mid-teens and I'm willing to bet, at least a few other peoples, on the occasions I came to suspect that their was nothing of innate or absolute value, meaning or worth 'out-there' in the world that I could anchor myself to, I came to question the whole enterprise of seeking value and meaning. It did not fill me with an overriding sense of my own potency to give the universe palpable meaning. In other words, if all things have value only in as far has we happen to value them, then this surely implies value-nihilism rather than the unconditional value of the agent that values.

Setting this concern to one side though, Korsgaard gives no compelling reason to believe that because we confer value on things – assuming that this does in fact make them valuable – that this underwrites or implies that we ourselves are valuable. As Julia Markovits points out[129], the successful conference of value from valuer to valued does not necessarily imply that the valuer is itself of value. The occurrence of infection makes penicillin valuable. Yet this does not make infection itself valuable. Markovits concludes that Korsgaard has got the direction of fit wrong. It is not our capacity to confer value from which we can derive our own fundamental value; we must begin with the fact that we are of value in order for the things we choose to have value in the first place, if the Kantian project is to work. However, this would take a significantly different argument to establish than the one Korsgaard offers.

However, let us assume for the moment that Korsgaard has got us as far as securing the value of our humanity. My own argument is that there is still the step, pivotal to establishing morality, from the value we each attach to our own humanity to the value we must attach to humanity *per se*, if we are to provide reasons for respecting other people as much as we do ourselves.

[129] Julia Markovits, *Moral Reason*, Oxford University Press (2014), p.105.

Unfortunately for Korsgaard, her case for this aspect of her theory rests on a highly dubious Wittgensteinian line of argument[130] – one that has come to be treated somewhat dismissively even by many of her supporters, as one of the weakest aspects of her position. In much the same way that Wittgenstein questions the coherence of the idea that a language could be private – i.e. that the meaning of a word could be fixed by a referent known *only* to the speaker – Korsgaard rejects the idea that reasons are in any sense private. Reasoning can be performed publicly, according to Korsgaard, which is how joint or collective decisions can be reached. An agent has no right to claim a reason for action as being strictly their own. Since any agent, by virtue of being committed to valuing their own humanity, has a practical reason to act in accordance with the value of their own humanity, that reason applies across all rational beings for the same reason.

Once again though, Korsgaard offers little argument for how this shared value sidesteps the motivational constraint that Internalism imposes. I believe the main crux of her argument to be that valuing anything ultimately commits us to valuing humanity *per se*, and thus to the ends of others that do not themselves conflict with the value of humanity. This however, is not the same as being motivated by the reasons other people have – merely being motivated by a reason to serve other peoples ends.

Finally, in terms of my general problems with Korsgaard's project; even granting that all reasons derive from our practical identities, there is still the question over whether *all* our practical identities actually stand in need of being provided with the kinds of foundations Korsgaard seems to think that they do. Further to this, even if they do need such foundations, it is not at all clear that this can only be provided by humanity. Here to the end of this subsection, I turn exclusively to my own criticisms of Korsgaard's account.

Epistemologists are familiar with the holistic approach to webs of belief. Each item in our web is supported by another item, which itself is supported by another to form a network of internal justification. Why should our practical identities not hang together in this interconnected way also? My practical identity as a teacher overlaps and shares values with my practical identity of being an uncle and friend. They bolster and support one another. My identity is more than just the totality of my practical identities. It is also constituted out of the coherence of all the identities that harmonize with each other to make what I consider to be me. Perhaps this is just what humanity

[130] Korsgaard, *The Sources of Normativity*, p.137-139.

really consists in – being an individual who has successfully integrated all their separate identities into a unified sense of self. Like Neurath's boat, there is no need for humanity itself to somehow serve as a foundational support, standing underneath our identities.

It is also a well-known principle that simple or lower-level phenomena can give rise to the emergence of higher order phenomena. One might argue that the purpose of simple neural activities like synapses firing in the brain – vital for the possibility of consciousness – is to give rise to that consciousness. The latter makes the former valuable, despite the supervening dependency operating in the other direction.

Korsgaard offers us no argument for why the value-systems of different practical identities, which are made intelligible by the value of humanity itself could not give rise to new systems of valuing that transcend and even surpass the value of humanity itself. Perhaps Shakespeare's or Mozart's value as great artists, or Einstein's and Newton's as scientific geniuses, gave them a legitimately higher set of ideals and values than those that could have been derived from their mere humanity alone. If this were so, then some practical identities may be able to give rise to novel, non-derivative values that contradict, at least to some degree, those dictated by our shared humanity.

5.3.ii Agent or Shmagent?

This subsection is entirely dedicated to what I believe is the single strongest objection that can be raised against both Korsgaard and the constitutivist project in general. It is given best expression by David Enoch in his *Agency Shmagency: Why Normativity Won't Come from What Is Constitutive of Action* paper (2006) and again in *Shmagency Revisited* (2010).

Remember that Korsgaard's goal, essentially is to say, that the normative force of certain principles of practical reason is guaranteed simply by virtue of those features being essential and inescapable to agency itself. Enoch starts with the deceptively simple questions; Why should I care about the way I am constituted[131]?

Even granting that there are elements of our constitution from which principles of practical reason can be derived – which both Enoch and I are prepared to do for the purposes of discussion – what difference does it make ultimately? Just because something is constitutive of agency how does this make it normatively non-arbitrary?

[131] David Enoch, *Agency, Schmagency: Why Normativity Won't Come from What is Constitutive of Agency*, Philosophical Review, Vol. 115, No. 2, 2006, p.178.

The point is this, Korsgaard is asserting that failing to adhere to the constitutivist standards of practical reason means that our 'actions' will not truly be actions and we will not truly be agents. This is analogous to saying that a chess player could not be said to truly be playing chess if their game-play is in no way directed toward luring their opponent's king into checkmate and at the same time protecting their own king from being checkmated[132].

However, Enoch's question is; what if I don't care that my 'actions' are not really actions? What if I am content to be a shmagent – i.e. a person whose conduct resembles that of an agent in every conceivable way, short of adhering the constitutivist standards of practical reason – rather than an agent? Enoch is saying that Korsgaard's account seems to be in need of providing a normatively grounded reason to be an agent in the first place. To put it another way, just because something is essential to being an agent it doesn't mean, as far as its *normative* status goes, it is anything other than arbitrary[133].

Korsgaard might be anticipated in responding to this challenge by saying that the norms of agency itself are self-validating somehow, and thus in no need of independent justification of the kind Enoch thinks necessary. Firstly, she might want to come back and say that simply by the fact that they are essential to agency itself and that which is essential to agency is precisely *what it is* to have a reason to do anything, is sufficient to establish their normative authority. Secondly, it could be asserted that on reflection, we invariably *do* care about what is constitutive of our agency and so asking why we should care is a moot question. Thirdly, to challenge the standards of practical reason is self-undermining. Just as any attack on the principles of logic (e.g. law of the excluded middle, non-contradiction, etc.) would have to employ logic; any attempt to challenge agency would require agency[134].

For Enoch though, none of these attempted defenses are in anyway satisfactory in eliminating our worries. Simply stating that constitutivist standards being essential renders them normatively non-arbitrary is not sufficient. You would need more of an argument to establish that the, at face value, legitimate question of why a person should care about being an agent, is not pertinent. Is it not conceivable that our constitution is antithetical to morality? No equivocation can be assumed between what is constitutive

[132] Ibid, p.185.
[133] Ibid, p.182.
[134] Ibid, p.184.

of agency and what has normative force. To insist from the outset that this is impossible by definition would seem to beg the question.

Regarding the statement that we do in fact invariably care about constitutive standards; first, empirically, it is not at all clear that we do actually care about what makes it possible to care about the things we do care about! Furthermore, how exactly does our desire for something establish its normative force? Can we not imagine loving someone and not caring that our love for them requires us to go on living? This would seem to be a rather straightforward example of the is/ought gap in action. The fact that we *do* care does not necessarily imply that we *should* care. The omnipresence of caring about constitutive standards does not go any way toward grounding their normativity.

Thirdly, one of the most prevalent defenses from criticism of constitutivism is that the very attempt to attack it is inconsistent. To undertake a skeptical 'attack' against constitutivism is to utilize one's agency, thus to make one's own argument untenable. Enoch asserts that in employing this defense constitutivists are falling into what Wright had earlier referred to as the 'adversarial stance'. It represents the nature of such a challenge in dialectical terms, where an actual or imagined interlocutor is offering their own position as superior or more plausible. In this way the constitutivist is basically making an *ad hominem* attack on anyone who tries to formulate a coherent critique of their thesis. Yet this is a misleading characterization of the nature of the challenge, according to Enoch. It is more that the skeptic is highlighting a problem that the constitutivist themselves must wrangle with for their own stance to be possible.

If one must personify the skeptic in the way that the adversarial stance would coax us toward doing, Enoch would say that it is flat wrong to say that they are not entitled to use the constitutivist's own weapons against them. The skeptic, after all, does not care if their position seems self-defeating from the perspective of the constitutivist; but the constitutivist *does* care if their own position is coherent. The simple fact that anyone who could potentially challenge the foundations of our agency must also act as agents, does not render those foundations normatively unquestionable in the way Korsgaard requires. I'll be returning to this issue a little later.

Let's return to the game analogy, so often used by constitutivists to bolster their claims. As I mentioned before, it seems out of place to say that someone is honestly playing chess without working toward the objectives constituent of a game of chess, even if they might be moving the pieces in accord with the rules of the game. Because of

this, moving toward the checking of your opponent has normative force as long as you can be correctly said to be playing chess.

For Enoch though, a vital part of this analogy is left out. For the constitutivist in order that the aims of chess have normative force, the player must have a reason to play chess in the first place. If you have no reason to play chess, you have no reason not to content yourself with playing shmess – an activity much like chess in the way the rules allow and disallow the movements of pieces, but where the goal is to achieve the optimally aesthetically pleasing arrangement of the chessmen on the board. Korsgaard still owes the player a reason to play chess instead of shmess.

Enoch's contention is that the constitutivist line only works if there is a foregoing, independent reason to play the game. This reason can't be garnered from the constitutive standards of agency itself. Additionally, if we need this independent reason, constitutivism can't give us the whole story of normativity. There will always be something missing from the kind of thoroughgoing constitutivism Korsgaard is trying to sell us.

Here we must be careful not to stretch the game analogy too far. For the constitutivist will say that the 'game' of agency is not like any other game. It is a game we are condemned to play[135]. It is inescapable and there in lies its *causa sui* normative force. We have no choice but to play[136] and so we stand in no need of a reason to play it.

But this appeal to unavoidability simply won't do the trick. For as Enoch points out, the kind of necessity an agent has to be an agent needs to be one of *normative* necessity. Let us assume that one is condemned to play chess in something sufficiently like the way we are condemned to be an agent – i.e. however we might conceive of it being true, it could not be otherwise than that the norms of chess *just do* apply to you. This still does nothing to establish the normative force of the goals of chess. Surely in such a situation we can imagine a player playing the game halfheartedly, without properly internalizing the aims of the game, and that there not be any irrationality involved in them doing so. Furthermore, there are many other non-optional aspects to the playing of chess, or any other goal-directed activity. Can normativity be derived from every non-optional constituent feature of an activity? If not, why only specific

[135] Ibid, p.188.
[136] As David Velleman observes, even to attempt to 'opt-out' of agency by committing suicide or wilfully rendering yourself unconscious is a move *in* the game and requires agency.

ones? Non-optionality does not in-and-of-itself entail normativity. We are still missing a vital piece of the solution.

In *Shmagency Revisited*, whilst recover some of the same ground, Enoch gives greater focus to the kinds of trends in responding to the shmagency challenge that had cropped up in the intervening literature. Many different versions of chess can be imagined, each with slightly different constitutive aims. Chess* for example is version where you must mate your opponent in an even number of moves; chess** is version where once your last pawn has been taken the king can move two spaces at once; Chess*** allows castling, queening and *en passant* but only when the total number of pieces on the board is prime, and so on…

The point is, in the early stages of the game, the majority of moves will be consistent with the constitutive aims of all four of these versions of chess. Thus the constitutive aims alone can't commit the player to whichever version of chess they are supposed to be playing. This is a question of what version the players intend to play. They can't be committed to the goals of all of them simultaneously!

One line of constitutivist defense that had emerged though, was that the shmagency challenge was based on a misunderstanding of the true source of normativity. It is not, Korsgaard might respond, though she never actually does, that normativity is entailed by the fact that agents must play the game of agency. It is that they can't help but *care* about playing the game of agency – the *intention to play* agency is part of agency. Hence, the suggestion that an agent could conceivably be aware of the standards constituent of agency, but not care about them is an untenable one for Korsgaard.

So Enoch's challenge supposedly only applies to those not *already* caring about playing chess or agency specifically. The fact that the skeptic asks why they should care about agency proves that they do! This is different to Korsgaard's assertion that if you don't care that you're not an agent then you're not an agent – yet none-the-less it is a response that a follower of hers might well deploy in their defense.

Enoch's reply to this is that it is both implausible to believe in the first place, but even if true, wholly irrelevant. First; implausibility. It seems a considerable stretch to our ordinary understanding of what constitutes legitimate examples of chess-play to say that actually *caring* about what is constitutive of it, is an essential part of doing it. Surely

we can imagine diverse players playing to an equally fine standard of Chess but with varying levels of heed or indifference as to the actual result of the match.

Also, nothing of any normative significance seems to hang on this point. Why does it matter if someone who doesn't care about the norms of chess or agency *counts* as being a chess-player or an agent? Given that the behavior of the player in question would be no different, it would appear to be a purely semantic point. The suggestion that shmagents can't be sufficiently 'like' agents if they lack the elements vital to being agents, seems too much to countenance. Even if they can't 'act' they can still 'behave', and to say behavior is nothing like action is surely absurd! Hence, without some further argument, shmagency (*qua* not caring about what is constitutive one's agency) *is* a valid option and we can't convict someone of irrationality because they are content to behave rather than act.

Then there's irrelevance. Even if it is true that we do invariably and unavoidably care about what constitutes our agency, nothing can be deduced from this regarding its normativity. That I *do* care about φ-ing says nothing as to whether or not I *should* care about φ-ing. This seems to be a rather straightforward failure to appreciate the is/ought gap.

It could be responded that you automatically have a reason to do what you care about. Yet this doesn't work either. Still more is needed. Whenever we apply reflective scrutiny we can ask of *any* desire we happen to find ourselves with if it is worth pursuing. The fact that you are bound to have certain desires does not remove the legitimacy of scrutinizing them. They remain normatively arbitrary facts about you. An Internalism that tries to solve this problem by saying that we have a reason to do *anything* we happen to care about, without any additional context or criterion of validity, would be extensionally inadequate and unworkable.

In *Shmagency Revisited* Enoch reiterates and develops the key point he made in the earlier *Agency, Shmagency*. The central strategy that followers of Korsgaard and other constitutivists are trying to employ is to make out that there is no conceivable ground for the skeptic to occupy. The skeptic who challenges constitutivist norms, or asks why they should care what they are, is not wrong but impossible.

By characterizing the disagreement in dialectical terms – in terms of the aforementioned 'adversarial stance' – even if only implicitly, the constitutivist is relying on saying that since the skeptic breaks their own rules, they somehow win by default.

They are saying that if the skeptical challenge is asked 'internally', the means for providing an answer is guaranteed because the skeptic must tacitly accept the conditions of agency that make asking the question possible. On the other hand, asking the question 'externally' is incoherent as one can't employ agency to seek a question without the tacit acceptance of the norms governing agency.

But this is not acceptable. Even if we concede that the skeptic is *in* trouble, it in no way implies the constitutivist is *out* of trouble! Given the apparent validity of the question of why anyone should care about the norms of agency (again, assuming for the sake of argument such things exist), or why we should opt for shmagency over agency; the burden of proof is surely on the constitutivist to provide a thoroughgoing treatment for why the question absolutely can't be asked externally. As Enoch concludes, no such thoroughgoing treatment has yet been presented and he is not expectant that it ever will be. Neither, for that matter, am I.

5.4 Score-Keeping

We are now in a position to assess how Korsgaard has done in meeting our three criteria for generating moral reasons. As I wrote before we started; of all internalist moral theories I felt Korsgaard's and the constitutivists in general had the greatest chance of meeting our three criteria for being a moral reason.

In principle and from the outset, she has her sights set on standards for the reasons she hopes to generate that are truly worth the name 'moral'. It is not her intension then that we need to assess, only her degree of success. The success of the three are intimately connected, but I will do my best to treat them piecemeal. My arguments for how the preceding discussion applies to this question and how I believe they are applicable to the success of providing moral reasons as I define them is distinct from the literature.

5.4.i Reason *As* Morality

First, let's look at weight. According to Korsgaard practical reason is grounded ultimately on our humanity. To have a reason to do anything is to have a reason to value our humanity and by extension, the humanity we share with all rational beings. In this way, one could say that Korsgaard's moral theory implies that practical reason itself is

fundamentally moral in character. Reasons to do anything entail reasons to be moral first and foremost.

Thus, if her form of constitutivism is successful she has demonstrated that our most fundamental, primal and strongest reasons to act at all are the moral ones. This is more than adequate to meet our criterion of weightiness.

5.4.ii Humanity *Per Se*

In the same vein as with weight, when discussing the right grounding, Korsgaard's view is that reason is fundamentally grounded on our own humanity and morality is based on our shared humanity. This sharing allows the reasons that other people have for valuing their own humanity are one-and-the-same with the reasons we each have to value our own. On Korsgaard's view, in acting out of duty to someone else I do not require that I desire their welfare in some form of instrumental sense. Instead, her constitutivism breaks down a supposedly illusory divide between what provides me with reason to value my own humanity and my reason to value others.

This would clearly be an example of the kind of transcendence of self-interested reasons that I take to be the typical hallmark of a moral reason and the right grounding. Additionally, that immoral acts are violations of other people's humanity is usually integral to what we think makes many, if not all acts of immorality, immoral, makes it a perfect candidate to ground moral reasons.

5.4.iii The E(x)ternal Question!

I have left dealing with the categoricity requirement until last as I believe my assessment of how well Korsgaard has dealt with this helps me to clarify how successful she has been overall.

The categoricity that Korsgaard hopes will emerge from establishing moral norms as being inextricably linked with the norms of practical reason itself is its greatest strength and its greatest weakness. That which makes our reasons moral, i.e. our duty to our shared humanity, is precisely what is supposed to ground their categoricity. It is our humanity that guarantees the inescapability and hence categoricity of our reasons, and this self-same character is what provides for its moral nature. All reasoning is ultimately moral reasoning as all reasoning is founded on practical identities as human beings. They stand together and fall together.

Therefore, if there is sound reason to question that the norms of practical reasoning are in fact categorical the justification for holding that the norms they recommend are ultimately moral crumbles with it. To put it another way, if normativity requires grounding that is external to the norms constitutive of agency, then the norms of agency can't give us a securely grounded normative reason that we *should* adhere to them.

5.5 On Reflection

In my opinion, if David Enoch has not completely and convincingly undermined the attempts of constitutivists like Korsgaard to show that the norms of agency itself can provide *causa sui* normative reasons to adhere to them, he has at the very least shifted the burden of proof resoundingly onto them to provide them. Until such proof is provided, I believe we are entirely justified in concluding that the question as to why we should care about the norms of agency, and any supposedly moral reasons which are entailed by them, may quite legitimately be asked 'externally'.

Though I have drawn heavily on the writing of other writers here, especially Enoch, my arguments as to how precisely these failures make the meeting of the three criteria possible are ideas of my own. The implications as to precisely why each failure makes Korsgaard's model unable to generate truly morally reasons that I am able to provide novel arguments as to why this is. My own original position in Korgaard is that they largely succeed in spirit but fail in substance. The kinds of reasons Korgaard is trying to provide are what I would be close to calling truly moral reasons as if she had succeeded, they would have meet the criteria. For this reason my position is that Korgaard's failure is only one of execution not of intention. The goal of her constitutivism is a sound one.

Korsgaard has not given us a sufficiently categorical reason to value our or humanity or anyone else's. In which case her entire project fails and no moral reasons have been provided. However, I will conclude by saying that if she had succeeded, her reasons would have been moral ones as I understand them.

Chapter Six

A Perennial Problem for Internalism

6.1 Introduction

We have now looked at three very prominent and influential internalist theories and, I hope, demonstrated how each fail to meet at least one of the necessary criteria for supplying moral reasons for action. What remains now is to explore the question as to whether these failures are specific to these theories or if they are representative of a broader general problem with Internalism itself. Is it possible that it is not only the case that certain internalists fail, but that *any* internalist moral theory is doomed to fail due to a fundamental constitutional feature of it?

The structure of this chapter will be as follows; in Sections 6.2-6.4 I will outline and expand on a mistake that every form of Internalism I have encountered makes. I refer to this mistake as *The Endemic Error*. I will argue that this endemic error is a natural consequence of Internalism's adherence to the motivational requirement, and as such leads to internalist theories, at least in practice if not strictly speaking in principle, to perennially commit it. In Section 6.5 I will make a small clarification regarding the scope of this thesis. Then in Section 6.6 I will explain how I believe *The Endemic Error* makes the prospects of any internalist theory meeting my three criteria highly dubious. Finally, in Section 6.7 I shall briefly examine the possibility that utilizing alternative normative concepts may hold the potential for Internalism to sidestep these problems and so emerge victorious. I will explain why I also find this an unlikely prospect. At which point, we will be ideally placed to draw our conclusions.

6.2 The Endemic Error

Almost every form of Internalism, certainly every that I am aware of, makes the same fundamental mistake. I refer to this mistake as *The Endemic Error*.

> *The Endemic Error*: What it is to be a normative reason is *nothing more* than that it could or would motivate an agent's actions under certain circumstances.

This section draws heavily on the insights of Derek Parfit. However, my formulation and the specic application of the Endemic Error are my own original

contributions. My reason for taking this error to be both fundamental and, in practice endemic to Internalism is simple. It is natural consequence of adhering to the motivational requirement. Recall,

> *The Motivational Requirement*: For there to be a reason that an agent φ, it must be possible that the agent could be motivated to φ.

The motivational requirement is the *sine qua non* of Internalism. By definition it is what all internalist theories have in common with each other. Furthermore, in practice the motivational requirement entails the *The Endemic Error* in as much as every actual internalist theory winds up committing it.

This section will be divided, roughly speaking, into two parts. The first will outline exactly what *The Endemic Error* is and why it tends so routinely to flow from the motivational requirement. The second will utilize insights from Derek Parfit to explain exactly why it is bad news for any theory that wishes to give an account of normative reasons for action.

So, the motivational requirement places an explicit limit on what reasons actually exist. It does not allow that there may be two classes of normative reason – i.e. things an agent should do and are capable of being motivated to do, and things and agent should do but could not be motivated to do. Instead it denies the existence and, in many cases, the very intelligibility of a reason that could not motivate agents to act under any circumstances – i.e. that there is anything an agent *should* do that they couldn't be motivated to do. From this it follows, if something is to count as a normative reason for some agent to act in the first place, it necessarily implies that there is, or could be, a motivational state that could cause or explain that action. In this way, nothing can possibly be a normative reason unless it is made one in virtue of some fact concerning the motivational states of some agent. To put it another way, the only things an agent *should* do are those things they actually *would* do, or would at least be motivated to do, under certain circumstances. In this way, the motivational requirement leaves nothing for a normative reason to be outside of its satisfaction *of* the motivational requirement. In practice this has almost universally lead to the internalist implicitly accepting that what it is to be a normative reason can be encapsulated entirely by its role in explaining agent action. This is *The Endemic Error*.

Perhaps an analogy will help to cement my point. When I consider this problem with Internalism, I am put in mind of *Mary's Room*. The reader will no doubt be aware of the thought experiment in epistemology of the super-scientist named Mary, who has learned and assimilated all of the scientific knowledge there is to know on the science of colours – i.e. photon behaviour, wave-lengths of light, neurological and psychological responses to colour, etc. However, the one thing missing is that she has been raised in a monochrome room the whole of her life and has never actually seen something that is red. The question raised is, if one day she is shown a ripe strawberry, say, does she now *know* something she didn't know before? If we answer 'yes' we acknowledge that there are things to know about colours that are outside of the empirical sciences; and, if we answer 'no' we are in need of some account of what it is that happens to Mary when she perceives the ripe strawberry, if it is not the acquisition of the knowledge of 'what red looks like'.

Similarly, the practical upshot of internalist's adherence to the motivational requirement has the implication that the quintessential nature of a normative reason is exhaustively provided by nothing over and above its functional role in explaining agent action. To put it in terms of the analogy with *Mary's Room*; if Mary knows absolutely everything there is to know about the motivational states of agents, is there anything about the normative reasons of agents that she doesn't know? If the internalist answers 'yes', it implies that there is something for a normative reason to be that does not depend on facts regarding motivational states, and hence the intelligibility of a normative reason that could not necessarily motivate an agent. On the other hand, if the internalist answers 'no', they acknowledge that there is nothing more to normative reasons aside from facts about motivational states – and therefore, they commit *The Endemic Error*. In fairness to internalists, I am not aware of any that actually offer answer to this question or one like it. However, the fact that they do not see this as an issue for them to contend with is indicative of the crucial blind-spot I see as being endemic to Internalism.

In a nutshell, I see the chief problem with normative reasons being conceived of in this way is that it get's the direction of fit wrong between the normative and the explanative. The internalist begins with the highly plausible intuition that the only reasons there could be are those that could motivate agents, and then tries to tailor a conception of normative reasons that will comply with it. Alternatively, we could say

they try to get us to accept as being adequate, a particularly skewed account of what a normative reason is, primarily because it best fits the plausible account of reasons and their relationship to motivations that they already have. Whilst I sympathize with the desire to get normative reasons to fit into this plausible account of reasons more generally, it is fundamentally wrongheaded and can't give us the kind of normative reasons that we would require to serve as moral reasons – which will be discussed in great depth in Section 6.6, below.

Before I go further, a brief note on terminology. Much of the following discussion will focus on my argument that the normative reasons that are generated by internalist theories and that internalists are comfortable calling normative reasons are not genuine normative reasons. To avoid confusion, unless otherwise stated, I shall be using the term 'normative reasons' to refer to reasons that are genuinely normative in the sense I think we intuitively think normative reasons should be. For the kinds of reasons internalist believe are adequate to be referred to as normative reasons – i.e. those that necessarily meet the motivational requirement – I shall be using the term 'internal reasons'. So, 'internal reasons' are those reason internalists believe are fit to be considered normative reasons, but which I maintain are not fit to be considered genuine normative reasons. I will not be denying that internal reasons *can* at the same time be examples of genuine normative reasons. I shall only be denying that their being normative reasons, *qua* normative reasons, is grounded by their being internal reasons – i.e. by their satisfying the motivational requirement. If they are also genuine normative reasons it has to be in light of some consideration beside simply satisfying the motivational requirement.

Internal reasons then are those that conform with and are grounded by the explanative reasons agents have for their actions by virtue of some fact or facts about their motivational states. However, it is my contention that if there were such things as normative reasons, the direction of fit would *have* to work in the opposite direction. It is the normative reason for action that must tell us how the agent should act – i.e. what motivational states it is fitting or appropriate for the agent to have. This means there would have to be more to what makes something a normative reason beside simply what it *does* or *could* cause an agent to do. *The Endemic Error* renders it impossible for Internalism to generate reasons that have this vital additional explanative power to them.

Though the above paragraph represent my own original formulation of what I believe constitutes and original problem with all forms of Internalism, I find Derek Parfit's discussion of this very issue in his *Reasons and Motivations* (1997) particularly helpful in further illustrating my main point. Here, Parfit tries to break down and examine the character of the relationship between internal reasons and motivational reasons, as the internalist (specifically Williams in this case) would characterize it. He says that the motivational requirement properly analyzed, can be interpreted in at least three different ways.

By necessity,

1) If agent *A* has a normative reason to *φ*…

   This entails,

2) If *A* deliberated in a fully procedurally rational way and had no false beliefs, they would be motivated to *φ*.

The first way the relationship between (1) & (2) can be characterized is what Parfit calls the *Analytically Reductive* way; the second he calls, *Non-analytically Reductive*; and the third, *Non-Reductive*[137].

Analytically Reductive basically amounts to saying that (1) & (2) essentially express the same piece of information just in a different way – just as 'bachelor' and 'unmarried man' picks out *exactly* the same thing via the same intension, but using a different term. The second, Non-Analytically Reductive, states that one fact has a kind of dependency on the other. This is much like the way facts regarding colours depend, to an almost exclusive extent, on facts about wave-lengths of light; where there is light of a certain wave-length there is invariably light of a certain colour, and *vice verse* – yet we do not think that the colour and the wavelength of light strictly refer to the exactly same thing, though they may well be co-extensive. The third however, Non-Reductive, is saying that (1) & (2) are causally connected in some way, but express very different kinds of facts about the world. A body having mass for example, and it distorting space-time might be said to be necessarily causally related, though 'a mass' and 'the distortion of space-time' are in no meaningful sense the same thing.

[137] Derek Parfit, *Reasons & Motivations*, Aristotelian Society Supplementary Volume, Volume 71, Issue 1, 1 July 1997, p.108.

Parfit believes that the kind of relationship between (1) & (2) that the internalist is committed to attempting by virtue of the motivational requirement, is almost invariably (and certainly is in Williams' case) some form of non-analytic reduction. An agent's normative reasons depend on what they could or would be motivated to do under the right circumstances. They essentially pick out *the same thing*, just in different ways. However, for Parfit the kind of things a normative reason has to be makes it impossible for this kind of non-analytic identification with mere facts about an agent's motivational states – i.e. the content of their psychology – to yield genuinely normative reasons.

(2) is simply a descriptive fact about the world. (1) on the other hand is essentially normative in character. It states what they *should* do. As such, for Parfit only the non-reductive kind of relationship between (1) & (2) makes any sense when you are talking about the relationship between normative and non-normative facts – between normative reasons and motivational/explanative reasons.

> '[I]t was conceptually possible that heat should turn out to be molecular kinetic energy. But heat could not have turned out to be a shade of blue, or a medieval king. In the same way, while it may not be conceptually excluded that experiences should turn out to be neurophysiological events, experiences could not turn out to be patterns of behaviour, or stones, or irrational numbers.'[138]

Parfit's point in the above section is that identity relations can only provide satisfactory, elucidatory and philosophically coherent explanations when the items being identified are in suitably similar logical categories. And furthermore, a reason that an agent should do something and a fact about what they could or would do under certain circumstances, aren't in suitably similar logical categories. For this reason, *The Endemic Error* means that Internalism is systematically incapable of furnishing us with a satisfactory analysandum of a normative reason.

Just parenthetically, Parfit also argues that even the most thoroughgoing internalist must be tacitly committed to the existence of one normative reason that can't be reduced to a fact about what we can be motivated to do. Surely, he says, there would have to be a primitive normative reason to do what we are motivated to do when fully informed and deliberating with full procedural rationality. If this were true, I would say

[138] Ibid, p.122.

that this sort of primitive reason is paradigmatic of what would have to lie outside of *Mary's Room*.

To believe that I have a reason to *φ* is to believe that I *should φ*, all else being equal – that I would somehow be guilty of an error, of practical irrationality, if I did not *φ*. But for Parfit, this is *not at all the same thing* as believing that if I did deliberate rationally I simply *would* be motivated to *φ*. All internalists implicitly trade on this kind of false suppressed conflation between what is quintessential to something being a normative reason and facts about the motivational states of agents.

To demonstrate the distinction between normative reasons and motivational states a little better, Parfit invokes Williams' famous gin & tonic example. If I am thirsty and I believe the glass in front of me to contain gin & tonic, when in fact it contains petrol, do I have a reason for drinking it? Surely whether or not I am actually motivated to drink the liquid depends entirely on what my beliefs are concerning the content of the glass. As such, my motivational state is determined by belief alone, not the fact of the matter re: the actual contents of the glass. Conversely, whether or not I actually have a normative reason to drink the contents of the glass depends only on the fact of what is actually in the glass, entirely regardless what I do or do not believe. So, in this case, what grounds my motivational state (my beliefs about what the glass contains) and what grounds my normative reasons (the fact of what the glass contains) are at variance with each other.

My own argument, making a more general point to Internalism, is that in trying to create a moral theory that conforms with some of our most plausible intuitions regarding the link between reasons for action and motivation, the internalist has been forced to sacrifice something quintessential to normative reasons. To plump for the primacy and desirability of the explanative adequacy of a theory of reasons at the expense of normative adequacy is a strategy that will never lead to anything like an account of genuinely normative reasons – only internal ones.

If there is more to being a normative reason than compliance with the motivational requirement, then it leaves open the possibility that agents might have reasons to act that can't motivate them. And this does raise the important question as to how such a thing can in any sense *be* a reason for an agent. However, it is ultimately a secondary concern when the goal is to try and give a satisfactory account of normative reasons, *qua* normative reasons. Indeed, as Parfit put it, to refer back to him for a

moment as we close this section, it is sufficient to say that an agent is practically rational in as far as they *are* motivated by the normative reasons they actually have. The possibility that some people may not be motivated by their reasons is just an ever-present possibility of life.

> 'Even if moral truths cannot affect people, they can still be truths.'[139]

6.3 A Possible Internalist Response?

This is all well and good, but we can't expect the internalist to leave it at that. If there were a way for them to establish a necessary relationship between normative reasons and facts concerning the motivational states of agents, which did not commit *The Endemic Error*, then Internalism might be salvageable.

In the previous section we looked at Parfit's analysis of the three different kinds of relationship that could be being expressed by,

By necessity,

1) If agent *A* has a normative reason to *φ*…

   This entails,

2) If *A* deliberated in a fully procedurally rational way and had no false beliefs, they would be motivated to *φ*.

We saw that Parfit took Williams' position, and by implication a substantial swathe of the internalist theorist who followed him, to be that the relationship between (1) & (2) was one of non-analytic reduction – i.e. the normative reason to *φ* and the fact that an agent could or would *φ* under the right circumstances essentially refers to *the same thing*, just in different ways. We then went over Parfit's reasons for considering this kind of reduction as inadequate to capture the essence of normative reasons. We also saw that the only kind of relationship between (1) & (2) that Parfit thought could be satisfactory would be a necessarily causal one, rather than one of identity. To use Parfit's analogy – it is not a significant goal of a mathematical proof to show that the property of being an even prime number is literally identical with the property of being the square root of four. It is only needed to show that they are, *of necessity*, instantiated by the selfsame particular – in this case, the number two.

---

[139] Ibid, p.111.

Strictly speaking then, Parfit's argument only applies to internalists who take what might called a *hard* naturalist line when it comes to normative reasons. A hard naturalist line would be one that attempts any kind of strict identity between normative facts or properties and natural facts or properties. In the case of Internalism, the posited identity is between normative reasons and facts about the motivations of agents. Therefore, if internalists were able to provide a non-reductive, necessary causal account of the relationship between (1) & (2) they would be able to sidestep *The Endemic Error* altogether. I will refer to any such non-reductive, softer naturalist attempt to do this as *soft* Internalism.

The best example I am aware of of an exploration of the viability of such a soft Internalist strategy comes in the form of a disagreement that took place between Schroeder and Parfit. Although Schroeder is of course a constructivist rather than an Internalist, and also he would not have approved of being classified a soft naturalist of the kind I have just mentioned. However, the points that get raised in the course of the discussion between the two are no less relevant to the current issue regarding Internalism and *The Endemic Error*.

In Volume 3 of *On What Matters* (2017), Parfit addresses Schroeder's answer to the former's 'triviality' objection. Parfit had argued elsewhere, in a way that has some echoes of Moore's *open question argument*, that if normative facts could be reduced to descriptive facts then there would be a kind of triviality to identity statements concerning them. This is similar to when he discussed in *Reasons & Motivations*, that the relationship between (1) & (2) would be either analytically reductive or non-analytically reductive. We can see straight away how an analytically reductive identity statement, like 'bachelor = an unmarried man', is trivially true. However, Parfit had also previously argued that a posited identity relationship between the normative reasons and some natural descriptive property (in this case, conduciveness to desire-satisfaction) could be trivially true even when it was *non-analytically* reductive.

He argues as follows; a naturalist might want to argue that,

(A) If some act would minimize suffering then, all else being equal, this act is what we ought to do.

Parfit takes this as implying one of two things. Either,

(B) If some act would minimize suffering then the fact that it would do so makes it also have the property of being what we ought to do. Or,

(C) If some act would minimize suffering then that is *the same thing* as it being what we ought to do.[140]

Parfit is quick to dispose of (C). He maintains that this is because this particular identity would be trivially true if it were. What minimizes suffering simply *is* what we ought to do according to (C), and would express an identity relation. However, we think that any statement of the form (A) takes must be informative – i.e. it must be capable of providing genuine information. However, no genuinely trivial statement could be informative in the way we think (A) is. Thus, Parfit dismisses (C)[141].

However, as Parfit acknowledges, Schroeder and others find his grounds for dismissal puzzling. Surely, to be *trivially* true, (C) would have to be uninformative and in order to be uninformative it would have to express an analytic identity. Yet if it were expressing a synthetic identify it could easily be informative. Since I am not aware of anyone who argues that (C) is analytic, Parfit's argument for its triviality would seem to be in need of further argument. However, since this specific point is not relevant to the rest of the discussion, it will have to wait for another time. The argument that Schroeder wants to make – i.e. that the relationship between the normative and the descriptive could conceivably be one of causal necessity rather than identity – is not committed to anything like (C), and so is not undermined by Parfit's triviality objection. Though some of these points are relevant to naturalism, it is not crucial that they be applicable to that debate – see the Section 6.5 below, 'The Scope of This Thesis'.

Where the disagreement between Parfit and Schroeder really comes in is with (B), or with any sentence similar to (B). Just for our purposes, let's use a sentence similar to (B) but that is friendlier to Schroeder and any potential soft Internalists.

(D) The fact that some act φ promotes some agent A's desire δ makes it have the property of being the thing that A ought to do.

Like (B), (D) is not intended as a statement of identity. It posits a necessary causal relationship between A's normative reasons and a fact about the world – specifically, that A's φ-ing will promote δ. Parfit takes Schroeder to be trying to assert that sentences of the form (A) can, *mutatis mutandis*, be rendered into sentences of the form (D)[142]. Furthermore, if (D) or some other sentence very like (D), turns out to be true it would provide an account of a normative reason that complies with the

[140] Derek Parfit, *On What Matters Vol. 3*, Oxford University Press (2017), p.137.
[141] Ibid, p.140.
[142] Ibid, p.143-159.

motivational requirement whilst simultaneously sidestepping the Endemic Error. This being the case then, all that Schroeder or some soft Internalist needs to do is to provide a convincing account of *how* the normativity of *A*'s reason to *φ* tracks, by necessity, the fact that *φ*-ing is conducive to promoting *δ*.

This, however, is *not* what Schroeder attempts to do. Instead, he maintains that in order to establish the relationship between the normative reason to *φ* and the desire promotion of *φ*-ing that is posited by (D), and in a way that is not reductive, it is not necessary to posit that the normative reason is anything over and above its conduciveness to promote *δ*. In other words, he thinks that the truth of (D) and like sentences do not depend on either an identity relation *or* a causal relation. He takes Parfit as attempting to force him onto the horns of a false dilemma.

By Schroeder's lights, (D) is more akin to a straightforward description of what a normative reason *is* – not anything it is identical to or caused by. For him, there are not two aspects to (D) – one normative and one descriptive. For an agent to have a normative reason *just is* the fact that some action would promote a desire of that agent. There is no need to invoke some superfluous normative property or power to explain how it is a reason, which then needs to be worked in to the account[143].

However, Parfit considers this an illegitimate move. He writes that this supposed capacity, that some action promotes an agent's desire, to account for how this makes it a reason that the agent *should* do it might just as easily be treated as a property in its own right. This is what Parfit refers to as 'the explanative property'. He goes on to argue that this explanative property and its relationship with the fact that *φ*-ing is conducive to the promotion of *δ*, would then itself stand in need of a normative explanation. Schroeder is moving the mystery on rather than actually resolving it. As such, Schroeder's purely descriptive approach to account for the relationship between normative reasons and motivational states would have no hope of being re-purposed and deployed by internalists to avoid *The Endemic Error*. Internalists and constructivists alike would be back to square one.

In a typical show of generosity and fairness to his opponent though, Parfit acknowledges that Schroeder does not actually consider himself backed into a corner in this way and has a kind of argument to prevent it. As I said earlier, Schroeder does not consider himself a soft naturalist of the kind who is vulnerable to this sort of attack,

[143] Ibid, p.150.

since he does not believe there are any truly irreducible normative properties. In a sense, he does not think there is any mysterious aspect to normativity that requires explaining. Normativity for Schroeder is nothing over and above the facts concerning certain actions' conduciveness to desire-promotion. Statements like (D) rather than 'explain' *that* an action's promotion of an agent's desire makes it normative reason for that agent, are simply definitional. They merely express that the fact that the action promotes the object of the desire *is* the normative reason.

Parfit however, rejects this possible line of defense. He does this by making some important distinctions into how an explanation of this kind works – what it can do and what it can't do. He uses the example of the identity relation often posited between the pre-scientific concept of heat and that of molecular kinetic energy. The reason I specify 'pre-scientific' understanding of heat, is that we want to allow for the fact that 'heat is identical to molecular kinetic energy' to be informative. The identity statement as we are discussing it here is the informative statement that the property of being molecular kinetic energy is necessarily co-extensive with the pre-scientific property that, among other things; melts things, triggers combustion, causes mercury to expand, etc.

Now, Parfit acknowledges that in one sense, heat is reducible to molecular kinetic energy. However, Parfit wants to argue that they are not identical in what he calls a 'description-fitting' sense. By this he means that while they do refer to the same thing they do so in a crucially different way – which is what allows the statement of their identity to be informative and not trivially true[144].

Having molecular kinetic energy explains why something has the properties of something that is hot – i.e. melts things, etc. Yet being the property that explains why something melts is not the same thing as melting or being melted, according to Parfit. The property that *explains* a property is not the same thing as that property itself. A property and what explains it, at least metaphysically speaking, is not identical with its effects or powers. To think otherwise is to commit a kind of fallacy of composition.

Likewise, being something that explains why an agent is motivated to act a certain way – i.e. the normative reason – and an agent actually being motivated to act a certain way are not at all the same things. That an agent should act under certain circumstances might serve as an explanation of why they acted, but that is not the same as the cause of their action. This is the false conflation rife in Internalism that I

[144] Ibid, p.153.

mentioned in the previous section, between the normative reasons for action and explanative reasons for action. Parfit called this the 'lost property problem'[145]. He believed that any theory that tries to pull off the kind of move Schroeder is trying to here will make such a fallacy – and hence, simply will not be able to achieve the desired satisfactory account of the relationship between normative reasons and facts about motivational states. Whichever way you cut it, ontologically speaking, Schroeder's account of desire-conduciveness just isn't rich enough to provide a fully satisfactory account of normative reasons.

This attempt by Schroeder is the best I am aware of to provide an argument that could be easily re-purposed by the kind of soft internalist we've been discussing, to sidestep *The Endemic Error*. However, for the same reasons that Parfit articulates, I believe it is simply untenable. As far I am concerned then, the internalist is reduced (no pun intended!) to the following three options,

i) Attempt a non-analytic reduction of normative reasons and facts about agents motivations, based on some kind of identity claim. This will inevitably incur *The Endemic Error*.
ii) Attempt a non-reductive, 'deflationary' and purely descriptive account of normative reasons, where 'normativity' is nothing outside or beyond conduciveness of certain actions to promote certain desires. This however, will inevitably incur the missing property objection.
iii) Provide an account of the necessary causal relationship between normative reasons and facts about the motivational states of agents.

Drawing my own conclusions from the debate then, for obvious reasons, I consider both the (i) & (ii) unacceptable. As for (iii) – Yes, absolutely! If any internalist theorist can provide a compelling case as to why or how an agent's normative reasons necessarily track their motivational states, no one would be better pleased than I. However, the fact that no such compelling explanation has been forthcoming leads me to suspect that no such account is possible. In turn, this failure leads me to make the clarification of the following section.

[145] Ibid, p.141.

6.4 The Endemic Error and Internalism: Principle vs. Practice

I am aware that some of my arguments could lead the reader to misidentify what I take the essence of *The Endemic Error* to be and what its precise implications for Internalism are. My position is that the motivational requirement places a particular kind of constraint on the kinds of things normative reasons can be when viewed by the lights of Internalism, and that this constraint renders it impossible for internalist theories to generate genuine moral reasons. Furthermore it is this requirement, specifically *qua* a constraint, which in practice invariably, at least to my knowledge, leads to internalist theorists committing the error. This is my own argument that I make specifically to distinguish what I am saying from the discussion of the Parfit/Schroeder debate.

However, some of the arguments I have deployed in Sections 6.2 & 6.3, specifically those relating to Parfit's points contra Williams and Schroeder, could lead the reader into thinking that I am mistakenly taking Internalism as being the view that motivational or explanative reasons are either *identical with* the normative reasons they ground or are else *sufficient* to ground normative reasons, instead of merely being necessary. If I were to be doing this then I think it would be accurate to say that the true target of my critique would be Constructivism rather than Internalism. I have two points to make in response to this.

The first is that in practice almost every single theorist who does in fact accept the motivational requirement, when pushed, will reveal themselves to have tacitly accepted that motivational reasons are in some sense identical with normative reasons or are sufficient to ground them. In other words, all internalists in practice are constructivist to a salient degree and enough to render their theories susceptible to the charge of *The Endemic Error*. If the reader can think of a counterexample of this phenomenon, I would be indebted to them to furnish me with it, for I have not encountered one in all my studies.

The second thing I have to say in response relates to the formulation of Internalism in principle. Am I willing to concede that when we attend strictly to the letter of the thing that Internalism is not *necessarily* committed to *The Endemic Error*? Yes, I am prepared to concede this. Indeed, in closer the prior section, I indicated one avenue an internalist could pursue to avoid the error. However, this is a concession that inspires rather than disheartens me to any degree. I maintain that in the mouth of an Internalist, 'normative reason' is invariably relegated to being nothing over and above

what an agent could or would do in certain circumstances. In practice, when carving out the necessary role for motivations in grounding reasons they leave no room for normativity to be outside of this. Even when I have granted that Internalism allows, in principle, for normative reasons to be something apart from the motivational reasons that ground them, *no account of this 'something' is ever forthcoming*! It is highly conspicuous by its absence.

All of this is extremely telling re: Internalism's chances of ever providing genuinely normative reasons. The lack of any richer account of what a normative reason is beyond its motivational grounding seems to imply that internalists consider their job having been done in this regard merely by establishing the motivational requirement. In order to avoid this charge an internalist need only provide some account of what it would be like for an agent to have a normative reason to act that could not motivate them – even if they then go on to make their central, *additional* claim that *all* such accounts would be lacking as they can't meet the motivational requirement. The absence of any sense of an independent account of a normative reason, which could in principle elucidate an aspect of the normative that transcends their motivational role leads me to conclude that they take the establishment of the motivational requirement as being sufficient to account for what normative reasons essentially are. It is *this* mistake of taking such a limited view of what constitutes an adequate account of normative reasons, rather than asserting that internalists necessarily identify normative and motivational reasons, which is the true essence of *The Endemic Error*.

However, I accept that my suspicion that this conspicuous absence of an independent notion of normative reasons – albeit very strong and I hope from this thesis, evidently justified – is not enough to enable me to conclude in the current work that Internalism is fundamentally undermined by *The Endemic Error*. Yet, at the same time I am encouraged to show that very thing is the case in a future work. I shall have to content myself here though, that the burden of proof may have been shifted somewhat onto the internalist to show that they can in fact provide some account of normative reasons that does not reduce to mere motivational or explanative adequacy. To my knowledge, this point regarding the discrepancy between the stated goals and the actual practice of internalists has not been highlighted before.

6.5 The Scope of This Thesis

Just parenthetically, I think it important to pause to clarify some things about the scope of this thesis. I do not wish to attempt to over-reach or to salt the earth for other non-internalist theories, which I hope might provide a satisfactory moral theory one day.

This is not intended as any kind of anti-naturalist or anti-reductivist thesis. I am receptive to the possibility that normative reasons could be given a thoroughgoing and exhaustive naturalistic account, and nothing written here is meant to imply otherwise. I am even open, though far less optimistic, to the prospect that a reduction of normative reasons to mere descriptive facts might be possible. However, my skepticism about this latter prospect is also informed by my belief that they would probably have to trade on the same kind of flawed conflation that I accuse Internalism of. That however, is a discussion for another time.

My target has always been and remains confined to Internalism. To repeat, my reason for this is only that motivational states are not suitable grounding for any such reduction. Internalism is limited from the outset as to what descriptive facts or parts of the world it may rely on to ground normative reasons. In practice (and possibly constitutionally) Internalism takes the limits of the normative to be the limits of the motivational. Facts about the motivational states of any given agent will never be rich or potent enough to capture the quintessential character of normative reasons. If this is true it follows that no internalist theory could ever have sufficient scope to ground moral reasons, as I have argued the kind of things moral reasons must be.

Let me reaffirm what this thesis is trying to argue for. Consider four different questions,

1) Are all normative concepts identical to some descriptive concepts, *where those descriptive concepts are just about our desires and motivations*?
2) Are all normative properties identical to some descriptive properties, *where those descriptive properties are just about our desires and motivations*?
3) Are all normative concepts identical to some descriptive concepts, *if those concepts needn't be about our desires and motivations*?
4) Are all normative properties identical to some descriptive properties, *if those properties needn't be about our desires and motivations*?

I would hope that it would be no surprise to anyone who has read this thesis, that my answer to both questions (1) & (2) would be a resounding 'No'!

My thesis is not about the kinds of identity relations that may or may not exist between descriptive and normative properties or concepts. I am not dealing with ontological questions here, though that is not to suggest that they are not fascinating or important. For the sake of argument, I'll concede that an identity relation could exist between the normative and the descriptive and that this might be all we need to finally solve the is/ought problem. My point is only that the descriptive properties that hold the key to demonstrating such a relationship can't be limited to descriptive properties or concepts that just concern our desires or motivations.

However, when it comes to questions (3) & (4) this thesis has no firm stance, and nothing I argue for within stands or falls depending on any answer I might provide to them. The failure of Internalism to provide moral reasons does not imply that no moral theory that tries to identify the normative with the descriptive will also fail. The return of 'no verdict' I give for question (3) means that the door is left open for a naturalist realist account of normative reasons. Equally, the same return of 'no verdict' on (4) leaves the door open for non-natural or robust realist accounts to succeed. These things do not concern us here, and I do not want my rejection of Internalism to be misconstrued as a more general assault on other areas of metaethics and moral theory. I am not sure where the ultimate truth resides, but if we can rule out which category of descriptive properties or concepts it definitely isn't, I consider that to be no small form of progress.

6.6 Implications for Genuine Moral Reasons

I have argued throughout this thesis that in order for a reason for action to be genuinely moral, it must meet three different criteria. The reason must be categorical; it must be of non-negligible weight; and it must be of the right grounding. In Chapters Three, Four & Five, we looked at individual theories that attempt to supply moral reasons that purport to supply moral reasons that meet the motivational requirement that internalists insist on. I tried to show how each of these theories failed to meet at least one of the three criteria and thus were not in a position to supply moral reasons.

Now however, in the light of the discussion in the foregoing sections of this chapter, I want to speak more specifically of why in practice *every* internalist theory fails to meet all three criteria at the same time. They fail because of the fundamental error Internalism continually falls into, in one way or another, of thinking that

normative reasons of the kind moral reasons have to be can be adequately analyzed or encapsulated by nothing over and above the motivational states of agents. To my knowledge, such employing this particular line of attack to elucidate exactly why Internalism failures to provide an independent standard for normativity condemns it to being unable to produce genuine moral reasons is completely original.

In this section we'll be examining each of the three criteria, and how *The Endemic Error* affects each in such a way that in order for an internalist theory to bolster one, it must make certain concessions to the others, which makes simultaneous satisfactory meeting of all three unachievable.

6.6.i Categoricity

Lets just go over again, quickly, what we mean by the categoricity of moral reasons. As we briefly went over in Chapter One, a reason is categorical (or sufficiently categorical-like) if it is a reason that an agent has whatever their desires are (or other motivational states).

How does this relate back to *The Endemic Error* in Internalism? Remember the problem we discussed in Chapter Two. Stated in its simplest possible form, the problem for Internalism is that an agent absolutely can't have a reason to do something they are not or could not be motivated to do, on the one hand; and on the other, for a variety of possible reasons, agents very frequently are not motivated to do what we would conventionally say they have strong moral reason to do. In which case, if normative and motivational reasons are not fundamentally in the same business, there can't be any sound reason for believing in a perfect or even sizable overlap between the two. Consequently, this means there is an ever-present possibility of mismatch between what there may well be normative reason for us to do and what reasons exist in virtue of their meeting the motivational requirement.

Internalism's only path to success would be by fudging the distinction between the grounding of normative and internal reasons. My contention though is that the dichotomy between internal reasons and moral reasons, is caused by their being of very different stripes and their having distinctive groundings. Because of this, there is no way to put it beyond doubt that a normative reason might exist where the requisite correlating internal reason might be entirely absent. Even if it turned out to be the case that everything for which agents had moral reasons for doing, they also had internal

reasons for doing, this would be only an instance of pure good fortune only. The kind of categoricity that moral reasons have to have can't work that way. It can't be the result of a complex of desire-grounded internal reasons, which just happen to come out as being invariably present. Again, this is to get the direction of fit the wrong way round. If an agent has a moral reason to do something it will be because it has something like the elusive genuine normativity or objective authority we want moral reasons to have. It is this characteristic of a moral reason that makes it beholden to an agent's subjective motivational set to incorporate it – i.e. it is the fact that there is a moral reason to do something that makes it fitting that an agent should desire to do it or be able to be motivated to do it. However, if the only moral reasons that exist are those already contained in or generated out off the elements of an agent's SMS, this idea of there being anything like a standard of genuine normative reasons evaporates. The best-case scenario for Internalism is that it can provide 'categorical' reasons that have been contrived to turn out as merely always true.

But there's more. Not only does Internalism have a problem generating categoricity purely in itself, but I also want to argue that the only means open to Internalism to even try and provide categorical or sufficiently categorical-like reasons, is by further compromising their weight or grounding.

For an internalist theory to have a hope of meeting both the motivational and categoricity requirement it must give an account of reasons wherein it could not be the case that an agent might lack a desire or other motivational state that could motivate them to act as they do in fact have a categorical moral reason so to do. Since moral reasons must always apply to an agent, where they do apply, the theory must not allow for the possibility that they might lack the motivation[146]. This means that the desires necessary to motivate action in alignment with moral reasons have to be so perennial that they necessarily become generic to all agents. Yet, given the idiosyncrasies of actual agents, any desire that is so generic that it can't fail to be present means that either its weight or its grounding will invariably be compromised.

Furthermore, as we saw was the case with both Schroeder and Korsgaard; the only way the internalist theorist can render reasons that both meet the motivational requirement and are easy enough to come by that they will be present regardless of the idiosyncratic make-up of any given agent, is to transform them into something barely

[146] I am here discounting instances of *akrasia*.

recognizable or serviceable as a reason. The idiosyncratic nature and potential for difference between agents means that the prospect of locating an internal reason that is sufficiently widespread or universal enough to be fit for purpose will be nearly impossible. If there were such a reason though, it would have to be, by necessity, one so peripheral to the main idiosyncratic concerns of any given agent that its weight, relative to the more personal concerns of the agent, will be such that it will easily be outweighed by the personal concerns of the agent in question. Generic reasons like these, which could reliably and routinely be expected to be outweighed could not service as moral reasons.

The categoricity requirement will never be met because the only viable source of a reason's categoricity lies in it being irreducibly normative in character. By placing the final impetus on a reason being motivationally efficacious, internalists make any normativity this reason might or could have contingent on a fact regarding its motivational potential. This means that for most internalist theories, there will always be the possibility that the subjective motivational content of an agent's psychology will never be able to lead to them being motivated to do what they should be motivated to do. The categoricity of moral reasons demands that where a moral reason applies, it could not be otherwise that an agent should have a reason to do it, regardless of whether they are motivated to do it.

Even in the case of Korsgaard, where agents can't help but have the kinds of reasons to do moral things, the reasons are still only accepted in her moral ontology because they pass the motivational test. It is their power to motivate that determines whether or not they are normative, not whether they are normative that dictates whether they should motivate. On this model, the normativity of moral reasons is stripped away and is thus no longer fit for purpose to be a moral reason.

6.6.ii Weight

As was just discussed in the preceding subsection, a perennial problem for creating internal reasons that are, or sufficiently approximate, categorical reasons for action is that they must be made so generic or perennial that they lack sufficient strength to outweigh the kinds of counter-veiling reasons that are furnished by the idiosyncrasies of any given agent.

In terms of the requirement of weight, taken in and of itself, however, the price for committing *The Endemic Error* is just as harmful when internalist theorists attempt to provide a satisfactory account of weight or strength of a reason. The key way to understanding this problem is not by asking what grounds the weight of the reason, *per se*, but what regulates it. In the case of the former, somewhat predictably, the internalist's account of the grounding of a reason's weight must include a motivational state. On the other hand, when it comes to the weight of that reason, the internalist must provide an account of how one reason that meets the motivational requirement can be stronger or weaker than another reason that likewise meets the motivational requirement – i.e. once the motivational requirement is taken as met, what is the variable that accounts for differing reasons' weights?

The options available to any top-to-tail internalist theory are somewhat limited. According to one view, which Schroeder dubbed Proportionalism, the weight of the reason is somehow proportional to the strength of the motivational state that grounds it. The stronger the desire, say, an agent has to achieve or attain *x* then, all things being equal, the weightier an agent's reason is to undertake an action that promotes their achievement or attainment of *x*. Yet, as Schroeder's fundamental insight teaches us, though he does not frame it this way explicitly, motivational strength/efficacy and normative weight are not the same animal at all – they are in quite different categories. A moral theory that allowed a psychopath's thoroughly overriding desire to torture innocent people, to imply a massively weighty normative reason to do so would of course be totally unacceptable. It is both an actual fact and an ever-present possibility that motivational states will exist that are stronger than an agent's motivation to do what they have upmost moral reason to do. This makes the motivational strength of a reason unfit for the purpose of grounding normative strength. Internalism is thus incapable of providing an adequate account of the weight of moral reasons so long as weight is determined by some quality of motivational states themselves.

This, as we've seen, is what led Schroeder to wisely abandon Proportionalism in favour of Hypotheticalism. Instead, the weight of a reason is determined by the 'appropriateness' of placing weight on it in our deliberations. However, it is this aspect of Schroeder's theory that has drawn the greatest amount of attention and criticism since the first publication of *Slaves of the Passions*. He seems to be invoking some

objective standard for the appropriateness or weightiness of reasons that is wholly disconnected from any quality of the desire that grounds it.

Likewise in Gauthier, we saw an attempt to ground all moral reasons on the long-term interests of agents. At first glance, this does seem a promising strategy if we assume the plausible thesis that all agents have the weightiest reason to look after their own interests. However, as all moral reasons are ultimately grounded in Gauthier's system by the same interest – i.e. the interest of obtaining the best situation for oneself, when living in society with other agents who are doing likewise – it is not the intensity of this self-interest that can account for the varying weights of Economic Man's diverse set of moral reasons. For example, it is hard to imagine any system of law or convention that Gauthier has in mind, where and agent should not have weightier reason to not brutally murder their neighbor than to simply steal a loaf of bread from them. Ultimately, how can Gauthier account for this difference in weight? Inevitably, it will have to be some other quality belonging to acts of murder or petty theft that will account for their different ranking within the hierarchy of crimes of a given social system – and the acts will have these qualities independently of the motivational states that ground them. The point is this, both Schroeder and Gauthier are forced to 'outsource' from the motivational states that ground reasons. The role of regulating the weightiness of reasons has to be transferred to some other extraneous property of those reasons or position they hold in a hierarchy.

They general point I am making is this; in attempting to provide a satisfactory account of reasons' weights, the internalist is placed onto the horns of a dilemma. On the one horn, they can commit the mistake of something we might classify as Proportionalism, in the broadest sense. In which case they will walk straight into *The Endemic Error* that renders motivational efficacy incapable of providing normative strength. On the other horn, they are required to shift the burden of providing weight from the motivational state to some other element that is independent of the motivational state – in which case that theory's account of weight, taken in-and-of-itself, while not strictly incompatible with Internalism, ceases to be in anyway incompatible with Externalism. In which case, there is no longer any advantage to being an internalist when striving for an adequate account of reason weight.

When it comes to providing an acceptable account of the weight of moral reasons then, the internalist must either fail or abandon any remnant of what led them to want to be internalists in the first place.

6.6.iii Right Grounding

The Endemic Error is most evident when it comes to the right grounding requirement. Facts about motivation or potential motivation might be able to express *that* an agent does or could do something. However, they can't capture why they *should* act a certain way. The grounding of a moral reason is precisely the kind of thing that could explain why an agent should do something even when they are not motivated to do so – i.e. why their lack of motivation, or potential to be motivated, represents a moral failing or shortcoming of some kind.

As I outlined in Chapter One, the right grounding for a moral reason must be at least in part constituted by something that is essential to that act being the act it is. An act of murder being an act of murder, rather than some other kind of homicide, must have certain elements about it – e.g. the certain *mens rea* of the perpetrator, the unwillingness of the victim to die, there being no legal justification for the act, etc. The precise definition of any immoral act is not crucial here. The point is, any action we would consider there is a moral reason for us to do or not do has some key element or nexus of elements that make it that kind of act. It is this element or nexus that must ground the moral reasons we have. Motivational reasons are contingent on the psychologies of the agent. What makes a certain act a certain act, does so necessarily – since they involve what is essential to that act. For this reason, the contingent motivational states of agents are inadequate to serve as the necessary grounding moral reasons must have.

To put it starkly, there is no *prima facie* reason to believe that a person will always lack overwhelming motivation or the potential to do something morally wrong. Yet it makes absolutely no intelligible sense to say that a moral reason not to murder an innocent person could be absent. What makes an act morally wrong is what the act essentially *is*, not what any agent's motivational disposition is toward the act. Our very process of forming moral opinions regarding certain action occurs in isolation from considerations of the motivational dispositions of any agent who might carry them out.

Normative reasons, if there ultimately are such a things, are something constituted by the essence of the action or state of affairs itself.

Where internalists consistently go wrong is by giving motivational or explanative adequacy of their theories primacy and then extrapolating outward from there in the hope of capturing moral reasons. They fall into the perennial trap of thinking that the grounding of a motivational reason can be sufficient to supply a moral reason's normative grounding also. What motivates agents and what makes things morally wrong simply isn't the same thing. Believing that they must be is to sets the internalist up for failure from the outset.

6.6.iv To Summarize i-iii

The problem is systemic. So long as internal reasons are required to 'serve two masters' – i.e. simultaneously meeting the three criteria that are essential for moral reasons to be moral reasons, *and* being motivational efficacious, they will always fail to meet one of the criteria in some important regard. The only options available to mitigate this systemic problem require compromises that further jeopardize the internalist theory in question from meeting one of the other criteria. Where-as alternative strategies to avoid these problems involve abandoning those elements of the theory that make it internalist in the first place. *The Endemic Error* renders the motivational 'requirement' into more of a motivational *constraint* on any moral theory it's possible to generate while complying with it. It's a constraint that makes the job of the internalist theorist, in providing reasons for action that are genuinely moral, impossible.

6.7 On the Prospect of Employing Alternative Normative Concepts

There has been a great deal of interest taken in recent years in more closely examining the actual normative concepts with which theorists have become comfortable using when discussing questions of the kind we have been covering in this thesis. We have been using normative concepts such as 'reason', 'categorical', 'weight', and of course, 'normative' itself. It is hard to imagine anything akin to this discussion without them. I believe it is only fitting to examine this with an, albeit, brief discussion of this avenue to saving Internalism from what I see as its endemic shortcomings. What we'll be looking at here harkens back to what I wrote about in Section 1.5, 'Conceptual Requirements vs. Commonly Held Intuitions'.

Some of the fruitful discussions, Eklund (2017)[147], has pondered has been to question whether or not there might be something inherently unfit for purpose about these concepts; that the tools we are using to resolve the problems are incapable of helping us solve the issues, and in fact might be contributing to the apparent problems in the first place. It has been suggested that there could be better normative concepts to use than these ones – or that these concepts could be adjusted so that they can do the job better. Eklund uses the term 'conceptual engineering' for projects of this kind. Could a re-engineered notion of 'moral reason' be the key to dissolving the apparent conflict?

In a sense, we have already looked at shades of this idea with our discussion of Copp in Chapter One. Recall, Copp argues for a teleological view of ethics. Morality serves the utilitarian function of facilitating or ameliorating the problems that inevitably result from human interactions. It is from its success or failure to do this that morality gains, or fails to gain, its legitimacy. Similarly, in the writings of error theorists like Mackie and Joyce, once their arguments have been made, they are left with the task re-imagining something *like* morality to fulfill its still vital role in human life. For Mackie this took the form of his loosely defined 'rule-right-duty-disposition utilitarianism' and for Joyce it was his Fictionalism. A common strand to all three is that is that the reasons for action they generate are given weight by the desirable results they give rise to. Whether we literally consider this morality, as with Copp, or just sufficiently morality-like as with most error theorists, they are considered as close to moral reasons as we are likely to get.

However, the kind of conceptual engineering Eklund and others are discussing is one that involves alterations to the very concepts being employed to bolster our theories. Could it be that categoricity, weight and grounding do not actually play the kind of crucial role in moral reasons that I have argued they do, or alternatively, that with a different notion of them and their function some form of Internalism might be able to create moral reasons?

In Section 1.5, I acknowledged that the right grounding condition, rather than being a conceptual necessity for moral reasons may well be nothing more than a very strongly held intuition concerning the way moral reasons must be grounded. For this reason, if a moral theory could meet the other two requirements, I might be willing to accept that it could in fact be sufficient for moral reasons. However, I argued for the

[147] Matti Eklund, *Choosing Normative Concepts*, Oxford University Press (2017).

conceptual necessity of both categoricity and weight. A moral theory that can't give rise to moral reasons that can't fail to apply to agents, and also generate at least some reasons of tremendous weight should not be accepted. In which case, if an alternative satisfactory conceptualization of both of these normative concepts could be found – alternatives that *can* be met by internalists – my thesis would be undermined and Internalism might be workable after all.

Now, in a manner of speaking I have already conceded that a slightly adapted conceptualization of categoricity is acceptable to me. I have acknowledged that to meet what I call the categoricity requirement a moral reason doesn't necessarily have to be categorical in the strict sense that an agent could have it even if they have no motivational state that would motivate them to comply with it. If it were the case that agents could not be otherwise constituted so that they will always have some motivational state or desire to motivate them to comply with their moral reasons, I take this as sufficiently categorical-like to have met the requirement. In this sense I have been operating throughout this thesis with a moderately re-engineered concept of categoricity. And I remain open to this avenue as the internalist may utilize it.

However, as I have shown in Chapters Three-Five, some of the best attempts do this have fallen short in some way. Additionally, it is in the nature of *The Endemic Error*, as it rears its head in Internalism, that its effective focus on explanative adequacy over normative adequacy will always tend to mean that whatever overlap exists between motivational reasons and the kinds of normative reasons moral reasons have to be will be a contingent one. In turn, this will make the prospect of mismatch between moral reasons and the motivational states of agents an ever-present danger.

In his *Choosing Normative Concepts*, Eklund provides a pretty exhaustive analysis of the different prospective ways alternative normative concepts might be utilized to solve the problems that myself and Externalist or Realist theorists continually raise. He is even critical of Parfit's argument for the irreducibility of the normative to the natural (or descriptive) that we covered in Section 6.2, arguing that it rests on a failure to distinguish between normative properties and normative concepts. It is not necessary for natural property terms to be identical to normative properties. It is only necessary that that we can be confident that natural and normative concepts can be used co-referentially. If this is the case, I believe Parfit's argument would be undermined, as well as my accusation of *The Endemic Error*, at least to a small degree. There may be a way

for internalists to avoid what I have argued would result in the strict dichotomy between Reasons Externalism & Error Theory – between 'Externalism or Bust!'

I must confess, at present I do not have an adequate response to this criticism and for now it also will have to be the dedication of a future work. That being said however, I am likewise not disheartened. For at the end of Eklund's thorough analysis of the options available to those who would re-engineer our normative concepts, he is forced to conclude that their just appears to be something 'ineffable' about normativity that does not seem to be able to be captured using alternative normative concepts. This remains my own position and I believe is in line with my general criticism against internalist theories, and, indeed, any theory that would sacrifice normative adequacy for descriptive adequacy.

It is possible that other theorists just do not attach to moral reasons the same characterizations that I think essential to it, and so would be happy to call the moral reasons they generate 'moral' in some sense that they find acceptable. I just can't imagine any theory of this kind convincing me. An analogy used by Schroeder in *Slaves of the Passions* might help here. Schroeder asks us to imagine an atheist friend who announces that they now believe in God. When they are quizzed a little further however, we discover that they have changed their definition of 'God' to just mean nature and the universe. We might not quibble with them over a word, yet at the same time I do not believe it is truly accurate to say that this person does believe in God. There are certain things we demand 'God' to be to be God at all. When it comes to moral reasons, this is essentially what I have been arguing throughout.

Those certain things that moral reasons must be are difficult to define. It maybe that for the time being, expressions like 'ineffable', 'queer', 'normative oomph', 'to-be-done-ness' are the best we can muster. Whatever term we use however, they fundamentally grasp something common, in that if moral reasons do exist there must be something about them that makes it appropriate that *we* comply with *them*, because of whatever it is they are or what qualities they possess. Fore-limiting what moral reasons there are or could be by making them comply with what we are, in terms of how we are motivationally constituted, is to abandon the fundamental part of what moral reasons should mean to us. Some theorists might be comfortable describing a set of reasons that lacks this quality 'moral'. I however, never would be.

I am convinced however, that both the inescapability of moral reasons as and where they apply to agents, provided by their categorical character, and the important role they should play in at least some, if not many of our deliberations, is essential to anything we should call morality. It is possible that there are other characteristics that also have this essential role in moral reasons. It is also possible that the characteristics as I have described them might be better articulated. However, I believe that the characteristics of moral reasons as I've outlined them strike at the heart of something crucially true of ethics, which Internalism will always fail to capture.

In closing this chapter, and the thesis as a whole, I will simply re-state my long-held stance that the search for moral truth and the search to find our innermost motivational cogs and springs are enterprises of such a completely different stripe, that it seems constantly surprising to me that the internalist attempts to sully the former by over-burdening it with considerations of the latter

Conclusion

We are now in a place where we can draw our conclusions. But first let's recap on how we have got here.

In the Introduction I stated the three criteria I believe any reason that could be called 'moral' must meet. They must be categorical, of non-negligible weight and have the right grounding. Furthermore, I posited that Internalism not only hasn't produced a moral theory that meets these three criteria but that it is incapable of providing reasons that meet these three criteria. If we accept that Internalism is so incapable, which I believe we should, we would be forced to conclude one of two things. Either there are, in the final analysis, no truly moral reasons for action, or moral reasons are external reasons. I made it clear that the purpose of the thesis was not to advocate for either of these two positions – only that it must be one of them as they are the only two positions consistent with what characteristics moral reasons must have. It is external moral reason or no moral reasons – Externalism or Bust!

In Chapter One I clearly outlined each of the three criteria in detail and defended each of them from possible criticism, making clear why each of them are essential to anything we wish to class as a moral reason.

In Chapter Two I briefly explained why each of the three criteria presented a *prima facie* problem for Internalism to incorporate, but that this did not mean it was not a challenge that could be met in principle. I did however make it clear what the minimum desiderata would have to be for any internalist theory to say that it had in fact generated a genuine set of moral reasons.

In Chapter Three I began my assessment of three different though highly prominent internalist moral theories, with the Contractarianism of David Gauthier, and how this specific theory fails to meet the required criteria.

In Chapters Four & Five, I gave the same treatment to the Neo-Humeanism of Mark Schroeder and the Neo-Kantianism of Christine Korsgaard, respectively.

In Chapter Six I argued for my thesis that internalist theories consistently fail in practice because they commit what I refer to as *The Endemic Error*. This is a mistake perennial to Internalism as it results from adherence to the motivational requirement, which invariably leads them have a stunted and diminished idea of what constitutes a normative reason. I further outlined how *The Endemic Error* makes has made all

internalist theory do date unable to meet the three criteria and give rise to genuine moral reasons – and furthermore, why this perennial failure suggest that the prospect of any internalist theory in principle being able to do so is highly dubious.

If my own original arguments for *The Endemic Error* and the implications I argue they have for any internalist moral theory are correct, then I believe the original dichotomy that provides this thesis with its title follows. Either reasons for action that can and do meet my three criteria exist or they do not. If they exist then these will be either internal reasons or external ones. Since I have ruled out the possibility that they are internal, it would mean that if they exist at all they could only be external reasons. Of course, it is possible, and I am genuinely receptive to the possibility that this is the case, that no reason can meet the three criteria. In which case, moral reasons are left with nothing *to be*, and we must conclude that they do not actually exist. This would mean accepting that all references to moral reasons are guilty of positing the existence of entities that do not exist – The Error Theory.

I believe it is clear that it is either Externalism or Bust! In which case, we must either redouble our efforts to find external moral reasons or accept that moral reasons are a fiction. In which case, we shall just have to find the best way for us to live together in their absence.

Bibliography

Bukoski, Michael – *Korsgaards's Arguments for the Value of Humanity*, printed in Philosophical Review, Volume 127, No. 2 (2018).

Copp, David – *Toward A Pluralist And Teleological Theory Of Normativity*, Philosophical Issues, 19, Metaethics (2009).

Copp, David – '*Contractarianism and Moral Skepticism'*, presented in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, Cambridge University Press (1991).

Eklund, Matti – *Choosing Normative Concepts*, Oxford University Press (2017).

Enoch, David – *On Mark Schroeder's Hypotheticalism: A Critical Notice of Slaves of the Passions*, Philosophical Review, Volume 120, No. 3 (2011).

Enoch, David – *Agency Shmagency: Why Normativity Won't Come from What Is Constitutive of Action*, Philosophical Review, *Volume* 115, No. 2 (2006).

Enoch, David – *Shmagency Revisited* (2010).

Foot, Philippa – *Morality as a System of Hypothetical Imperatives,* reprinted in *Foundations of Ethics*, Edited by Russ Shafer-Landau & Terence Cuneo (2007).

Gauthier, David – *Assure & Threaten*, presented in *Ethics*, Vol. 104, No. 4 (Jul., 1994).

Gauthier, David – *Morals By Agreement*, Oxford University Press (1986).

Gauthier, David – '*Why Contractarianism?*', presented in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, Cambridge University Press (1991).

Harman, Gilbert – *Moral Relativism Defended,* reprinted in *Foundations of Ethics*, Edited by Russ Shafer-Landau & Terence Cuneo (2007).

Hume, David – *Concerning the Principles of Morals*, Oxford University Press (1999).

Hobbes, Thomas – *Leviathan* (1985)

Joyce, Richard – *The Myth of Morality*, Cambridge University Press (2001).

Joyce, Richard – *The Evolution of Morality*, New York: MIT Press (2006).

Joyce, Richard – *The Error in 'The Error in the Error Theory'*, The Australasian Journal of Philosophy 89(3) (2011).

Korsgaard, Christine – *Skepticism About Practical Reason,* reprinted in *Foundations of Ethics*, Edited by Russ Shafer-Landau & Terence Cuneo (2007).

Korsgaard, Christine – *The Sources of Normativity*, Cambridge University Press (2014)

Langton, Rae – *Objective and Unconditioned Value*, printed in Philosophical Review 116 (2007).

Mackie, J.L. – *Ethics: Inventing Right and Wrong*, Penguin Books (1990),

Markovits, Julia – *Moral Reason*, Oxford University Press (2014).

Parfit, Derek – *Reasons & Motivations*, Proceedings of the Aristotelian Society, Supplementary Volumes, Volume 71 (1997).

Parfit, Derek – *On What Matters*, Volume 3, Oxford University Press (2017).

Prichard, H. A. – *Does Moral Philosophy Rest on a Mistake?* Mind, New Series, Vol. 21, No. 81 (Jan., 1912), pp. 21-37 Published by: Oxford University Press on behalf of the Mind Association.

Prinz, Jesse – *The Emotional Construction of Morals*, Oxford University Press (2013).

Sayre-McCord, Geoffrey – '*Deception and reasons to be moral*', presented in '*Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*', Cambridge University Press (1991).

Scanlon, T.M. – *What We Owe to Each Other*, The Belknap Press of Harvard University Press (1999).

Schroeder, Mark – *Slaves of the Passions*, Oxford University Press (2013).

Smith, Michael – *The Moral Problem*, Blackwell Publishing (2011).

Street, Sharon – *A Darwinian Dilemma for Realist Theories of Value*, Philosophical Studies 127 (2006).

Street, Sharon – *In Defence of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters*, Philosophical Issues, Volume 19, 2009.

Williams, Bernard – *Internal & External Reasons*, re-printed in *Moral Luck*, Cambridge University Press (1999).

Williams, Bernard – *Ethics & the Limits of Philosophy*, Routledge (2006).

Wittgenstein, Ludwig – *Philosophical Investigations*, Blackwell (1999).