A DEFENCE OF MORAL ANTI-RATIONALISM

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University. 06/23.

The contents of this thesis are entirely my own work. To the best of my knowledge, all sources I have used and any assistance received in preparing this thesis have been acknowledged.

Word count: 62,500.

Josef Holden

30/06/2023

ACKNOWLEDGEMENTS

There are many people I need to thank for their help. They include: Jesse Hambly, Domi Dessaix, Shang Yeo Long, Devon Cass, Nicholas Southwood, Dale Dorsey, Alan Hajek, Donald Nordblom, James Willoughby, Kirsten Mann, Lachlan Umbers, Oliver Rawle, Jessica Isserow, Ross Pain, Chad Lee-Stronach, Erik Zhang, Joseph Moore, Jessica Holden, and Mark Kortink. Thanks also to the participants in the Moral Philosophy Work in Progress Reading Group at the ANU and the participants in the ANU-Humboldt-Princeton Summer Institute on Normativity 2018 for their helpful comments on early drafts of some of the material in this thesis.

My greatest thanks go to two members of my supervisory panel. I am very grateful to Philip Pettit for all the time and encouragement that he has given me. The comments that he has provided on various drafts, and the many conversations that we have had about my work, have greatly improved the quality of this thesis.

I could not have asked for a better primary supervisor than Seth Lazar. He has provided tremendously helpful comments on various drafts of each chapter of this thesis, as well as valuable advice on big picture issues concerning what the thesis should look like. In addition, I can't thank him enough for his continuous support throughout this process.

Finally, I want to thank the Australian Government Research Training Program and the Princeton University Centre for Human Values for their financial support.

ABSTRACT

Morality can require us to make sacrifices. In extreme cases, it can demand that we sacrifice our happiness, our projects, our relationships and even our lives. In more mundane cases, it can forbid us from doing what we want to do. In either sort of case, we might ask: 'Why should I care about what morality demands? Why should I do what morality requires?' This thesis is a defence of *moral anti-rationalism* – the view that an agent can have sufficient reason, all things considered, to act immorally. I begin by setting the stage. Chapter One discusses a particular kind of reason for action, which I call an *excellence-based reason*. Chapter Two and Chapter Three defend moral anti-rationalism. More specifically, I argue that both prudential reasons and excellence-based reasons can provide an agent with sufficient reason, all things considered, to act immorally. Chapter Two focuses on prudential reasons and Chapter Three focuses on excellence-based reasons. Chapter Four and Chapter Five respond to two arguments that have been given for *moral rationalism* – the view that, if an agent is morally required to perform an action, then they have decisive reason, all things considered, to perform that action. Chapter Four discusses what I call the *blameworthiness defence* of moral rationalism. Chapter Five concerns a claim that has motivated various philosophers to try to vindicate moral rationalism. This is that it would be bad if moral anti-rationalism were true, and that we have reason to hope that it is not. The arguments that I give in these chapters do not rely on my previous arguments for moral anti-rationalism. My aim is just to show that these particular arguments fail to vindicate moral rationalism.

TABLE OF CONTENTS

INTRODUCTION AND PRELIMINARIES

Morality can require us to make sacrifices. In extreme cases, it can demand that we sacrifice our happiness, our projects, our relationships and even our lives. In more mundane cases, it can forbid us from doing what we want to do. In either sort of case, we might ask: 'When morality demands these sacrifices, why should I care? Why should I do what morality requires?'

These are not moral questions. If they were, they would answer themselves. These questions also cannot be answered by the simple assertion that we ought to sacrifice our interests because morality requires us to do so – or, alternatively, by the claim that our interests have already been given their morally correct weight. We know these facts. What we want to know is why we should care about them. These questions concern the *authority* of morality. To get a better grip on such questions, and to get a better sense of why they matter,

> Suppose that your life is at a crossroads. You must decide whether to Φ or Ψ. And suppose that, after reflecting on the options, you have come to believe that you morally ought to Φ. Despite this, you want to Ψ. This may simply be because you believe that Ψ-ing would be more fun than Φ-ing, or it may be because you believe that Ψ-ing would be better for you than Φ-ing, or better for someone you love. Suppose, in addition, that your beliefs are true. Taking all these considerations into account, should you Φ or should you Ψ?

There are at least two ways to go here. According to *moral rationalism*, it is always true that you should Φ. This is true regardless of the considerations that favour Ψ-ing. Moral rationalism claims, in short, that:

> **MR:** If you are morally required to Φ, then you have decisive reason, all things considered, to Φ.

Moral anti-rationalism is the denial of moral rationalism. On this view, non-moral considerations can sometimes justify genuinely immoral actions. As such, it is possible that, in the above case, the considerations that favour Ψ-ing make it reasonable for you to Ψ. In short:

> **AR:** Even if you are morally required to Φ, you can have sufficient reason, all things considered, to Ψ.

To make these views clearer, it is worth explaining what is meant by the claim that an agent has sufficient or decisive reason, *all things considered*, to perform an action. Here is an intuitive

way to get at this idea. When we are trying to decide what to do, there are different questions we can ask. We can ask, for instance, what we morally ought to do. If there is an action we want to perform, we can also ask whether performing that action is morally permissible. We could ask, instead, which available action would be best for me. That is, which action would be prudent. There are, of course, countless additional examples. We could ask which available action would be best for my career, or would help me lose the most weight by the end of the month, or would most impress an attractive stranger.

Each of these questions presumably has an answer. But, in at least certain cases, the answers to these questions will conflict. It may be immoral to act in the way that would be best for you, and imprudent to act as you morally ought to act. In cases of conflict, there seems to be a further question we can ask: Which of these actions should I *actually perform*? Should I ultimately act morally or prudently? The answer to this question is the answer to the question of what you ought to do, all things considered. It is the 'all things considered' ought in the sense that it takes every relevant consideration properly into account.

Suppose that you perform an action that you lack sufficient reason to perform. You will then have made a *mistake*. You will be acting in a way that you should not actually act. Put differently, you will be acting in a manner that is decisively disfavoured by the reasons that bear on your action. Your action will, in this sense, be *practically irrational*. If, on the other hand, you have sufficient or decisive reason to perform an action, then performing that action is practically rational. It involves no such mistake.

With all this in mind, we can now see more clearly what moral rationalism and moral anti-rationalism claim. Moral rationalism is the view that, if it is true that you are morally required to perform an action, then it is also true that you *actually ought to perform that action*. Moral requirements always have the preponderance of reasons on their side. In the sense noted, moral requirements are requirements of rationality, and, in the same sense, immorality is always irrational.[1] Moral anti-rationalism denies these claims. You can sometimes make immoral choices without making a mistake, and without acting against the preponderance of reasons.

[1] I sometimes use the terms 'rational' and 'irrational' in these senses, but I often use the terms 'reasonable' and 'unreasonable' instead. To stay closer to ordinary usage – where to call an act irrational is a criticism similar to calling the act 'foolish' (cf. Parfit 2011, 33; 1984, 318) – I only use the terms 'rational' and 'irrational' when discussing cases where the agent knows all of the relevant facts.

We can now also see more clearly why the debate between these views matters. This is because it concerns one of the most important normative questions. This is the question of the extent to which we ought to *actually live* moral lives; the extent to which we ought to be moral.

I

This thesis is a defence of moral anti-rationalism. I argue that there are two kinds of non-moral reasons – prudential reasons and excellence-based reasons – that can provide an agent with sufficient reason, all things considered, to act immorally.

Here is the plan: I will spend the rest of this chapter setting the stage. This will involve further clarifying the view that I intend to defend, discussing how I will argue for this view, and making some of my assumptions explicit, such as how I will understand prudential reasons for action.

Chapter One does not directly concern the debate between moral rationalism and moral anti-rationalism. It instead discusses a particular kind of reason for action, which I call an *excellence-based reason*. Since these are not often discussed, I begin by offering a detailed account of what these are. I then argue that excellence-based reasons are genuine and normatively significant reasons for action; that is, that they make a real difference to how we ought to live our lives, all things considered.[2] Finally, I argue that excellence-based reasons are neither moral nor prudential reasons.

Chapter Two and Chapter Three defend moral anti-rationalism. In particular, I argue that both prudential reasons and excellence-based reasons can provide an agent with sufficient reason, all things considered, to act immorally. Chapter Two focuses on prudential reasons. I begin by discussing and motivating an argument that has been made before. I then offer an original argument for moral anti-rationalism and respond to some objections. In Chapter Three I turn to excellence-based reasons. I first argue that the same line of reasoning that supported moral anti-rationalism in the previous chapter also supports the claim that excellence-based reasons

[2] As opposed to considerations that *merely* count in favour of an action from some point of view – or according to some domain, set of rules, or institution – but do not make a genuine difference to how an agent should live their lives, all things considered. This distinction is sometimes labelled as the distinction between *authoritative* and *formal* normativity. For an informative discussion of this distinction, see Baker (2018a, section 1.1). Unless otherwise stated, when I use a term like 'reason', I mean an authoritative reason; a reason that really matters to how you should live your life. The same holds of claims about value. When I say that something is 'valuable', I don't just mean that it is valuable from some particular point of view, or good according to some standard. I mean that it is genuinely valuable. I return briefly to this issue below when I discuss normative pluralism.

can provide an agent with sufficient reason to act immorally. I then discuss some distinct challenges that excellence-based reasons raise for the plausibility of moral rationalism.

Chapter Four and Chapter Five respond to two arguments that have been given for moral rationalism. Chapter Four discusses what I call the *blameworthiness defence* of moral rationalism. I reject a key premise of this argument, which claims that a person can only be morally blameworthy for freely and knowingly performing an action if they lacked sufficient reason, all things considered, to perform that action. Chapter Five concerns a claim that has motivated various philosophers to try to vindicate moral rationalism. This is that it would be bad if moral anti-rationalism were true, and that we have reason to hope that it is not. If someone is still in the grip of this thought, they may be reluctant to accept my previous arguments. I first discuss whether, even if it would be bad, this is a good reason to reject moral anti-rationalism. I then argue that, even if arguments of this type are legitimate, the idea that the truth of moral anti-rationalism is undesirable is unpersuasive. If anything, it is the idea that moral rationalism is true that is unappealing. These two chapters are standalone chapters. This is so both in the sense that they are self-contained papers that can be read on their own and in the sense that the arguments I give in these chapters don't rely on the arguments for moral anti-rationalism given in the previous chapters. My aim in these chapters is just to show that these particular arguments fail to vindicate moral rationalism.

II

Since the terms 'moral rationalism' and 'moral anti-rationalism' are used to refer to various distinct views, it is important to be clear about how I am understanding them. To start with, I do not reject the following claim, which has also been called moral rationalism:[3]

> **MR*:** If you are morally required to Φ, then you have a reason to Φ.

The view that I argue against is:

> **MR:** If you are morally required to Φ, then you have *decisive* reason, all things considered, to Φ.

[3] This is, for example, how Russ Shafer-Landau (2003, chapter 8) uses the term.

If MR* is false, then this entails that MR is false. If you lack any reason to Φ, then you lack decisive reason to Φ. The reverse is not true. It is possible to hold that an agent always has a reason to do what morality requires while claiming that, in certain cases, an agent lacks decisive reason to do what morality requires.

It is important to emphasise that MR is the claim that, if an agent is *morally required* to Φ, then they have decisive reason to Φ. It is not the claim that agents always have decisive reason to do what is *morally best*, or to do what is most strongly supported by the moral reasons that bear on the available actions. Many moral rationalists accept that agents can have sufficient reason to perform a morally inferior action so long as the action is morally permissible.[4] Of course, it might turn out that what we are morally required to do just is what is morally best, and then these claims would be equivalent, but I will not assume this to be the case. When I say, for instance, that an agent 'acts morally', what I mean is that she does what is morally required of her. An 'immoral' action is a morally impermissible action. And when I say that morality conflicts with (e.g.) prudence, what I mean is that morality requires an agent to perform an action that would be bad for her, not that the best available moral action would be bad for her. When there is the potential for confusion, I will often use the following terminology: When an agent merely does what she is morally required to do, I will say that she acts in a way that is *morally decent*. And when an agent lives her life doing only what is morally required of her, I will say that she has lived a *morally decent life*.

There are three ways of denying MR that are worth noting here. I will first discuss two forms of moral anti-rationalism. The first view, which I will defend, claims that:

> **Weak Anti-Rationalism:**[5] An agent can have sufficient reason, all things considered, to act immorally.

If an agent has sufficient reason to act immorally, then it cannot be true that they have decisive reason to act as morality requires. There is a stronger version of moral anti-rationalism which claims that:

---

[4] I discuss a number of moral rationalists who endorse this view in Chapter Two. One of the most prominent defences of the idea that agents can lack decisive reason to do what is morally best is Susan Wolf's *Moral Saints* (1982). I discuss some of her claims in the next chapter. While Wolf herself is a moral anti-rationalist – she endorses this view near the end of *Moral Saints* (435-439) and elsewhere (e.g. 1992) – many aspects of her discussion of the unattractiveness of a moral saint's life, at least as she understands this, have been taken on board by those sympathetic to moral rationalism, such as Carbonell (2009; 2013).

[5] I borrow this label from Dorsey (2012).

> **Strong Anti-Rationalism:** An agent can have decisive reason, all things considered, to act immorally.

On this view, acting as morality demands can be a mistake.[6] This stronger claim entails the weaker claim. If an agent has decisive reason to act immorally then they also have sufficient reason to act immorally. But the weaker claim could be true even if the stronger claim is false. This would be so, for instance, if the *dualism of practical reason* were true. This is (roughly) the view that (1) an agent always has sufficient reason to act either morally or prudentially, and (2) that morality and prudence sometimes conflict. On this view, an agent always has sufficient reason to act as morality demands – acting morally is never a mistake – but, when morality and prudence conflict, they also have sufficient reason to act immorally. Even though I only defend the weaker version of AR, nothing that I say rules out the stronger version.

There is a potential concern about this focus. On the face of it, the stronger version of AR may seem to be significantly more important or interesting than the weaker version. In at least some respects, this is surely right. It is one thing to learn that those who have lived morally decent lives could reasonably have lived otherwise, and another to learn that those who have lived morally decent lives have made a mistake. But even if the stronger version raises some fascinating issues that the weaker version does not, the issues that are raised by the weaker version are far from being unimportant or uninteresting. We can see this by noting that, when we are in the throes of the 'Why be moral?' question, what we are typically grappling with is whether some attractive option that we are drawn to, but that we believe to be morally impermissible, can nonetheless be justified. It is much rarer to be tempted to do what we believe to be morally required but to wonder whether *this* option can be justified. In this respect, the weak version of moral anti-rationalism captures the aspect of the 'Why be moral?' question that has the greatest existential grip on us. It also addresses the issue – whether immorality can be justified – that is likely to make us wonder about the normative significance of morality in the first place. A similar rationale can be given for the focus on MR rather than the more striking MR*. When we are in the grip of questions about the authority of morality, what usually matters to us is whether immorality is a mistake, not whether there is anything, however small, to be said in favour of acting as morality requires.

---

[6] This stronger view was recently defended by Dorsey (2016, Chp. 6). Though my argument for AR is different than his – and though, as discussed in Chapter Four, I disagree with him about how to respond to the blameworthiness defence of MR – Dorsey's work has had a significant influence on me. Strong Anti-Rationalism is also an implication of certain subjectivist and egoistic theories of reasons, which I discuss below.

There is a third way to deny moral rationalism that I mention just to put aside. Some hold that claims about what an agent has sufficient or decisive reason to do, all things considered, are incoherent, confused, or empty.[7] This view is known as *deflationary* or *normative pluralism*. According to this view, while we can make sense of the idea that, for instance, morality requires us to Φ and prudence requires us to Ψ, we cannot make sense of the idea that there is an answer to the question of what we have sufficient or decisive reason to do, all things considered, when morality and prudence conflict. To put this another way, the concept of what we ultimately or actually or just plain ought to do – of the ought *simpliciter* – is incoherent, confused, or empty. If deflationary pluralism is correct, then moral rationalism doesn't even get off the ground. This is not a unique problem for moral rationalism. If deflationary pluralism is correct, then moral anti-rationalism, as I have understood it, doesn't make sense either. When an anti-rationalist denies that an agent has decisive reason to act morally, the basis of this claim is not the thought that the very idea of an agent having decisive reason to act morally is confused. It is that the agent has sufficient reason to act in some other way; that acting immorally is *positively reasonable*. The rejection of pluralism is a commitment that these views share.

In this thesis, I will simply assume that claims about what we have sufficient or decisive reason to do, all things considered, are coherent.[8] I will also assume that we have a grasp on the plausibility of verdicts of this kind. It is intuitive, for instance, that if it is morally impermissible for an agent to physically assault an innocent stranger, but that assaulting the stranger would be best for the agent due to the calories it would burn, then the agent has sufficient reason, all things considered, to refrain from assaulting the innocent stranger.

These assumptions make a difference to how I will understand the relationship between moral rationalism and moral anti-rationalism in this thesis. I will assume that, if an agent lacks decisive reason, all things considered, to perform an action, then it follows that the agent has sufficient reason, all things considered, to perform an alternative action. It follows from this claim that, if we can show that moral rationalism is false – that there are cases where an agent lacks decisive reason, all things considered, to act as morality demands – then we will also have shown that moral anti-rationalism is true – that there are cases where an agent has sufficient reason, all things considered, to act immorally. This kind of reasoning is misguided according to normative pluralism because, on this view, any claim about what someone has

[7] See Copp (1997; 2007), Tiffany (2007) and Baker (2018b) for defences of this view.
[8] See McLeod (2001) and Dorsey (2016, 1.3) for arguments against Copp's and Tiffany's versions of normative pluralism.

sufficient or decisive reason to do, all things considered, is untrue. Because I am assuming that there is this connection between the falsity of moral rationalism and the truth of moral anti-rationalism, some of the arguments that I present for moral anti-rationalism are arguments against the claim that an agent has decisive reason to act as morality demands.

III

This section will briefly discuss three ways that moral anti-rationalism has been defended. This will be useful because it will allow me to explain how my argumentative strategy will differ from these.

SUBJECTIVISM

One well-known style of argument for moral anti-rationalism starts from a claim about our reasons for action. This is that our reasons for action are determined by, or depend upon, our desires or ends. This could be the desires that we in fact have, or the desires that we would have under ideal conditions. The most prominent version of this view is the *Humean Theory of Reasons*, which, in its simplest form, states that, for a person to have a reason to Φ, they must have some desire that would be served by them Φ-ing.[9] If we accept a theory of reasons along these lines, then we can argue for moral anti-rationalism as follows:

> **P1** If an agent has no desire that would be served or satisfied by acting as morality demands, then an agent has no reason to act as morality demands.
>
> **P2** Acting as morality demands can fail to serve or satisfy any of our desires.
>
> **C** An agent can have no reason to act as morality demands.

A classic example of this style of argument is found in Philippa Foot's (1972) *Morality as a System of Hypothetical Imperatives*.[10] Foot begins by arguing that morality is *inescapable* in the sense that its demands *apply* to an agent regardless of her desires. She (1972, 307-8) writes:

---

[9] Good discussion of the Humean Theory of Reasons include Finlay & Schroeder (2017), Schroeder (2007), Railton (2006) and Smith (2004).

[10] For some other arguments of this kind for AR, see Brink (1989, Chp. 3); Railton (1986a, 166-71; 1992*)*; and Sobel (2007a, 14-16; 2016, Chp. 1).

> When we say that a man should do something and intend a moral judgement we do not have to back up what we say by considerations about his interests or his desires; if no such connection can be found the "should" need not be withdrawn. It follows that the agent cannot rebut an assertion about what, morally speaking, he should do by showing that the action is not ancillary to his interests or desires.

In short, an agent can be morally required to Φ whether or not she has any desires that are served by her Φ-ing. This sense of inescapability, however, does not secure MR (or MR*), since it doesn't follow from the fact that some requirement applies to you that you have any reason to comply with the requirement. As Foot (1972, 308) notes, this same kind of inescapability is also found in various other domains that nobody thinks have 'automatic reason-giving force', such as the rules of etiquette, or the rules of a club. Foot then claims that, while moral requirements apply to someone regardless of their desires, it is not the case that agents (necessarily) have a reason – let alone decisive reason – to act as morality requires. Foot's (1972, 310) argument for this claim appeals to something like the Humean theory of reasons:

> 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. Nor will his action necessarily be irrational. Irrational actions are those in which a man in some way defeats his own purposes, doing what is calculated to be disadvantageous or to frustrate his ends. Immorality does not *necessarily* involve any such thing.

There is a lot to be said for arguments of this kind, but it is not how I will defend AR in this thesis. I will not assume that an agent's reasons for action depend on her desires or ends. Indeed, I will assume not only that moral requirements apply to an agent regardless of her ends, but also that they give her good reasons to act regardless of her ends. This assumption can be justified on methodological grounds. It grants important claims to the moral rationalist and makes moral anti-rationalism harder to defend.[11]

[11] This is not to say that moral rationalism is logically incompatible with the Humean Theory of Reasons. It is just difficult to plausibly square these claims. Smith (1994, Chp. 6) and Schroeder (2007, Chp. 6) are two examples of Humean's who defend MR.

RATIONAL EGOISM

Another classic argument for AR – perhaps *the* classic argument – is the argument from rational egoism. According to rational egoism, we always have decisive reason, all things considered, to act in our own best interests.[12] Since our interests may not depend on our desires – and since it is possible, in any case, to not care about our own welfare – this theory is distinct from subjectivist theories of reasons. In one simple form, the egoistic challenge to MR goes like this:

P1 If acting immorally is in an agent's best interests, then the agent has decisive reason to act immorally.

P2 Acting immorally is sometimes in an agent's best interests.

C Therefore, an agent sometimes has decisive reason to act immorally.

Arguments of this kind have a long history. In Plato's *Republic*, Glaucon tells the story of Gyges of Lydia, who was said to have found a gold ring on a corpse in a hollow bronze horse after an earthquake broke open the ground near to where he was tending his sheep. While at a monthly meeting that reported to the king about the state of the flocks, Gyges accidently discovered that, if he turned the setting of the ring towards himself, he would become invisible. Upon realising this, he arranged to become a messenger who is sent to report directly to the king, and 'when he arrived there, he seduced the king's wife, attacked the king with her help, killed him, and took over the kingdom' (Book II, 360B; 1992). Reflecting on this case, Glaucon states:

> Now, no one, it seems would be so incorruptible that he would stay on the path of justice or stay away from other people's property, when he could take whatever he wanted from the marketplace with impunity, go into people's houses and have sex with anyone he wished, kill or release from prison anyone he wished, and do all the other things that would make him like a god among humans. (Book II, 360B-360C; 1992)

This is not a mere prediction about human behaviour. What Glaucon ultimately wants is to be convinced by Socrates that, despite the prudential benefits of such injustice, a person who acted in these ways would be making a mistake. This is clear from the way he later frames the issue:

[12] For a general overview of rational egoism, see Shaver (2021, section 3).

> Indeed, every man believes that injustice is far more profitable to himself than justice. And any exponent of this argument will say he's right, for someone who didn't want to do injustice, given this sort of opportunity, and who didn't touch other people's property would be thought wretched and stupid by everyone aware of the situation. (Book II, 360D; 1992)[13]

This is a very powerful challenge to the authority of morality. At least in my own case, it is not difficult to get in the frame of mind where it seems right that Gyges is not making a mistake; where it seems that he does have sufficient reason to do the things 'that would make him like a god among humans.' How could it be irrational or unreasonable to attain everything that you have ever wanted when the opportunity to do so arises? Perhaps it is not wretched to pass up this opportunity, but it certainly doesn't seem stupid to seize it.

In certain respects, my argument for moral anti-rationalism is closer to the argument from egoism than it is to the argument from subjectivism. This is because, as I discuss below, I take prudential reasons to be genuine reasons for action. But, with the same justification as above, I will assume that P1 is false. I will assume that agents have good reason to act as morality requires even if this is bad for them.

THE DUALISM OF PRACTICAL REASON

I will discuss one more theory about our reasons for action, which I briefly mentioned above. This is the *dualism of practical reason*. According to this view, as I will understand it here, agents always have sufficient reason, all things considered, to do what is best them. But they also always have sufficient reason, all things considered, to do what morality requires.

---

[13] This is also clear from Socrates' response, which does not claim that agents *wouldn't* act as Gyges acted. Interestingly, Socrates does not respond to Glaucon's challenge by rejecting P1. He rejects P2. As I understand it, the argument is that acting as Gyges does will make you worse off because following these instincts will make you a slave to your appetites. You will lose rational self-control and corrupt your soul. He states, in Book X, that 'justice in her own nature has been shown to be best for the soul in her own nature. Let a man do what is just, whether he have the ring of Gyges or not' (Book X, 612B; 1992). Replies of this kind to the egoistic challenge also have a long history. It seems to have been common, in early modern British philosophy, to say things that at least sound like endorsements of rational egoism, but then to avoid moral anti-rationalism by denying that morality and prudence ever conflict. Joseph Butler is a well-known example. He believed that, for religious reasons, conflict between morality and prudence was 'impossible'. He also seems, at least in certain passages, to endorse rational egoism. Perhaps the most famous example is this: 'Let it be allowed, though virtue or moral rectitude does indeed consist in affection to and pursuit of what is right and good as such; yet, that when we sit down in a cool hour, we can neither justify to ourselves this or any other pursuit, till we are convinced that it will be for our happiness, or at least not contrary to it' (Sermon 11:20; 2017, 101-102). For a good discussion of Butler on rational egoism – and for a good general discussion of rational egoism from Hobbes to Sidgwick – see Shaver (1998). I largely take the idea that morality and prudence conflict for granted in this thesis, though I do offer some examples that illustrate the plausibility of this claim.

Although it is unclear whether he held the view in this form, the dualism of practical reason is most closely associated with Henry Sidgwick. In a well-known passage, he writes that:

> No doubt it was, from the point of view of the universe, reasonable to prefer the greater good to the lesser, even though the lesser good was the private happiness of the agent. Still, it seemed to me also undeniably reasonable for the individual to prefer his own. The rationality of self-regard seemed to me as undeniable as the rationality of self-sacrifice. (1907/1962, *xviii*)

If the dualism of practical reason is correct, then moral anti-rationalism can be defended with the following argument:

P1 If acting as morality demands is not best for an agent, then the agent has sufficient reason to act immorally.

P2 Acting as morality demands is not always best for an agent.

C An agent sometimes has sufficient reason to act immorally.

The dualism of practical reason is more plausible than rational egoism. This is because it maintains the most compelling aspect of egoism – the claim that it is reasonable to act in your own interests – and rejects the least compelling aspect – the claim that it is *unreasonable* to act in ways that are not in your own interests.

Unlike the previous two arguments, I believe that something close to the dualism of practical reason is true, and that, in a suitably refined form, the above argument is correct.[14] My arguments for moral anti-rationalism in this thesis, however, do not rely on the dualism being correct. While everything that I say is compatible with the claim that agents always have sufficient reason to act as morality demands, nothing that I say entails that agents never have decisive reason to act as morality demands even when doing so would be bad for them.

IV

As some of my remarks may indicate, I take the debate between moral rationalism and moral anti-rationalism to be a substantive normative debate. Not everybody sees it this way.

[14] Parfit (e.g. 2016) is an example of someone who has defended a refined form of the dualism. I do not agree with the details of his account, but I won't discuss this here. This is something that I hope to explore in future work.

According to *conceptual rationalism*, it is a non-negotiable feature of our concept of a 'moral requirement' that moral requirements provide agents with decisive reason to act. Michael Smith (1994, 87), for example, writes that 'Our concept of a moral requirement is indeed the concept of a categorical requirement of rationality or reason.'[15] Not all moral rationalists are conceptual rationalists. Sarah Stroud (1998, 170), for example, writes that 'I view the issue of morality's putative overridingness as a substantive one. There is no guarantee that simply in virtue of what it is, and regardless of our particular conception of it, morality takes precedence over other commitments.'[16] On this view, which we can call *substantive rationalism*, whether morality always provides agents with decisive reason to act is an open – although perhaps easy to answer – question. An important difference between these two brands of moral rationalism is what the upshot would be if MR could not be vindicated. If moral rationalism is a substantive claim, then, if agents can lack decisive reason to act as morality demands, it will turn out that, as AR claims, some genuinely immoral actions are reasonable. If moral rationalism is understood as a conceptual claim, then the failure of MR would put us on the road to moral error theory.[17]

Like the substantive rationalist, I believe that the denial of moral rationalism does not (necessarily) involve any conceptual confusion. The egoist who believes that morality requires self-sacrifice, but denies that he should care about this fact, may be making a mistake, but he does not seem to be confused. His mistake, if there is one, appears to be a normative mistake about how important morality is to how we should live our lives. But this is deep water that I will not explore here.[18] The main reason for this is that many of my arguments in this thesis do not turn on whether we understand moral rationalism as a conceptual or a substantive claim. Chapter One – which discusses a particular kind of non-moral reason – does not directly concern the debate between moral rationalism and moral anti-rationalism at all. Chapter Five discusses a particular style of argument for moral rationalism that is perhaps more likely to be made – and typically has been made – by substantive rationalists, but similar arguments could also be made by conceptual rationalists, and the same response could, *mutatis mutandis*, be

---

[15] Darwall is another example of a conceptual rationalist. See, for instance, (2006a 93-95).

[16] Another example of a non-conceptual rationalist is Nagel (1986, Chp. 10).

[17] Michael Smith (1994, Chp. 3.2; 2018) has some helpful discussions of the relationship between conceptual rationalism and error theory. Richard Joyce (2001, Chp. 2) is an example of an error theorist who defends error theory by arguing that MR (or, more precisely, MR*) is a conceptual commitment that cannot be vindicated. (Though Joyce focuses on MR*, the presupposition that is supposed to be problematic is 'bindingness'. This idea seems to fits better with MR than MR*, since having a decisive reason to Φ suggests that you *must* Φ in a way that merely having a reason to Φ does not.)

[18] But see Dorsey (2016; Chp. 2) for detailed and, I think, convincing responses to a number of arguments – including Smith's arguments in *The Moral Problem* – that have been given for conceptual rationalism.

made to both views. Chapter Four is similar. As I will discuss in that chapter, the premise that I reject in this argument has been found plausible by both conceptual and substantive rationalists, as well as by certain moral anti-rationalists. My argument against this, if correct, should lead anyone to reject this premise.[19] The chapters where this commitment is most apparent are chapters Two and Three. In these chapters, I offer positive arguments for the claim that an agent can have sufficient reason, all things considered, to act immorally. Any such argument presupposes that moral rationalism can be coherently denied. Even so, these arguments are still relevant to the plausibility of conceptual rationalism. This is because if, when we think it through carefully, there are cases where it seems plausible that an agent has sufficient reason to act immorally, then – just as this would give us reason to reject substantive rationalism – this would give us reason to reject the conceptual claim that our concept of a 'moral requirement' presupposes that agents always have decisive reason to act as morality demands. This is so in the exact same way that the fact that there seem to be cases of reasonable self-sacrifice gives us reason to reject not only rational egoism as a normative claim but also the claim that our concept of a 'reason' presupposes that agents have a reason to Φ only if Φ-ing would be in their best interests.

V

My focus will be on *objective* or *fact-relative* reasons for action. These reasons depend on how the world actually is or will be. Objective reasons for action can be contrasted with *subjective* reasons for action. Subjective reasons depend on an agent's beliefs – or perhaps her evidence – about how the world is or will be.[20]

[19] If the argument is correct, then we should also reject Darwall's argument for conceptual rationalism, which is based on alleged conceptual connections between immorality, blameworthiness, and an agent lacking sufficient reason to do the thing that she could be fittingly blamed for doing. Portmore, who I also discuss in this chapter, argues for moral rationalism on the basis of the same apparent connections. It is harder to say whether he is a conceptual or a substantive rationalist. In (2011, 44), he writes that one of the premises in his version of this argument – that blameworthiness entails lack of sufficient reason (BELS) – is 'not a conceptual truth.' Presumably, you could not get conceptual rationalism from an argument with non-conceptual premises. In (2014, 241), he writes that since '*moral rationalism* (MR) is a conceptual thesis, I argue for it on the bases of two other conceptual theses.' One of these other theses is BELS. I'm guessing that, between these two works, Portmore changed his mind and moved from substantive to conceptual rationalism.

[20] There is another possible category. We could also talk about reasons that depend on the objective probabilities that the world will be a certain way rather than an agent's beliefs or evidence about how the world will be. For simplicity, I am ignoring this category in this thesis.

The details here are complicated and controversial, but the basic idea is intuitive.[21] Suppose that, in apparent self-defence, Jordan kills somebody who he believes and has every reason to believe is a lethal threat. It seems plausible to say that, in some sense, Jordan has not acted wrongly. After all, it is not wrong to kill a lethal threat, and Jordan had every reason to believe that this person was a lethal threat. Suppose further, however, that this person was not in fact a lethal threat. They were merely an innocent bystander who had been perfectly set up to appear as a lethal threat. Given this, it also seems plausible to say that, in another sense, Jordan has acted wrongly. After all, it is wrong to kill an innocent person, and Jordan did kill an innocent person. This situation can be summed up by saying that Jordan had a subjective moral reason to kill this person, but no objective moral reason to kill them.

It is clear that this distinction also applies to self-regarding reasons.[22] Suppose that Juliet is deciding between two job offers. She believes and has every reason to believe that one of these jobs will bring her happiness, and that the other will bring her nothing but misery. In fact, however, the opposite is true. It seems plausible to say that, in one sense, given Juliet's beliefs and evidence, choosing the job that will in fact make her miserable is the prudent decision. In another sense, it seems plausible to say that choosing the job that will in fact make her miserable is the imprudent choice.

Since I am focusing exclusively on objective reasons, I will say about cases like these that Jordan and Juliet have *no* reason to perform the actions that they plausibly have subjective reason to perform.[23]

---

[21] There are debates, for example, about which – if any – of these notions is fundamental. There are also arguments that there are only objective reasons and requirements, or only subjective reasons and requirements. Peter Graham (2010; 2021), for example, has provided compelling arguments (to my mind, anyway) that, at least as far as morality is concerned, we should be objectivists about obligation and permissibility. For a good overview of these issues, see Sepelli (2018). My focus on objective reasons should not be objectionable to anybody except those who hold that objective reasons do not exist – or, perhaps, that they do not matter. Even if you endorse this view, however, it seems likely that there are subjective counterparts to the reasons that I discuss. Further, in all the cases that I discuss, the agents know all the relevant facts, so there is no ignorance or uncertainty involved.

[22] I will briefly discuss how this distinction applies to excellence-based reasons in the next chapter, which are neither self-regarding nor other-regarding.

[23] There may be some potential for confusion here arising from my choice of terminology. The distinction between subjective and objective reasons discussed in this section crosscuts the distinction between subjectivist and non-subjectivist theories of reasons discussed earlier. Suppose that the Humean Theory of Reasons is true. An objective reason would then be a reason to perform an action that would *actually* serve some desire. A subjective reason would be a reason to perform an action that you believe and have every reason to believe would serve some desire. These two notions can clearly come apart. For one thing, our own minds are sometimes opaque to us, and we do not always know what we desire. For another, we do not always know which actions would serve our desires. In these cases, we may have a subjective reason to ϕ, but no objective reason to ϕ.

VI

The idea that there is a tight connection between immorality and blameworthiness recurs throughout this thesis, so it is worth saying something about it upfront. I assume that, if an agent freely and knowingly fails to act as she morally ought to act, then she is blameworthy for doing so.[24] To say that an agent is blameworthy for Φ-ing is to say that certain reactive attitudes are a fitting response to her Φ-ing. Paradigmatically, these include indignation, resentment, and guilt. This is not to say that we necessarily *ought* to feel or express these attitudes in response to wrongdoing. This will depend on, for instance, whether we have standing to blame the wrongdoer, and perhaps also on considerations such as the consequences of blaming the agent. It is just to say that these attitudes would *make sense* in response to the wrongdoing. This is so in the same way that it makes sense to fear an animal you believe to be dangerous, and it doesn't make sense to fear an animal you believe to be harmless. The claim that resentment, indignation, and guilt are fitting responses to immorality is intuitive. If I spread vicious rumours about you just for kicks, it makes sense that you would feel resentment towards me, and my coming to feel guilty about these actions would be fitting. [25]

This idea plays a role in various arguments for and against moral rationalism. It is, for instance, a premise in the blameworthiness defence of moral rationalism. I also rely on this connection to help identify whether some consideration is a moral reason or requirement or a non-moral reason or requirement. To explain this, I first need to explain how I am understanding the idea that a reason is a 'moral reason'. A moral reason to Φ, as I will understand it, is a *pro tanto* moral requirement to Φ. In other words, if you have a moral reason to Φ, then, all else being equal, you are morally required to Φ. If, for example, you have a moral reason not to spread vicious rumours just for kicks, then, all else equal, it is morally impermissible for you to spread vicious rumours just for kicks. A non-moral reason to Φ, on the other hand, is a reason to Φ of

[24] Here and elsewhere, I use the phrase 'freely and knowingly fails to act as she morally ought to act' as a convenient way of saying that the agent does not have an adequate excuse for failing to act as she morally ought to act (cf. Portmore 2011, 43; Darwall 2006a, 93). It is fairly clear that, if an agent does not know that an action is wrong, this fact can prevent her from being an appropriate target of blame for performing that action. It also seems clear that she needs to have some kind of control over whether or not she performs the action. The details here are difficult. For some illuminating discussion of excusing conditions, see Pettit (2018, Chp. 6; *Forthcoming*). I have tried to set up the cases that I discuss in such a way that, on any plausible account of excuse, the agent has no valid excuse for acting immorally.

[25] These claims about blameworthiness are not only intuitive, but also widely endorsed. This way of thinking about blame and blameworthiness is usually traced back to Strawson (1962). It has been developed and defended by, among others, Wallace (1994), Darwall (2006a), and Graham (2014).

which it is not true that, all else equal, you are morally required to Φ. Unlike a moral reason to Φ, a non-moral reason to Φ will never generate a moral requirement to Φ.[26]

With this understanding in hand, we can use the connection between immorality and blameworthiness to help us identify whether a reason or requirement is moral or non-moral. Start with the claim that a reason is a non-moral reason. Suppose that we have a case where an agent has a reason to Φ, and all else is equal, but the agent freely and knowingly fails to Φ. If Φ-ing was a moral reason, then it would follow that the agent is blameworthy for failing to Φ. This is because (1) an agent is blameworthy for freely and knowingly acting immorally, (2) a moral reason is a pro tanto moral requirement, and (3) all else is equal. If the agent is *not* blameworthy for failing to Φ – if resentment, indignation, and guilt would not be fitting – then we can conclude that Φ-ing is a non-moral reason. Suppose that the agent is blameworthy for failing to Φ. We will then have strong evidence that the agent had a moral reason to Φ. As we will see in the next chapter, this is one of the methods that I use to argue that prudential and excellence-based reasons are non-moral reasons.

Based on how I am understanding moral reasons, a concern about my argumentative strategy that I gestured at above may resurface here. This is that I will move straight from the claim that there is a non-moral reason that gives an agent sufficient reason, all things considered, to act contrary to a moral reason to the claim that the agent has sufficient reason, all things considered, to act contrary to a moral requirement. As stated earlier, I am not assuming that moral anti-rationalism is vindicated if there are cases where a non-moral reason provides an agent with sufficient reason to Ψ despite the agent having most moral reason to Φ. This move fails because, even though non-moral reasons do not generate moral requirements, they can still be *morally relevant* – they can morally justify an agent in failing to perform the morally best action, or the action that is most strongly supported by moral reasons. They can, in other words, make an action that would otherwise be morally impermissible morally permissible. As Portmore (2011, 122) puts it: 'non-moral reasons can, and sometimes do, prevent moral reasons, even those with considerable moral requiring strength, from generating moral

[26] We can also understand this distinction with reference to a moral version of Joshua Gert's (e.g. 2004, Chapter 4, section 2) distinction between reasons with rational requiring strength and reasons with (mere) rational justifying strength. A reason is a moral reason if it has moral requiring strength – that is, if it can make it morally impermissible to fail to act on the reason. A reason is a non-moral reason if it lacks moral requiring strength – if it cannot make it morally impermissible to fail to act on the reason. As I discuss below, I do not assume that non-moral reasons lack moral justifying strength – I do not assume, that is, that non-moral reasons cannot make an action that would otherwise be impermissible permissible.

requirements.' More on this in later chapters. In the early chapters, I just want to establish that prudential and excellence-based reasons are non-moral reasons in the above sense: they do not generate moral requirements.

This way of carving up the conceptual space seems to me the most natural, but we could do it differently. We could say, for instance, that a reason that can make an action morally permissible but not morally required is a kind of moral reason, or at least that there are moral reasons of this kind. There is a sense, after all, in which these reasons count in favour of an action, morally speaking. They move what would otherwise be a morally impermissible action closer to being morally permissible.[27] This way of labelling the reasons would make my view more complicated to explain, but it wouldn't ultimately make a difference. My main argument for moral anti-rationalism is based on the idea that there are reasons that make a difference to how an agent should live, all things considered, but which do not make a difference to *either* whether an action is morally required or morally permissible. These are reasons that don't count in favour of an action in any sense, morally speaking. Again, more on that later.

VII

The next chapter contains a detailed discussion of excellence-based reasons. As noted, these are not the only non-moral reasons that I argue can give an agent sufficient reason, all things considered, to act immorally. I also argue that prudential reasons can make immoral actions reasonable. In contrast to excellence-based reasons, I will keep my discussion of how I am understanding prudential reasons brief. The main reason for this is that I do not intend to say anything unique or controversial. I want my claims about prudential reasons to be as ecumenical as possible so that my arguments for moral anti-rationalism based on prudential reasons have as much dialectical force as possible. There are other reasons. Prudential reasons are more widely discussed than excellence-based reasons, and many already accept that these are genuine reasons for action that can conflict with morality. This makes a long discussion of the nature of prudential reasons less interesting. It also makes it less important for the purposes of defending moral anti-rationalism. This is because, given that prudential reasons are

[27] My understanding of moral and non-moral reasons is heavily influenced by Portmore (especially 2011, pages 120-124). As far as I can tell, my claims are compatible with his. He does consider the idea (2011, 122) that there are moral reasons that can't morally require us to perform an action, but that can make an action morally permissible, but he ultimately puts this idea aside.

relatively uncontroversial and intuitive, if moral rationalism can only be vindicated by rejecting the idea that agents have reasons to promote their own interests, then that is itself a strong argument against moral rationalism. To put this differently, even if I can only vindicate the conditional claim that, if there are prudential reasons for action, then we should reject moral rationalism, I will have provided a strong argument against moral rationalism.

As I will understand it, an agent has a prudential reason to Φ if and only if, and because, Φ-ing would positively contribute to that agent's own wellbeing. The better that Φ-ing would be for the agent, the stronger her prudential reason to Φ. And if Φ-ing is better – or perhaps noticeably better – for the agent than any other available option, it would be prudent for her to Φ. If she nonetheless Ψ's, then her action is imprudent. We can have many different reasons, including moral reasons, to promote our own wellbeing. What makes prudential reasons distinct, as Worsnip (2018, 236)[28] puts it, is that they are 'distinctively and fundamentally about the promotion of the agent's own well-being.' In other words, it is the mere fact that my Φ-ing would positively contribute to my own wellbeing that fully and non-derivatively explains why I have a prudential reason to Φ. This, of course, is not to say that you cannot have prudential reasons to benefit others. But it is to say that, unless benefiting them somehow also contributes to your own wellbeing, you will not have a prudential reason to benefit them. And when you do have such a reason, it is entirely because benefiting them benefits you. In contrast to morality, the interests of others do not fundamentally matter from the point of view of prudence.

Any claim that an agent has a prudential reason to Φ presupposes a view about what is good for that agent. Many of my examples concern whether an action would contribute to an agent's happiness. I am not assuming that hedonism is correct – the view that *only* happiness and suffering make an intrinsic difference to wellbeing – since the claim that an agent's happiness is an important component of her wellbeing is compatible with many distinct theories of welfare. Even if one thinks that other things also matter – such as loving relationships, knowledge, autonomy, or even being a morally good person – most of my examples should be acceptable. This because most of my examples are extreme and simplified cases. In many of these, an agent has a choice between Ψ-ing and Φ-ing, where Ψ-ing would lead to a lifetime of happiness and fulfilment and Φ-ing would lead to a lifetime of misery and dissatisfaction.

[28] I first came across this quote in Fletcher (2021, 13).

While other things may matter to how well an agent's life goes, no plausible theory of welfare could have the implication that it is better for the agent herself to Φ in such a case.

The claim that facts about how an action would contribute to an agent's wellbeing give that agent reasons for action is not a conceptual claim. It is a normative claim. But this does not make it any less plausible.[29] It seems obvious that the mere fact that Juliet will be happy if she takes job A, but miserable if she takes job B, gives her a reason to take job A and a reason to turn down job B. Though I briefly return to this issue in Chapter Two, for the most part I just assume that prudential reasons, so understood, are genuine reasons for action. I return to prudential reasons for a different purpose in the next chapter. Here, I argue that, based on the attitudes that are fitting responses to imprudence, we can distinguish prudential reasons from both moral reasons and excellence-based reasons. This is part of my argument that excellence-based reasons are both non-moral and non-prudential reasons for action.

[29] This is to agree with Darwall (2016, 257-263). He writes that, while 'pretty much everyone agrees that an agent's good gives her reasons', this is 'a substantive normative conviction and nothing that is conceptually guaranteed.'

CHAPTER ONE: EXCELLENCE-BASED REASONS

In 1840 – when he was twenty-seven years old – Soren Kierkegaard became engaged to Regine Olsen.[30] He had been pursuing her for two years. As his journals make clear, he was, throughout this time, deeply in love with her – or at least deeply infatuated with her. In 1838, for example, he wrote:

> Thou, my heart's sovereign, 'Regine', treasured in the deepest privacy of my bosom, at the source of my most vital thought… Everywhere in the face of every maiden I see traits of thy beauty, but it seems to me as though I must have all maidens in order to extract, as it were, from all their beauty the totality of *thine*…. Thou blind god of love! ... Shall I find here on earth what I seek, shall I experience the *conclusion* of all the eccentric premises of my life, shall I clasp thee in my arms.[31]

As his journals also make clear, Kierkegaard's feelings for Olsen never diminished. He wrote about her consistently and affectionately until the day he died.

Upon her acceptance of his proposal, however, Kierkegaard began to worry that he had made a mistake. For a time, he tried to bury these feelings, and to act as if he was happy. But internally he often found himself 'debating whether I could become engaged to her – and there she was, my fiancée, beside me.' After about a year, he called off the engagement. He writes that, during this time, he 'suffered indescribably'. It seems fair to say that, for the rest of his life, Kierkegaard continued to suffer as a result of this decision.

There seem to be some strong considerations against Kierkegaard's decision. It is plausible, for example, that he would have lived a richer life if he had married Olsen. And it is almost certainly true that he would have lived a happier life.[32] But there also seem to be considerations for his decision. One such consideration weighed heavily for Kierkegaard himself. He believed that the demands of a married life would undermine his ability to become a great writer. This

---

[30] This story is drawn from the discussions of Kierkegaard's life in Hannay (1982) and Lowrie (2013). I should note, as a disclaimer, that I have simplified and somewhat idealised the details to suit my own purposes.

[31] This passage, and the other quotes, come from Kierkegaard's journal. They are quoted in Lowrie (2013, 161-166).

[32] Though an aesthete in his youth, Kierkegaard was a virtual recluse during his productive years. As a consequence, he lived a lonely life. He was also a frequent target of vicious articles from the tabloids of his day, as well as being – due to his strange physical attributes and questionable fashion choices – a constant subject of caricature. When he did leave his house, he was often openly ridiculed. The children of the town where he lived, for example, would throw stones at him while chanting 'Either/Or! Either/Or! Either/Or!'

concern does not appear to have been baseless. If he had married Olsen, it seems very unlikely that he could have lived the lifestyle that allowed him to reach the heights that he did. Following the break-up, for instance, Kierkegaard threw himself into his work with a single-minded intensity that is hard to fathom. This often involved working for sixteen hours a day. Within two years, he had drafted his first masterpiece – *Either/Or* – which comes in at over eight hundred pages. And he never really slowed down.[33] At the least, this lifestyle seems incompatible with being even a minimally decent husband.

Whether or not we think that, all things considered, Kierkegaard's decision to leave the love of his life was the right one, it is deeply intuitive that the fact – assuming it was a fact – that leaving Olsen was necessary for Kierkegaard to become a great writer counts in favour of his decision to leave. That is, that this fact gave him a genuine reason to leave.

In the following sections, my aim is to offer an account of this kind of reason. In slogan form, I will argue that people have a reason to achieve excellence in valuable activities. I will call these considerations *excellence-based reasons* (EBRs). For the purposes of illustration and manageability, I focus on the achievement of aesthetic and intellectual excellence. After presenting and defending my account, I then argue – assuming this account is correct – that EBRs are neither moral nor prudential reasons.

I

I will begin with some preliminaries.

Objective Reasons

As noted in the previous chapter, my focus will be on objective reasons for action. These reasons depend on how the world actually is or will be. This contrasts with subjective reasons for action, which depend on an agent's beliefs or evidence about how the world is or will be.

That this distinction can be applied to EBRs is clear. Suppose that Kierkegaard believed and had every reason to believe that he could become a great writer only if he broke off his engagement with Olsen. But suppose that, for one reason or another, he would in fact never

[33] Within five years, he had published – in addition to *Either/Or – Repetition*; *Fear and Trembling*; *Philosophical Fragments*; *The Concept of Anxiety*; *Stages on Life's Way; Concluding Unscientific Postscript*; and around twenty-one standalone essays.

become a great writer if he performed this action. Perhaps he would (unforeseeably) be so overcome by regret that he would not be able to write at all, or perhaps he was fated to die of consumption on his twenty-ninth birthday. It seems plausible to say – assuming that EBRs really do count in favour of actions – that Kierkegaard, in these cases, has a subjective EBR to leave Olsen, but no objective EBR to leave her. Again, since I am focused purely on EBRs in the objective sense, I will say about cases like these that Kierkegaard has *no* EBR to leave Olsen.

WHY CARE?

It might be wondered why I spend so much time discussing EBRs. This concern is particularly pressing given that, in the next chapter, I argue that moral anti-rationalism can be vindicated on the basis of prudential reasons alone. It is worth making some remarks about why the claims in this chapter matter, both in the context of the plausibility of moral anti-rationalism and more generally. Let me start with the latter.

Concerning the claim that EBRs exist, the answer is relatively obvious. If EBRs exist, they can make a difference to the kind of life that an agent can reasonably live. What may be less obvious is why anybody should care about my specific account of EBRs. After all, others have discussed what I take to be the same kind of reason. Parfit (2011, 389), for example, writes:

> On some value-based objective theories, there are some things that are worth doing, and some other aims that are worth achieving, in ways that do not depend, or depend only, on their contributions to anyone's well-being. Scanlon's examples are 'friendship, other valuable personal relations, and the achievement of various forms of excellence, such as in art or science.' These we can call *perfectionist* aims. On such views, it would be in itself good in the… reason-implying sense if we and others had these valuable personal relations, and achieved these other forms of excellence.

And Nagel (1979, 129-30) – during an examination of different kinds of value – writes:

> The fourth category is that of perfectionist ends or values. By this I mean the intrinsic value of certain achievements or creations, apart from their value *to* individuals who experience or use them. Examples are provided by the intrinsic value of scientific discovery, of artistic creation, of space exploration, perhaps. These pursuits do of course serve the interests of the individuals directly involved

> in them, and of certain spectators. But typically the pursuit of such ends is not justified solely in terms of those interests. They are thought to have an intrinsic value…. [M]any things people do cannot be justified or understood without taking into account such perfectionist values.[34]

In addition, various writers have discussed considerations that are at least in the same ballpark as EBRs. In *Moral Saints*, for example, Susan Wolf (1982) argues that we have reasons to cultivate 'personal excellence'. Another example is Bernard Williams' (1981a) famous discussion of Gauguin's decision to abandon his family – the consequences of which will be 'grim' for them – to sail to Tahiti and realise his potential as a painter. According to Williams (1981a, 23), if Gauguin fails as a painter, then he has 'no basis for the thought that he was justified in acting as he did. If he succeeds, he does have a basis for that thought.' On one natural interpretation, what would give Gauguin the basis for this thought is that, if he succeeds, then he has achieved aesthetic excellence, and, if he doesn't, then he hasn't.[35] These writers claims are close to mine in another respect. Both Wolf and Williams believe that these

---

[34] See also Scanlon (1998, chapter 2). Note that both Nagel and Parfit use the term 'perfectionist' to refer to these kinds of considerations. I do not use this terminology because I do not want EBRs to be associated with perfectionist theories of, or claims about, morality and prudence (for discussions of these views, see Dorsey 2010; Foot 2001; Haybron 2007; Hurka 1993; and Wall 2021). There are two reasons for this. First, as noted, I will later argue that EBRs are non-moral and non-prudential. Second, these views are generally – though not always – grounded in a claim about developing our natural or essential human capacities. This is not how I ground, or understand, EBRs. It is also very likely – as will hopefully become clear throughout the paper – that EBRs and perfectionist reasons, in this sense, will come apart. A person could have *no* EBR to develop her natural or essential capacities, and she could have an EBR *not* to develop certain (otherwise important) natural or essential capacities. There is no doubt, however, that perfectionists have often been motivated by the same kinds of considerations that motivate EBRs. Indeed, some, such as Rawls (1971, 325), define perfectionism simply as the view that we ought to maximise the 'achievement of human excellence in art, science and culture'. If perfectionism is understood in this way – and without reference to essential capacities or human nature – then EBRs and perfectionist reasons will likely coincide. As such, we are not that far apart. In my view, these writers are right to recognise that these considerations are important, but wrong to think that, since they are important, they must be moral or prudential. They are also wrong to think that, since they are important and not recognised as such by other moral and prudential theories, they provide counter-examples to those theories, or reasons to prefer perfectionism over those theories.

[35] Though Williams and I agree that whether Gauguin succeeds in reaching his potential as a painter makes a significant difference to whether his decision was justified, Williams would probably reject EBRs as I understand them. This is because of Williams' view about the nature of our reasons for action (see 1981b and 1995; see also Shafer-Landau 2003, Chp. 7). On his view, an agent cannot have a reason to Φ unless they are already motivated to Φ, or would be motivated to Φ if they went through a process of sound deliberation starting from their existing motivations. Since Gauguin, as Williams imagines him, is motivated by the desire to realise his gifts as a painter, his justification depends on whether he succeeds in realising these gifts. But there is a sense in which this is so only because this happens to be the content of his desire. If Gauguin had the potential to paint aesthetically excellent paintings, but couldn't be motivated to pursue this aim, then he would not have a reason to create aesthetically excellent paintings. Given this, it not, on this view, the aesthetic excellence of the work itself that grounds his reasons to sail to Tahiti and realise his gifts. On my view, it is the excellence of the work itself that gives Gauguin a reason to paint, and this is so regardless of whether he is or can be motivated to paint. As I discussed in the previous chapter, I do not assume in this thesis that our reasons for action depend on our desires.

considerations are normatively significant non-moral reasons that create problems for moral rationalism.

The first thing to note is that, while others – like Parfit and Nagel – have discussed what I take to be the same kind of reason – and while these discussions have been suggestive and compelling – their claims have tended to be vague and schematic. In this chapter, I hope to offer and defend a more detailed account that unambiguously tells us what an EBR is, when a person has one, and why they have it. Second, even though various writers – like Wolf and Williams – have discussed reasons that are close to EBRs, there are important differences between my view and many of these other views. Further, as I will argue below, my account has various advantages over these other claims.

Suppose we accept that there are excellence-based reasons. We might still wonder why we should care whether these reasons are non-moral or non-prudential. When it comes to living as we ought to live, a reason is a reason. Note first that, while it seems basically right that, at the time of acting, a reason is a reason, this claim does not show that whether EBRs are non-moral or non-prudential makes no difference to how we should live. For one thing, whether an agent was guided by moral or non-moral reasons seems to make a difference to how we should treat them after they have acted. It seems plausible, for example, that we should be less inclined to end a friendship with a person who failed to show up to our wedding for moral reasons – for instance, because they were helping someone in need – than for non-moral reasons – for instance, because they recognised that they would derive more pleasure from watching a *Law & Order* marathon than from attending the wedding. Note also that, even if reasons being moral or non-moral made no difference to how we should act, this would not show that this question is insignificant. This is because, as I will discuss later in this chapter, whether a person acts contrary to a moral or a prudential or an excellence-based reason makes a significant difference to what it would be appropriate for us to feel about that person, and to what it would be fitting for that person to feel about themselves.

Turning to the debate between AR and MR in particular, whether EBRs are non-prudential reasons is relevant to the plausibility of moral rationalism. As I argue in Chapter Three, EBRs create distinct and compelling problems for moral rationalism that prudential reasons do not. Some may be moved to reject MR due to these problems even if they are not moved to reject it due to the problems that prudential reasons create for MR. And even if one is convinced to

reject MR by the arguments from prudential reasons – as I hope they will be – reflecting on EBRs can still provide additional reasons to reject MR; it can strengthen the case against MR.

II

As noted, I claim that there are reasons to achieve excellence in valuable activities. In this section, my aim is to clearly spell out how I understand these reasons. I will first discuss what I mean by 'achieving excellence'. I will then discuss what I mean by 'valuable activities'. Finally, I will explain the sense in which these reasons are reasons for action.

ACHIEVING EXCELLENCE

When I say 'achieving *excellence*', I am referring to accomplishments of the highest standard. It seems clear, for example, that David Hume was an excellent philosopher, and that Henry James was an excellent novelist. More generally, it seems clear that Hume achieved intellectual excellence, and that James achieved aesthetic excellence. Other philosophers and novelists, even good ones, did not achieve excellence in this sense; they did not reach the same heights.[36]

This understanding of excellence commits me to certain claims. For one, I am committed to the claim that certain aesthetic and intellectual achievements are *better than* others. I assume that, when we compare two scientific monographs, or two paintings, there is some kind of criterion that allows us to truly say that one is better than the other. There will, of course, be many borderline cases. But there also seem to be many clear cases. It seems obvious, for example, that *Nighthawks* is aesthetically superior to a painting of a stick figure done by a five-year-old, and that *On the Origin of Species* is intellectually superior to the average high school biology paper. Similarly, I assume that we can truly claim that certain aesthetic and intellectual works are *good*, and that others are *bad*.

It is worth saying something about what I mean by 'aesthetic' and 'intellectual'. The first thing to note is that these categories are fairly artificial. They are not designed to carve nature at its

[36] This is not to deny that those who produce good, but not great, work have reasons to do so, or even that these reasons can be explained in a similar way to how I explain reasons to produce excellent work. This may be so. In any case, focusing on the highest achievements can be justified pragmatically. This is because I am ultimately interested in considerations that can potentially counterbalance moral requirements. Given this focus, it makes sense to concentrate on the strongest versions of the non-moral reasons that we can have. This is simply because the stronger the reason, the more likely it is to counterbalance other reasons. And, as should become clear in this chapter, we have stronger reasons to produce better work than we do to produce worse work, and hence stronger reasons to produce excellent work than non-excellent work.

joints, or anything like that. They are merely a somewhat descriptive shorthand for certain kinds of works. By 'aesthetic' excellence, I have in mind achievements in various art forms. This includes, among other things, great novels, plays, films, paintings, photographs, and musical compositions. My use of 'intellectual' is perhaps even more diverse. I have in mind achievements in, for example, philosophy, biology, history, and mathematics.

One thing that aesthetic and intellectual excellence have in common, I believe, is that both are valuable. But I do not claim that they are valuable for the exact same reasons. It seems highly plausible that what made Emily Dickinson's work excellent is not identical to what made Marie Curie's work excellent.

Note also that I use the terms 'aesthetic value' and 'intellectual value' ecumenically. They refer to *whatever* makes it the case that artworks *qua* artworks are valuable, and whatever makes it the case that intellectual works *qua* intellectual works are valuable. There are some natural candidates. It might be, for example, that something is aesthetically valuable if it is beautiful, and intellectually valuable if it is true. But these claims can be denied. For one thing, they may be too narrow. Regarding aesthetic value, Gardner (2003, 236) writes:

> [T]he concept of beauty… has undergone a dramatic reversal of fortunes in the history of aesthetics. Classical aesthetics took it for granted that beauty is the only, or at least the fundamental, aesthetic quality. Some modern writers propose by contrast that "beautiful" is merely a catch-all term, roughly equivalent to "aesthetically commendable", and that there is, as ordinary language implies, a limitless plurality of aesthetic qualities, encompassing elegance, grace, poignancy and so on.[37]

Similar claims can be made about intellectual value. When we consider what makes an excellent philosophical work excellent, for instance, truth seems insufficient and, arguably, unnecessary. During his discussion of Marx's work, Jonathan Wolff (2002, 101) writes that:

> [W]e value the work of the greatest philosophers for their power, rigour, depth, inventiveness, insight, originality, systematic vision, and, no doubt, other virtues too. Truth, or at least the whole truth and nothing but they truth, seems way down the list…. To put it bluntly there are things much more interesting than truth.

[37] For a nice overview of aesthetic value, and of general aesthetic reasons for action, see King (2022). For a discussion focused more squarely on aesthetic reasons, see McGonigal (2018).

As far as I can tell, nothing that I say relies on any particular conception of aesthetic or intellectual value being correct.

I should also point out that, though I focus on these two broad categories of excellence, I do not take them to be exhaustive. I believe there are also reasons to achieve, among other things, athletic excellence.

Note next that, on my view, excellence-based reasons are reasons to *achieve* excellence. This is a success condition. If an action did not – or would not – lead to excellence, then you did not – or do not – have an excellence-based reason to perform that action. Suppose that John has a deep desire to be a theoretical physicist. In pursuit of this goal, he spends night after night working tirelessly on a paper. But suppose also that John is terrible at mathematics. As a result of this deficiency, the paper that he ends up producing is thoroughly mediocre. It may be true that John had reasons to spend his time as he did, but he did not have excellence-based reasons to do so. This is because he did not actually produce anything excellent.

As this may suggest, I take excellence-based reasons to be *teleological*. They are reasons to perform actions that bring about a particular end. In this case, excellent aesthetic and intellectual works.

VALUABLE ACTIVITIES

I now turn to the 'valuable activities' condition. On my account, what explains why we have reasons to bring about this end, and to create these works, is that such works are *valuable*. I shall now clarify this.

To begin with, when I say that intellectual and aesthetic excellence are valuable, I mean that they are intrinsically valuable, or valuable for their own sake.[38] This claim is difficult to argue for, but it is worth saying something in its defence. This is that it is very doubtful that instrumental value alone can fully account for the value that we often take these works to have. For one thing, the instrumental value of some excellent works is far from obvious. It is hard to

[38] These two phrases are often treated as synonymous. A number of philosophers have argued, however, that they actually express two different concepts. Roughly, the idea is that something is only intrinsically valuable if it is valuable solely in virtue of its intrinsic properties, whereas something can be valuable for its own sake even if it is valuable partly, or even wholly, in virtue of relational or extrinsic properties. For discussion of these issues, see Korsgaard (1983), Kagan (1998), and Langton (2007). When it comes to the value of great novels, or great philosophical works, it is the latter sense that seems most apt. One reason for this is that it is intuitive that some of the properties that make an aesthetic or intellectual work valuable *qua* aesthetic or intellectual work are relational. An example is *originality*. I will, however, continue to these terms interchangeably, since nothing turns on this issue as far as my claims go.

believe, for example, that the existence of certain highly theoretical mathematical proofs makes any real difference to anything. At the least, it is hard to believe that their existence makes any significant difference. Yet, they seem to be significantly valuable. Hence, there is a mismatch between their instrumental value and their overall value. These works being intrinsically valuable explains this mismatch.

Additionally, the existence of excellent aesthetic and intellectual works can have instrumental costs. It is not obvious that these costs will always be outweighed. Even if they are not, this does not seem to strip these works of all their value. Let me give just one example. If you have the desire to achieve excellence, but lack the talent to do so, the existence of excellent works can be extremely depressing. When I was younger, I wanted to be a novelist. I distinctly remember, after reading John Fowles' *The Magus*, feeling dejected and disillusioned. I felt that, no matter how hard I tried, I would never be able to produce anything *that* good. Looking over my own previous attempts, all of the words suddenly seemed dead on the page. I have had many similar experiences reading philosophy, as have others. Discussing Leibniz, Diderot writes:

> Perhaps never has a man read as much, studied as much, meditated more, and written more than Leibniz… What he has composed on the world, God, nature, and the soul is of the most sublime eloquence…. When one compares the talents one has with those of a Leibniz, one is tempted to throw away one's books and go die quietly in the dark of some forgotten corner.[39]

Suppose that, whenever somebody read *The Magus* or the *Theodicy*, it had this sort of effect on them. This would likely make these works instrumentally disvaluable. Nonetheless, it still seems that they would have value. If this is right, then the value of excellent works cannot be purely instrumental.

The 'valuable activity' rider serves a further purpose. It allows us to explain why we do not always have reasons to achieve excellence. That this is plausible can be made vivid with an example. During the 1980s and 1990s, Gary Ridgway – aka the Green River Killer – went on a two-decade killing spree. Many of his victims were runaways and prostitutes, which made them particularly vulnerable. He would begin by luring his intended victim into his car. To gain their trust, and to come across as a nice guy, he would often show them pictures of him playing

[39] This passage is from Diderot's *Encyclopedia* article on Leibniz. It is quoted in Look (2020).

with his son. Once they were in his car, he would have sex with the victim and then strangle them from behind using either ligatures or his bare hands. Following this, he would dump their body in the woods surrounding the Green River in Washington – hence the name. In the days that followed, he would often return to the body and have sex with it again.[40]

Ridgway was an excellent serial killer, for at least three reasons. First, he was prolific. There are forty-nine confirmed victims. This makes him one of the most prolific American serial killers in history – at least according to confirmed numbers. It is widely believed – based on Ridgway's confessions and other evidence – that he actually murdered around seventy people. Second, he evaded the police for roughly twenty years. This included various task forces that were set up solely to figure out his identity. Finally, he is notorious. Ridgway is known and studied the world over. Many people have spent significant portions of their lives trying to understand him, and to come to terms with his crimes.

Though Ridgway was an excellent serial killer, this fact does not seem to count in favour of his actions. In other words, the mere fact that he was an excellent serial killer seems to do nothing to justify his actions. My view can explain this. Since murdering innocent people is not a valuable activity, Ridgway had no excellence-based reasons to be an excellent serial killer. Just to be clear, I am not here denying that Ridgway had reasons to murder his victims. This claim is not that counterintuitive. If we imagine him as having a certain sort of psychology, then it is plausible that he could have had prudential reasons to do what he did. The counterintuitive implication of claiming that Ridgway had excellence-based reasons to slay his victims is the implication that, all else equal, a better serial killer has more reason to slay their victims than a worse serial killer, simply in virtue of being better. That is hard to believe.

My view can, in the same way, explain other kinds of cases where a person has no reason to achieve excellence. Aside from disvaluable activities, it seems plausible that there are activities that simply *lack* value. This may be true, to use a classic example, of being an excellent grass

[40] For a comprehensive overview of Ridgway's life and crimes, see Anne Rule's (2004) *Green River, Running Red*.

counter. If counting grass has no value, then there are no excellence-based reasons to be an excellent grass counter.[41]

Reasons for Action

Thus far, I have focused on what I mean by 'excellence' and 'valuable'. I shall now say a bit more about the sense in which EBRs are reasons for action.

Above, I claimed that excellence-based reasons are teleological. As Portmore (2011, 79) writes:

> A teleological reason to φ is a reason to φ in virtue of the fact that φ-ing would either itself promote a certain end or is appropriately related to something else that would promote that end.

The relevant end in this case is aesthetically and intellectually excellent works. And what justifies our acting in ways that promote this end is that aesthetically and intellectually excellent works are non-instrumentally valuable. An excellence-based reason for action, then, is a reason to perform some action that leads to, contributes to, or promotes in some way the creation of intellectually or aesthetically excellent works. And an excellence-based reason against performing some action is that it detracts from, undercuts, or hinders in some way the creation of intellectually or aesthetically excellent works.

That, in general, certain actions can contribute to, or detract from, the achievement of excellence is obvious. If the next potential F. Scott Fitzgerald had an existential crisis and jumped off a bridge, he would not write any great novels. This fact would give him an excellence-based reason not to jump off a bridge.

---

[41] Note that these claims about value are not just the claims that grass-counting and serial killing lack moral value, or morally relevant value. These are supposed to be examples of activities that don't produce anything at all of genuine (see fn. 2) non-instrumental value. This could be denied. It could be argued, for instance, that serial killing or grass counting have, or can have, aesthetic or intellectual value. There do seem to be people who have an aesthetic interest in murder and death. In fiction, this is true of some of Thomas Harris' characters, such as the Toothy Fairy from *Red Dragon*. And Edgar Allan Poe famously wrote that 'The death of a beautiful woman is, unquestionably, the most poetical topic in the world' (2006, 548). If something along these lines is correct, then it could turn out that a serial killer like Ridgeway had EBRs to slay his victims that are explained in the same way as James' EBRs to write *The Golden Bowl*. How plausible this claim is depends on how plausible it is that murder can have aesthetic value. It could also be argued – without reference to other commonly recognised categories of value (e.g., aesthetic value) – that, when we reflect on grass counting or serial killing, these activities just seem genuinely valuable for their own sakes. Since EBRs can be accepted even if one rejects the specific examples of disvaluable and neutral activities that I have given, I will not attempt to argue against these views. I take it that, at the least, my claims are intuitive. The world does not seem better because Gary Ridgway was an excellent serial killer. It seems worse. And, though the world may not be worse because people undertake trivial activities, it does not seem better either.

We can be more specific. Some facts count for or against an action in a fairly direct manner. In order to write an excellent novel, a person has to sit down and put words on a page. This action directly contributes to the creation of an excellent aesthetic work. The fact that performing this action would lead to producing an excellent novel is an excellence-based reason to perform that action. A guaranteed way to fail to produce an excellent novel, on the other hand, is to spend all your designated writing time playing solitaire. The fact that this action would prevent a person from producing an excellent novel is an excellence-based reason for them not to spend their time in this way.

Other cases are less direct. Certain lifestyles, for instance, are likely to be more conducive to producing excellent work than others – at least for certain people. For some, such as Kierkegaard, the creation of excellent work may require a monkish existence that allows for single-minded focus. Others may require a more balanced lifestyle. And still others, like some of the romantic poets, seem to benefit from a life of hedonistic excess infused with occasional, though intense, bursts of productivity. A person will have an EBR to live whatever lifestyle leads them to produce excellent work, and they will have an EBR not to live any lifestyle that would prevent them from producing excellent work.

What these different examples indicate is that we can evaluate pretty much anything from an excellence-based perspective. This is because we can always ask – be it of a discreet action, a job, a relationship, a lifestyle choice, or anything else – whether something will contribute to, or detract from, the achievement of excellence.

III

I shall now contrast excellence-based reasons, as I understand them, with a few similar proposals. As well as helping to avoid potential confusion, this will bring out some of the advantages of my account.

To start with, we can distinguish excellence-based reasons from two views mentioned earlier. In *Moral Saints* (1982), Susan Wolf argues that always doing what is morally best is undesirable because it would undermine ‘personal excellence’. This claim suggests that, on Wolf’s view, we have reasons to cultivate personal excellence.

Though this claim may seem similar to mine, Wolf has something quite different in mind. She is worried about morality interfering with a person living ‘a healthy, well-rounded, richly

developed' life. This might include 'reading Victorian novels, playing the oboe, or improving [one's] backhand' (1982, 421).

That these are not excellence-based reasons can be seen by considering conflicts between the two. As with the Kierkegaard example above, achieving excellence for some people may require single-minded obsession. Living in this way would not be excellent in Wolf's sense – indeed, a person who lived such a life would have much in common with the moral saints that she finds so unappealing – but there may be excellence-based reasons to live in this way. That is, there could be excellence-based reasons *not* to live 'a healthy, well-rounded, richly developed' life.

Bernard Williams (1973; 1981) worries that obeying the demands of morality could undermine a person's ability to wholeheartedly commit to her personal projects. Without such commitments, Williams suggests, life would hardly be worth living. He (1981c, 12) writes: '[M]y present projects are the condition of my existence, in the sense that unless I am propelled forward by the conatus of desire, project and interest, it is unclear why I should go on at all.'

The reasons that we have as a result of our projects may sometimes coincide with excellence-based reasons. This is because, as the quote below indicates, a person's projects may include – as in the Gauguin example – creating excellent aesthetic or intellectual works, and achieving excellence in various other ways. But, as with Wolf's proposal, they can come apart. This is due to the fact that 'projects', for Williams, is an extremely broad category. Some examples include:

> The obvious kind of desire for things for oneself, one's family, one's friends, including the basic necessities of life, and in more relaxed circumstances, objects of taste. Or there may be pursuits or interests of an intellectual, cultural or creative character…. Beyond these, someone may have projects connected with support of some cause…. Or there may be projects which flow from some more general disposition towards human conduct and character. (1973, 110-11)

A commitment to some of these projects may well undermine, or detract from, a person's ability to achieve excellence. This could be the case if the project is exceedingly time-consuming, as some of the examples above are likely to be. Whenever a person's present projects compromise their ability to achieve excellence, there will be excellence-based reasons to abandon those projects. And this will be so regardless of how much these projects mean to that person.

More generally, excellence-based reasons can be distinguished from three other potential kinds of reasons. First, it is sometimes argued that *achievement* is itself intrinsically valuable.[42] My claim is not about achievement for its own sake. It is about a certain level of achievement. Completing a novel is an achievement, but, if the novel is terrible, there was no excellence-based reason to complete it.[43]

A second claim that could be made is that we have reasons to merely *try* to produce excellent work. It might also be argued that engaging in aesthetic and intellectual activities is valuable for its own sake. Both of these views are distinguishable from excellence-based reasons. If I spend my whole life studying and trying to create great films, then I have both tried to produce excellent work and engaged in aesthetic activity. But if all my work is derivative and sophomoric, then I had no excellence-based reasons to perform these actions.

I have tried to show that excellence-based reasons are distinct from some similar proposals. Even if this is right, it does not show that we should endorse excellence-based reasons either in addition to, or instead of, these other claims. I shall now suggest that, even if we accept these other proposals, we should also accept excellence-based reasons. I will also note some advantages that excellence-based reasons have over these other proposals.

The first thing to note is that these other conceptions fail to accommodate a powerful source of justification that excellence-based reasons captures. Suppose that two mathematicians – Madeline and Elizabeth – commit numerous immoral acts in their quests to achieve intellectual excellence. Suppose further that both commit identical immoral acts, and that most of their other actions are also identical. The only difference between them is that Madeline discovered a proof as powerful and elegant as Gödel's incompleteness theorem, whereas Elizabeth came up with a minor result that will quickly be forgotten. Now imagine that both are called to account for their immoral actions. Both can say various things in their own defence. For

[42] See, for example, Bradford (2013; 2015).

[43] As Bradford agrees (e.g. 2013, 205-210), the greatest achievements are not necessarily those that produce the most independent value. When we are evaluating how significant an achievement is, *difficulty* is one of the most important considerations. Consider two people. One must work tirelessly to play the violin well, and the other has prodigious natural talent. The one who works hard becomes a good, but not great, concert violinist. The other becomes an excellent concert violinist with very little effort. On the face of it, it seems that – due to the effort required – becoming a concert violinist is a greater achievement for the first person than it is for the second. But it is the second person who, on my account, has an EBR to become a concert violinist. This is explained by the greater aesthetic value of her work. As I discuss further below, it seems to me that, even if we want to say that there are achievement-based reasons, we should also say that there are reasons based on this aesthetic value.

example, both can cite all the considerations stated above: they both achieved something; they both tried to achieve something excellent; and they both were involved in intellectual activities. But Madeline can say something that Elizabeth cannot: *She actually discovered an important proof*. This seems like a significant additional consideration in favour of her actions. Excellence-based reasons captures this, whereas the other proposals do not.

Perhaps for similar reasons, excellence-based reasons allow us to explain certain cases that these other proposals cannot explain. Consider:

> Three people – Andrew, Blake, and Cameron – want to be writers. They are each the primary breadwinner for their young families. All three decide that they want to wholeheartedly pursue their ambitions. All believe that they cannot do this while caring for their children. As a result, each decides to abandon his family.
>
> Many years pass, and all three eventually die. At the end of their lives, this is what each has achieved: Andrew chased his dream in a disciplined and serious manner, and as a result he produced a substantial body of work. It is all garbage. Blake never really got going. Like most people who say they want to be a writer, he started pieces here and there but never got close to finishing anything. Finally, like Andrew, Cameron went about his work in a disciplined manner, and he produced a substantial body of work. Unlike Andrew, Cameron's writing is on the same level as Kafka's.

Why is it that, of these three people, Cameron's decision to pursue his ambition seems the most reasonable? My view explains this. It is because his actions actually led to the creation of excellent work. As such, he had reasons to do what he did that neither Andrew nor Blake had.[44]

The three alternative views discussed above can explain certain aspects of this case, but not others. Note first that it seems plausible that Andrew's decision is more reasonable than Blake's. He at least *tried*, and he at least achieved *something*. Blake, on the other hand, ultimately abandoned his family *for nothing*. His decision was terrible in every respect. All three views can arguably explain this difference. Andrew tried to produce excellent works, he achieved something, and he was involved in aesthetic activities. None of these things are true

[44] It is worth noting that Cameron's situation is very similar to Gauguin's situation in Williams' (1981a) example.

of Blake. This perhaps supports the claim that the considerations these proposals highlight are genuine reasons.

What these views do not explain, however, is why there is more to be said for Cameron's decision than for Andrew's. Both, after all, tried to achieve aesthetic excellence, both were deeply involved in aesthetic activities, and both achieved something by producing a substantial body of work.

These considerations, I hope, support the claim that it is necessary to postulate excellence-based reasons, even if there are also these other kinds of reasons. They also support another claim. This is that EBRs are *stronger than* these other reasons. Other considerations support this claim. One is that excellence-based reasons seem to fare better than these other potential reasons when they are weighed against paradigmatic moral and prudential considerations. Hopefully the mathematician case above suggests that this is so for moral considerations. Similar examples suggest that this is also the case for prudential considerations. Consider a person who, like Kierkegaard, sacrificed significant happiness and a loving relationship to become a great writer. Unlike Kierkegaard, however, he failed. It seems fairly clear that this person's attempt to become a great writer was a mistake. The fact that he achieved something; that he tried to produce something great; and that he engaged in aesthetic and intellectual activities seems to do little to counterbalance the costs. They would, I imagine, be cold comfort for the agent. The fact that Kierkegaard *actually wrote Either/Or*, on the other hand, seems to go a long way towards justifying his actions, and counter-balancing the costs. Even though both ended up unhappy as a result of their decisions, it is far less clear that Kierkegaard made a mistake.

A final consideration in favour of excellence-based reasons over these other proposals is worth noting. Aside from seeming stronger, the claim that we have excellence-based reasons is more intuitive than these other claims. It seems clear, for example, that Henry James and David Hume achieved valuable things. It also seems clear that they had reasons to produce the work that they did. It is far less clear that a person has a reason to try to produce something great if

they are just going to fail anyway, or that a person has a reason to achieve their goals if the results will be thoroughly mediocre.[45]

IV

I will now consider three possible objections to my account.

Distinguishing Different Kinds of Excellence

I have claimed that people have reasons to achieve excellence in certain activities. I have also claimed that, in other activities, people do not have reasons to achieve excellence. This was my claim about serial killers such as the Green River Killer, and also about those who perform certain trivial activities, such as the Grass Counter. Although I claim that these people lack excellence-based reasons, I do not deny that excellence can be achieved in these activities. On the face of it, this may seem like an unstable set of claims, and it might be wondered whether I am entitled to them.

In response to this concern, I will lay out more carefully how my account of EBRs allows us to non-arbitrarily distinguish between excellence in different activities. As I understand them, an excellence-based reason to φ is a teleological reason to bring about a certain kind of value. These reasons are teleological in the sense that what *explains* why a person has an EBR to φ is that φ-ing will lead to, promote, or in some way contribute to the creation of value. A mathematician has an EBR to develop her talent if doing so would lead her to produce excellent mathematics. And this is so *because* excellent mathematics is valuable for its own sake.

With this in mind, the reason that both the Green River Killer and the Grass Counter do not have EBRs – even though both are excellent at what they do – is relatively straightforward. On my account, the mere fact that performing some action would lead to the achievement of excellence is not enough to generate an EBR to perform that action. In addition, it is also

[45] Note that the same considerations also show that EBRs are distinct from alternatives other than those that I have focused on above. It might be claimed, for instance, that we have reasons to live up to our potential, or to develop our talents, or to develop our natural or essential human capacities. These reasons are unlikely to be extensionally equivalent with EBRs. An agent can live up to their potential, or develop their essential capacities, without achieving anything excellent. Such a person would not have an EBR to reach their potential, or to develop their essential capacities. Even if they were extensionally equivalent, these proposals would fail to capture the source of justification that is captured by EBRs. Kierkegaard may have had reason to leave Olsen if this allowed him to live up to his potential, or to develop his talents, but the fact that this was necessary for him to write *Either/Or* seems to itself count in favour of this decision.

necessary that the products of this excellence be valuable. In some cases, the products of excellence are not valuable. In these cases, a person does not have an excellence-based reason to achieve excellence. And I submit that the outcomes of both excellent serial killing and excellent grass counting are not valuable. If this is right, then neither the Green River Killer nor the Grass Counter have EBRs. This differentiates them from an excellent mathematician.

This claim is not arbitrary. It follows naturally from endorsing a value-based teleological account of reasons to achieve excellence. To maintain that this is arbitrary would be like maintaining that it is arbitrary for a utilitarian to claim that people ought to maximize pleasure rather than that people ought to maximize pleasure *and pain*. There may be a lot of things wrong with utilitarianism, but this is not one of them.

WHAT ABOUT THE VALUE OF EXCELLENCE ITSELF?

There is a clear sense in which excellence does not *itself* explain why people have excellence-based reasons. This may lead to a concern. This is that my account neglects an important source of value, which is excellence itself. It may also seem that, on the face of it, taking excellence itself to be intrinsically valuable would provide the simplest explanation for why people like Kierkegaard have reasons to achieve excellence.

It is worth stating first that I am not committed to denying that excellence is valuable for its own sake. But note that this claim really does seem to be open to counter-examples like Gary Ridgway. Now, it could of course be denied that Ridgway was excellent, or that any serial killer could be excellent. But this seems both arbitrary and wrong. If my account can avoid these counter-examples in a non-arbitrary manner – while still plausibly explaining the cases that we want to explain – then, all else equal, we should prefer my account to this one.

It might be worth emphasising at this point that, even if we deny that excellence is itself valuable or reason generating – even if we claim that agents can lack EBRs to achieve excellence – the connection between excellence and EBRs remains extremely tight in the areas where we do want to say that people have reasons to achieve excellence. This is certainly the case when it comes to the creation of significant aesthetic and intellectual value. There would not be excellent novels, excellent philosophical treatise, or excellent mathematics without excellent novelists, excellent philosophers, and excellent mathematicians. In these areas, it is also the case that the value of the best work *far exceeds* the value of mediocre work. The world would lose vastly more intellectual value if all of David Hume's works were incinerated than if all of David Icke's works were incinerated. Given this, even though aesthetic and intellectual

value, rather than excellence itself, explains why a person has an EBR, it is still the case that the particular people who have EBRs have them because they have the potential to be excellent.

Note also that, unlike some other kinds of value, aesthetic and intellectual value do not seem to be straightforwardly additive. It is at least plausible that there is more hedonic value in the world if thousands of people live slightly pleasurable lives than if one person lives a blissful life. But it is not plausible that there is more intellectual value in the world if it contains thousands of slightly above average undergraduate essays rather than *A Treatise of Human Nature*. Nor is it plausible that there is more aesthetic value if there are thousands of average remainder-bin novels in the world rather than *The Golden Bowl*. For this reason, those with the potential to be excellent are also *uniquely* placed to bring about significant intellectual and aesthetic value.

A Potential Counter-Example

I have claimed that a person has an EBR to φ if φ-ing would lead to the creation of excellent aesthetic or intellectual works. This account may seem to be open to a certain sort of counter-example. Consider a playwright – Emily – who makes substantial sacrifices to her own welfare to focus wholeheartedly on her art. Suppose that, as a result of these sacrifices, she is able to write a play that has all the aesthetic qualities of *King Lear*. But suppose also that, due to a crippling fear of rejection, Emily never shows this play to anybody. It sits in a chest in the attic for years after her death, and it is eventually incinerated when a fire burns down the house in which she lived.

On the face of it, it seems that this play has significant aesthetic value. If this is right, then my account suggests that Emily had a reason to write this play, and a reason to sacrifice her welfare. But it might seem that Emily actually had *no* reason to perform these actions. After all, neither she nor anybody else ever benefitted from the creation of this excellent play. Nobody, other than Emily herself, ever even experienced it, or knew of its existence.

There are at least two ways to go here. The first is to deny that Emily's play had significant aesthetic value. This is an implication of certain subjective theories of aesthetic value. On these theories, aesthetic value is determined wholly by people's actual psychological responses to an artwork – say, whether they enjoyed it. Since nobody ever enjoyed Emily's play, it does not have significant aesthetic value, and hence there was no EBR for her to produce it.

The second is to accept the implication but try to minimise its counter-intuitiveness. One strategy would be to claim that, though Emily had an EBR to write her play, it is very unlikely, in this case, that she had sufficient reason, all things considered, to write her play. And, so long as we keep this in mind, it is not that strange to claim that she had *some* reason to write it. Along the same lines, it could be pointed out that – even though her EBR is identical to theirs – Emily likely had less reason to write her play, all things considered, than many of the other people who have been discussed so far had to produce their work. This is because most of these people also had reasons that arise from the fact that people benefitted from the existence of their work. Emily lacked such reasons. But that Emily had less reason, all things considered, to achieve excellence than most should not lead us to deny that she had some reason based purely on the aesthetic value of her work. Aesthetic value, after all, is an important and significant source of value.

The subjective theories that would support the first response strike me as implausible. *King Lear* is a great play independently of whether anybody has ever read it.[46] Since I believe this, I cannot deny that Emily's play could be a great play. As such, I am inclined to accept some version of the second response. I do admit, however, that this objection has intuitive bite.

With this point in mind, it is important to emphasise that the idea that EBRs *are* genuine reasons has itself been given significant intuitive support. We have seen, for instance, that there are cases where it seems extremely plausible that a person has a genuine reason to achieve excellence. The Kierkegaard example that we started with is one such case. It is possible, of course, that a different kind of reason could explain our reaction to this example. Perhaps we think, for instance, that Kierkegaard merely had a prudential reason to leave Olsen and become a writer. It seems less likely, however, that we could plausibly explain why Madeline seems more justified than Elizabeth – or Cameron more justified than Andrew – without appealing to EBRs in particular. After all, it seems to be purely the intellectual and aesthetic value these people create that makes the difference. Even in Emily's case, it seems very plausible that, all else equal, she had *more reason* overall to sacrifice her welfare than an otherwise identical person who was destined to write a terrible play. The life of a person who makes these sacrifices and fails seems much more tragic and depressing – it seems like much more of a waste – than

[46] Note that I am not here rejecting subjective theories of aesthetic value in general. To see this, consider that two prominent subjective theories – dispositional accounts and accounts involving idealisation – do not imply that a play lacks aesthetic value if nobody ever reads or sees it. This is because it can still be true that people *would* evaluate a play a certain way, even if nobody ever *actually* evaluates it that way. See King (2022, 7-8) for discussion of this point.

the life of a person who makes these sacrifices and succeeds. And, returning to the Kierkegaard case, even if other reasons could in principle explain, or help to explain, why Kierkegaard's decision was justified – at least to the extent that it was – we seem to be omitting something important if we don't mention that he actually wrote aesthetically and intellectually excellent works. When we are considering cases like these, the success or failure of the pursuit seems to play an important role in determining whether someone has made a mistake. More simply, it just seems wrong to claim that the mere fact that Hume's actions led to *A Treatise of Human Nature* – or James' to *The Golden Bowl* – does not count at all in favour of these actions.

It may be worth pointing out also that, aside from intuitive support based on cases, there are persuasive general principles that support the claim that EBRs are genuine reasons, and hence that Emily had a reason to write her play. For instance, it is highly plausible, as a general deontic principle, that, if φ-ing would bring about value, then you have a reason to φ. Assuming that my claims about value are correct, this principle entails that EBRs are genuine reasons.

V

For the remainder of this chapter, I will assume – on the strength of the above considerations – that EBRs, as I have understood them, are genuine reasons for action. I will now argue that, so understood, EBRs are neither moral nor prudential reasons for action. Many of these arguments appeal to the appropriateness, or inappropriateness, of certain reactive attitudes. As previously noted, I assume that, if an agent freely and knowingly acts immorally, then indignation, resentment and guilt are fitting responses. My claim will be that, when a person freely and knowingly fails to achieve excellence when they could have done so – when they have an EBR to Φ – these attitudes are not fitting responses. This supports the idea that EBRs are non-moral reasons. I will defend the claim that EBRs are non-prudential reasons in a similar way. If some attitude is a fitting response to imprudence, but that attitude would not be a fitting response to an agent freely and knowingly failing to achieve excellence, then this supports the claim that EBRs are non-prudential reasons. To make this argument, we need to know which attitudes, if any, are fitting responses to imprudence. To answer this, it is useful to start by contrasting prudence with morality.

I will begin by discussing other-directed attitudes. The first thing to note is that, in contrast to immoral action, attitudes such as indignation and resentment seem entirely out of place as a response to another person's imprudent action. On the face of it, this claim may not seem

obvious, and this is especially so when we consider that indignation and resentment do seem to be appropriate in some cases that involve imprudent action. For instance:

> Jennifer is a gambler who spends almost all her time playing baccarat. As a result of this habit, she often loses significant amounts of money, and these losses are causing her life to fall apart. Not only is she about to be evicted, but she hardly sees her two children. In addition, she can barely afford to put food on the table, let alone buy the new clothes that her children desperately need.

Jennifer's actions are clearly imprudent. It also seems fitting to feel anger and resentment. Though these feelings do make sense in this example, it is not a good test case. This is because our reaction may well be explained by the harm that Jennifer is doing to others, and not by the harm that she is doing to herself. What we need is a case where a person is imprudent but doesn't harm anyone else. When we consider such a case, these emotions do not seem to be fitting. For example:

> Again, Jennifer is a gambler who spends almost all her time playing baccarat. She is almost out of money, and is about to lose her job and be evicted. If this happens, her life will be in ruins. Unlike in the previous case, nobody relies on Jennifer. Indeed, she has no connections to anybody.

As before, Jennifer's actions are clearly imprudent. Unlike before, it does not seem appropriate to feel indignation or resentment towards her. More than this, it would be odd if someone felt these emotions upon learning about Jennifer's situation. After all, indignation and resentment are negative emotions; they are expressions of *ill-will* towards a person. What sense does it make to feel ill-will towards a person when all they have done is act in ways that have made their own life worse? It seems completely out of place, for example, to resent somebody for the mere fact that their own life is falling apart.[47]

We can next ask which other-directed emotions do make sense in cases of self-harm. This is a difficult question, but two plausible candidates are *sympathy* and *pity*. It seems clear that it would be fitting, for instance, to feel sorry for Jennifer. Other attitudes, including less palatable ones, may also be fitting responses to imprudence. When a person consistently makes their own life worse, frustration doesn't seem out of place, and neither does a certain kind of contempt. There may even be cases where it makes sense to feel glad that a person's life is

[47] Joyce (2007, 12-14) makes some similar claims in his discussion of imprudence and retributive anger.

falling apart. This could be so if they have previously betrayed you, or wronged you in some other way. Whatever we should say about these claims, it at least seems right that, in most – if not all – cases of imprudence, sympathy and pity would be fitting responses, even if they would not be uniquely fitting. From this, we can make the following claims: If sympathy and pity seem to be appropriate responses to some action, then that gives us reason to believe that the action involved imprudence. More importantly for our purposes, if sympathy and pity do not seem to be appropriate responses to some action – and especially if they seem to be entirely out of place – that gives us reason to doubt that the action involved imprudence.

Self-directed attitudes such as guilt strike me as much less clear-cut. This is because certain components of the attitude of guilt also seem to be present in attitudes that can be fitting in response to self-harm. Self-loathing, for instance, can make sense both as a response to ruining your own life and as a response to ruining someone else's life. This is not to say that guilt is a fitting response to imprudence, but it does make isolating the relevant attitudes difficult. For this reason, I will primarily focus on other-directed attitudes.

With these claims in place, I will now argue that there are important differences between the attitudes that are fitting when someone acts contrary to excellence-based reasons and the attitudes that are fitting when a person acts immorally or imprudently. For ease of expression, when an agent could have achieved excellence, but fails to do so, I will say that they have acted 'imperfectly', or that they are 'imperfect'.

Immorality and Imperfection – Other-Directed Attitudes

I will first argue that how it makes sense to feel when somebody acts immorally is not the same as how it makes sense to feel when somebody acts imperfectly. More precisely, I will argue that the mere fact that somebody acts imperfectly does not warrant indignation or resentment. More than this, these reactions seem to be bizarre responses to a person's failure to achieve excellence.

As with imprudence, it is worth noting at the outset that indignation and resentment can be appropriate in cases that involve imperfection. This is because the achievement of aesthetic and intellectual excellence can have morally relevant instrumental benefits. Consider:

> Clara, a brilliant medical researcher, has been working for a number of years on the cure for cancer. She is close. The work is so far over everybody else's head that only she can bring it to completion. One night, while reading Schopenhauer, Clara

> has an epiphany. She suddenly feels that life is not worth living, and that death is a preferable state to being alive. As a result of this revelation, she abruptly abandons her work on the cure for cancer. It would be a disservice to people, she believes, to save them.

As curing cancer would be a towering intellectual achievement, Clara has an EBR to finish her work. It also seems fitting to feel indignation towards Clara for abandoning her project. I can imagine that, if somebody I loved was dying of cancer, I would feel this emotion strongly. Further, it doesn't seem inappropriate to feel indignation simply as a response to the extreme suffering that could have been prevented but will now occur.

It seems true, then, that indignation and resentment can be appropriate reactions to imperfection. This does not conflict with the claim that I am defending. This is that indignation and resentment are not appropriate reactions to *mere* imperfection. The above case, though it involves imperfection, is morally loaded. Clara's decision is – at least plausibly – immoral as well as imperfect.

Consider next a case that involves imperfection, but no obvious violation of moral reasons. Indignation and resentment would be, in this case, extremely odd reactions.

> Kate is a mathematical genius. She grew up in a family that did not value education at all, and she now shares this perspective. As a result, she never worked hard in school, and she dropped out at a young age. She now works as a bartender and is completely content with her life. This is what she wants to do. If she did go back to school, and if she applied herself, she would quickly be recognised as brilliant. She would end up making ground-breaking, once in a generation contributions to number theory. As is often the case in this area of mathematics, these results would have no – or almost no – practical effect. Even if Kate knew what she would achieve if she went back to school, she would not choose to take this path. She would prefer to keep bartending.

Kate has a strong EBR to go back to school. This derives from the fact that, if she does, she will ultimately produce excellent mathematics. By continuing on her current path, she is acting imperfectly.

The next question is this: Is it appropriate for us to feel either indignation or resentment towards Kate for being imperfect? The answer seems to be 'no'. Kate has simply done nothing to

warrant these reactions. The plausibility of this verdict is strengthened by considering that the *natural expressions of* indignation and resentment seem wildly out of place in Kate's case. These could include, for example, giving her the cold shoulder, or verbally abusing her. In general, it does not seem appropriate to feel any ill-will towards Kate for living the life she wants to live. These considerations support the claim that, in cases of mere imperfection, attitudes such as indignation and resentment are inappropriate.

To strengthen this conclusion, note that, unlike in isolated cases of imperfection, isolated cases of violations of many paradigmatic moral reasons *do* seem to warrant some degree of indignation and resentment. This is true, for example, of cases of deception, disrespect, and harm. This point can be put a bit differently. All else equal, harm, deception, and disrespect warrant indignation and resentment, but failing to become a great mathematician when you have the ability to do so does not.

Before moving on, it is worth noting that certain attitudes do seem appropriate in Kate's case, and towards imperfection more generally. We can appropriately feel a certain kind of *regret* that Kate is not going to become a mathematician. This would likely be rooted in the fact that, as a result, some exceptional mathematics will never be written. It is, we might say, a *shame* that Kate isn't pursuing mathematics. There are also what might be thought of as appropriate negative emotions, even if these don't include resentment and indignation. One example is *envy*. It can make sense, I believe, to envy another person for being more talented than you. This is especially the case given that, as it is often understood, talent is unearned. It can also make sense to *despair* of this fact, and to despair of the fact that you lack what this other person has. This will be particularly apt – and intense – in cases where you have a strong desire to achieve excellence, but do not have the talent to do so.

A Response

There is a potential response to this argument. This is worth considering both for its own sake and because doing so leads to a further argument that EBRs are non-moral reasons. The response involves arguing that complying with EBRs is *supererogatory*. If this is true, it would allow a person to deny that EBRs are non-moral, even if they accept that indignation and resentment are inappropriate responses to imperfection.

An act is supererogatory if it goes 'beyond the call of duty'. More carefully, an act is supererogatory if it is (i) not morally required; (ii) morally permissible; and (iii) morally better

than other available actions that are also morally permissible.[48] To illustrate, suppose that Julia and Jane – who are strangers to one another – come across a burning car on the side of the road. They both approach this car from different sides. The driver is dead, but two children are alive in the back seat. It is clear to both Julia and Jane that the children will perish before any additional help can arrive. It is also clear to both that, if they try to intervene, there is a high chance that they will themselves be engulfed by flames before they are able to save the children. Despite this, Julia rushes into the car and manages to save the child closest to her before the car is fully engulfed. Jane, on the other hand, does not try to save the other child, who burns to death.

Plausibly, while both Julia and Jane's actions were morally permissible, Julia's action was supererogatory. I take it as obvious that Julia's action was permissible, but we may also think that people are not morally required to take grave risks in order to save another person's life. If so, Jane's action was also morally permissible. This also suggests that Julia's action was not morally required, since it would have been morally permissible for her to allow the child to die. I take it as obvious, in addition, that Julia's action was morally better than Jane's. If all this is correct, then Julia's action was supererogatory.

This example also allows us to see two important features of supererogatory and non-supererogatory action. First, acting in a supererogatory manner is morally *praiseworthy*. There are certain attitudes that it is appropriate to feel towards Julia which are not appropriate to feel towards Jane. Second, acting in a non-supererogatory manner is not morally *blameworthy*. Though Jane is not deserving of praise, she is also not deserving of indignation or resentment. Since she acted in a way that was morally permissible, these attitudes would be inappropriate.

We can now see how this response works concerning EBRs. Clearly, both Jane and Julia had moral reasons to attempt to save the children. Nonetheless, it is inappropriate to blame Jane for failing to comply with these reasons. If EBRs are like this, then it will be the case that, even though people have moral reasons to achieve excellence, it will not be appropriate to feel indignation or resentment towards them for failing to do so. This would explain how EBRs could be moral reasons despite these emotions being inappropriate in Kate's case.

[48] This is a rough characterisation of supererogation, but it should suffice for our purposes. For detailed discussion, see Heyd (1982; 2019).

Though this would explain Kate's case, it is not plausible that complying with EBRs is supererogatory. This is because complying with EBRs is not morally praiseworthy. The reactive attitudes that are appropriate in Julia's case – and in other paradigm cases of supererogation – make little sense in cases of aesthetic and intellectual excellence. To see this, reflect on how we feel about Julia, or how we feel about Nelson Mandela, and how we feel about David Hume and Henry James. No doubt we feel great admiration for Hume and James, but it is phenomenologically very different to how we feel about Julia or Mandela. We would not, for example, consider it appropriate to award Hume and James certain prizes, such as heroism awards or the Nobel Peace Prize. On the other hand, a heroism award seems entirely appropriate in Julia's case, and the Nobel Peace Prize seems like appropriate recognition for Mandela. We also tend to think of those who consistently act in a supererogatory manner as *saints*. This is not at all how we feel about Henry James for consistently writing excellent novels.

THE FALSE ATTRIBUTION ARGUMENT

It seems, then, that complying with EBRs is not supererogatory. A somewhat similar argument suggests that EBRs are also not non-supererogatory moral reasons. If both of these claims are correct, then EBRs are not moral reasons at all.

In brief, the argument is this: If EBRs are moral reasons, then, all else equal, we should consider a person who complies with excellence-based reasons to be a morally better person than somebody who does not. We do not consider a person who complies with EBRs to be a morally better person than somebody who does not. Therefore, EBRs are not moral reasons.

With one qualification, the first premise seems correct. The idea is just that, if two people are equally morally good in every respect except one, and that one difference involves doing more morally good things, then it seems natural to say that the person who does more is a morally better person. It seems clear, for instance, that we have moral reasons to assist people in need. Suppose that, with one exception, two people perform identical actions with identical intentions, and suppose that both are presented with identical opportunities. The exception is that one of these people helped someone in need when the opportunity presented itself and the other did not help someone in need when the opportunity presented itself, and nor did they perform any other action that they had moral reason to perform in its place. It seems plausible

to claim that, as a result of this difference, the person who acted beneficently is a morally better person than the one who did not.[49]

The premise does require a qualification. Sometimes we have a moral reason to perform an action – or are morally required or permitted to perform an action – due to our own past voluntary behaviour. One example is *promises*. If I promise you that I will φ, then – at least typically – I have a moral reason to φ that I would not have if I had not promised you that I would φ. Call these *voluntary moral reasons*. The important point, for our purposes, is that it is far from obvious that I am a morally better person than I would otherwise be – or than another person is – merely in virtue of complying with voluntary moral reasons. Suppose that, as part of my daily routine, I make various promises to people that are very easy to fulfil. And imagine that the only difference between me and some other person is that I fulfil one of these promises at the end of each day. It is plausible that I have a moral reason to fulfil these promises. If so, then I comply with an extra moral reason. But it seems wrong to say that this makes me a morally better person.

A more accurate version of the premise, then, is this: If EBRs are *non-voluntary* moral reasons, then, all else equal, we should consider a person who complies with excellence-based reasons to be a morally better person than somebody who does not. This qualification, though important, does not make a difference to my argument. This is because, whatever sort of reason EBRs are, they are not voluntary. A person cannot simply choose whether they have the potential to write the next *Treatise of Human Nature.* As such, if EBRs are moral reasons, then they are non-voluntary moral reasons. Given this, they fall under the purview of the first premise. If EBRs are moral reasons, then, all else equal, we should consider a person who complies with excellence-based reasons to be a morally better person than somebody who does not.

According to the second premise, we do not consider a person who complies with EBRs to be a morally better person than somebody who does not. To see this, consider two examples. The first involves intellectual value and the second aesthetic value.

[49] It is perhaps simpler to think of this as an example of comparing your own actual actions with alternative actions that you could have taken. If, when reflecting on your life, you determine that you could have helped more people than you in fact did, then it seems natural to conclude that you could have been a morally better person than you in fact are (assuming that you didn't perform alternative actions that were also supported by good moral reasons).

> For twenty-two hours a day, two people – Naomi and Veronica – perform exactly the same actions, with exactly the same intentions. These actions also have identical consequences. The only difference in their lives is this: For two hours a day, Naomi does some morally neutral activity, such as watching television.[50] Veronica, on the other hand, spends two hours writing obscure and practically ineffectual, but excellent, articles on metaphysical problems arising from vagueness.
>
> The situation is exactly the same as above, with one difference. Instead of working on vagueness, Veronica spends two hours a day working on obscure and practically ineffectual, but excellent, surrealist paintings.

Given that the works that Veronica produces are intellectually and aesthetically excellent, respectively, she has EBRs to produce this work. Now, assuming that EBRs are non-voluntary moral reasons, it seems to follow that Veronica is a morally better person than Naomi. After all, this view implies that Veronica spends two extra hours a day complying with non-voluntary moral reasons. Everything else is equal. But, in both cases, it seems clear that Veronica is not a morally better person than Naomi. They seem to be equally morally good.

There is a more general point here, which doesn't rely on the specifics of the above argument. We simply do not take the fact that somebody has achieved excellence as a painter, or as a metaphysician, to be *any* kind of evidence that they are a morally good person, or even a morally *decent* person. This is perplexing if EBRs are non-voluntary moral reasons. After all, we would take the fact that somebody has acted beneficently, honestly, or kindly as evidence that they are a morally good person, or at least a morally decent person.

PRUDENCE AND EBRS – OTHER-DIRECTED ATTITUDES

I will now discuss whether sympathy and pity are appropriate responses to imperfection. If imperfection is simply a kind of imprudence, then we should expect that they often, if not always, will be. I will argue that this is not the case. I will first argue that, in certain cases, pity and sympathy are strange response to imperfection. I will then consider an example that seems to support the idea that pity and sympathy can be an appropriate response to imperfection, and I will argue that this case is deceptive.

[50] What exactly counts as morally neutral is going to vary from theory to theory, so I am just using 'watching television' as a placeholder.

As with indignation and resentment, there are examples where pity and sympathy seem to be entirely out of place as a reaction to imperfection. Consider:

> Carmen has the potential to be an excellent biologist. While at university, however, she decided that she did not want to become a researcher. She decided instead to become a public servant. She now lives a comfortable middle-class life with a loving partner and beautiful children. She is happy, and she has never regretted her decision.

Carmen has acted imperfectly. She had an EBR to become a biologist, but she did not do so. Further, in making this choice, we can assume that she did not pick another option that would also achieve excellence. We can suppose that she ended up with a fairly average job, and a fairly average life.[51]

If it is true that imperfection consists in failing to comply with prudential reasons, then it should be the case that it makes sense to pity or sympathise with Carmen. But this does not seem to make sense. She is living the life that she wants, and this life does not itself seem to be inherently pitiable. We can again get a better sense of the strangeness of pity or sympathy in this case by considering their natural expressions. Two obvious examples include offering Carmen assistance or consoling her. Both seem out of place.

There is a possible response to this argument. It might be claimed that the above example is a faulty test case. This is because Carmen's life is otherwise going well. This is what makes pity and sympathy inappropriate. Any imprudent elements in her life are swamped by the good things.

I am not convinced by this response, for the following reason. If we strip Carmen of other clear prudential goods – but keep her life otherwise the same – then it still seems that some degree of pity or sympathy is fitting. If, for example, we keep the example otherwise the same but imagine that Carmen is unhappy, or that she is being deceived, then feeling some degree of sympathy towards her seems to make sense. This remains the case even if, overall, her life is going well for her. In contrast, when I consider the mere fact that Carmen chose not to become a biologist – and hence acted imperfectly – I do not feel any sympathy towards her. If others

[51] 'Average' in the sense of non-remarkable and non-exceptional, not in the pejorative sense where this means something like pedestrian.

share this response, then this gives us reason to doubt that imperfection consists of failing to comply with prudential reasons.

It is worth noting that this sort of argument seems to generalise. When we consider certain facts about people's lives, in isolation, it is often clear that at least some degree of pity or sympathy is appropriate. Whatever else is happening, it is fitting to feel sympathetic towards somebody who is unhappy. The same plausibly holds for certain common items on objective lists. If we learn, for example, that somebody does not have any friends or loving relationships, then some degree of sympathy seems to make sense. Again, this seems true regardless of other facts about their life.

This does not seem true of mere imperfection. The mere fact that a person could have been a great dancer, but is not, does not arouse sympathy in me. I feel that I need to know other facts, such as – at minimum – whether they *want* to be a dancer. If they do, then this response makes sense. But, if not, then the fact that they had EBRs to be dancer does not seem to make sympathy fitting. This gives us further reason to doubt that EBRs are prudential reasons.

ANOTHER DECEPTIVE CASE

I have argued that there are cases where pity seems like an inappropriate reaction to imperfection. It might be objected that this alone does not show that EBRs are non-prudential. After all, I have not established that sympathy and pity are always fitting responses to imprudence. This is true, and the objection may seem to be supported by cases where pity does seem to be a fitting response to imperfection.

If there really are such cases, then this objection is a good one. I am sceptical that there are. It seems to me that, in cases where this appears to be true, the appropriateness of our reactions is explained by something other than mere imperfection. One example is when people *desire* to achieve excellence but fail. In these cases, the frustration of the desire, rather than the mere fact that the person fails to achieve excellence, seems to explain the appropriateness of sympathy and pity.

It is difficult to demonstrate that this is true of every possible case. What I shall do here is discuss just one particular kind of case. These seem to be *prima facie* compelling examples of pity being an appropriate response to imperfection. If I can show that this is not so in these cases, this should at least make us doubt, when we feel this attitude in response to other cases, that it is really the imperfection doing the work.

The kind of case that I shall discuss are those involving a person who could achieve excellence but who instead wastes their life doing other things. Consider:

> Alex has the potential to be a great musician. If he consistently applied himself, he would end up creating works of profound beauty. Unfortunately, Alex has fallen on hard times. He is a heroin addict and is struggling to hold down a string of demeaning jobs. When he does have spare time, he is too messed up to practice.

Alex acts imperfectly, and pity seems to be an appropriate response. The question is whether our feelings track the imperfection itself. This is doubtful. Other features of the case seem to better explain the appropriateness of pity. These include, most saliently, the fact that Alex is living a miserable life, and perhaps even that he is not living up to his potential. Admittedly, if he did live up to his potential, he would achieve excellence. In this sense, it is a pity that he is not achieving excellence. The appropriateness of this response seems to ultimately be explained, however, by Alex not living up to his potential – which is plausibly something that he does have a prudential reason to do – not by the mere fact that doing so would achieve excellence.

This claim may appear unconvincing. It is supported by considering a certain sort of counter-factual instability. Suppose that Alex is living the same life, but imagine that his potential is lower than the level required to achieve excellence. This version of Alex is not acting imperfectly. Should we feel any less pity or sympathy toward him as a result? The answer seems to be 'no'. The facts that warrant this response remain in place. Namely, his life is not going well, and it could be a lot better. It seems irrelevant that the first version of Alex could be a great musician, but the other could not.

This claim should be qualified slightly. There are attitudes that are fitting in the first version but not the second. In the first version, it is, for example, *regrettable* that Alex is not living up to his potential in a way that it is not in the second. In this version, excellent music is not being created as a result. Again, however, I don't see why we should feel more sympathetic towards this version of Alex for this reason. People don't deserve more sympathy from us just because they are more talented.

EBRs and Prudence – A Further Argument

To this point, my argument that EBRs are non-prudential reasons may seem less compelling than my argument that EBRs are non-moral reasons. For this reason, I shall now offer a further argument for the claim that EBRs are non-prudential.

We can start by noting that, if we claim that EBRs are prudential reasons, then we are also committed to the claim that a person would benefit from writing a great novel even if they do not care about this novel – and even if they find the thought of creating it *undesirable* – and even when its creation would bring them nothing but misery. This is because there is just no reason to believe either that achieving excellence will always positively contribute to our happiness, or that we always desire to achieve excellence – or would desire to achieve excellence if we knew all the relevant facts and were thinking clearly. This point alone is enough to show that, on many prominent theories of welfare, EBRs cannot be prudential reasons. These theories include mental state theories, desire-satisfaction theories, and most hybrid theories. If one of these theories is true, then a person can have an excellence-based reasons to Φ, but no prudential reason to Φ. And, if a person can have an excellence-based reason to Φ, but no prudential reason to Φ, then excellence-based reasons cannot be prudential reasons.

The idea that something cannot be good for an agent if it leaves them cold or miserable is powerfully expressed by Peter Railton (1986b, 9):

> It does seem to me to capture an important feature of the concept of intrinsic value to say that what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him.

Not everyone finds this thought persuasive. There are various theories of welfare that can accommodate the idea that excellence-based reasons are prudential reasons. Perhaps the most promising candidate is some kind of objective list theory. It could be argued that achieving excellence is an objectively good thing for a person, whatever their desires or mental states. This view will of course commit one to the idea that there are alienating prudential goods, but this alone won't convince an objective list theorist that excellence-based reasons are non-prudential reasons. After all, every defender of an objective theory of welfare accepts that

certain prudential goods can be alienating – indeed, this is plausibly what distinguishes objective from subjective (and hybrid) theories of welfare.

Note, however, that – whatever we think of alienating prudential goods in general – EBRs are alienating in a particularly stark manner. To see this, consider two items commonly included on objective lists: autonomy and knowledge. Assuming these are genuine goods, both can be alienating. The freedom to make our own choices can be a source of crushing anxiety, and learning certain truths can cause near suicidal despair. These are not things we always want. Despite their potential to alienate agents, knowledge and autonomy remain, in some important sense, *connected* to agents. Knowledge is something that a person *possesses*, and autonomy is something that an agent *has*. This connection makes it plausible that, if autonomy and knowledge are intrinsically valuable, my having those goods is good for me even if I don't want them. Excellence-based reasons, however, are not like this. The value of achieving excellence that these reasons track is aesthetic and intellectual value. This value is not something that I possess or have. It exists completely independently of me and my life. This lack of connection makes the claim that achieving excellence necessarily benefits an agent much less plausible.

VI

This chapter had three main aims. The first was to give an account of excellence-based reasons. An agent has an excellence-based reason to Φ if and only if, and because, Φ-ing would lead to, contribute to, or in some way promote the creation of intellectually or aesthetically excellent works. The second aim was to motivate the claim that excellence-based reasons, so understood, are normatively significant; they make a difference to how an agent should actually live their lives. The final aim was to show that excellence-based reasons are neither moral nor prudential reasons. This allows for the possibility of genuine conflict. An agent may have a normatively significant reason to achieve excellence despite them having no moral reason to achieve excellence, and there may be a lot to be said for an agent achieving excellence even if this achievement fails to make the agent's own life better in any respect.

CHAPTER TWO: MORALITY AND PRUDENCE

The purpose of the next two chapters is to argue that certain cases strongly support moral anti-rationalism. The cases that I discuss involve prudential and excellence-based reasons. As I have already argued that these are genuine non-moral reasons, I assume that, when an agent faces a conflict between morality and prudence, or a conflict between morality and excellence, the agent has genuine reason to act immorally. The claim that I defend is that, in some such cases, the agent also has sufficient reason, all things considered, to act immorally.

This chapter focuses on prudential reasons. I will begin by discussing what I call *acute conflicts* between morality and prudence. In these cases, an agent faces a mutually exclusive choice between acting morally and acting prudently. What makes the conflict acute is that, if the agent acts as morality demands, her life will be ruined.

I

Consider:

> *The Phone Call:* Mary is an agent with relatively typical desires, projects, and attachments. She has a partner she loves and close ties to various family members and friends. She cares about her career and aspires to continue to advance in her field. Overall, Mary likes her life and feels that it is going well. She has many of the things that she wants and is on track to attain those things that she desires but currently lacks.
>
> One night Mary suddenly wakes. For reasons she can't identify, she is overcome by feelings of dread and unease. As if she is being guided by an external force, she gets up and begins to walk towards the kitchen. As soon as she enters, her home phone – which nobody uses anymore – begins to ring. When she answers, a sinister voice tells her the following: Unless Mary drinks the entire bottle of undiluted bleach that she keeps under the sink many innocent strangers will die in seemingly random but excruciating ways. These deaths will never be connected, and will certainly never be traced back to this phone call. In addition, if Mary refuses to drink the bleach, the phone call itself will later seem to her like only a bad dream. As a result, she will not be haunted by this decision.

Let's suppose this threat is real, and that Mary knows this.[52] Some claims about this case are obvious. It is obvious, for example, that it is prudent for Mary to refuse. The idea that her life will go better for her if she drinks the bleach than if she doesn't is absurd. Other claims are less obvious. Is it the case, for instance, that, given the number of people who will die if she refuses, Mary is morally required to drink the bleach?

The question that I will focus on is this: Is it plausible to interpret *The Phone Call* in a way that is compatible with moral rationalism? To see what a rationalist-friendly interpretation of this case would look like, it is useful to first consider what an anti-rationalist interpretation of this case would look like. This would involve two claims. The first is that it is reasonable for Mary to refuse to drink the bleach.[53] This claim would likely be motivated by the fact that, if she drinks the bleach, Mary will lose everything that matters to her. The second claim is that Mary is morally required to drink the bleach. The motivation for this claim is likely to be that, if she doesn't, then many innocent people will die in excruciating ways. A rationalist-friendly interpretation of the case must deny at least one of these two claims. This leaves three options available to the moral rationalist:

> (i) *Reasonable and Morally Permissible:* In line with the anti-rationalist interpretation, this view claims that it is reasonable for Mary to refuse to drink the bleach. Unlike the anti-rationalist interpretation, it claims that refusing to drink the bleach is morally permissible.
>
> (ii) *Unreasonable and Morally Required:* This view agrees with the anti-rationalist interpretation that Mary is morally required to drink the bleach. Unlike the anti-rationalist interpretation, it denies that it is reasonable for Mary to refuse.
>
> *(iii) Unreasonable and Morally Permissible.* This view claims that the anti-rationalist interpretation gets both verdicts wrong. It claims, that is, that it is

[52] Perhaps through supernatural means.

[53] As a reminder, I use the term 'reasonable' to mean that an agent has sufficient reason, all things considered, to perform an action. I use 'unreasonable' to mean that an agent lacks sufficient reason, all things considered, to perform an action – or, in other words, that they have decisive reason, all things considered, *not* to perform an action.

morally permissible for Mary to refuse to drink the bleach, but that doing so would be unreasonable.[54]

It is important to emphasise that the issue that concerns us here is not whether the moral rationalist *can* explain every purported counterexample to moral rationalism in a rationalist-friendly manner. Since the three views above are coherent, I don't doubt that they can. But the mere fact that a view is coherent is not a reason to believe it. The question that matters is whether a rationalist-friendly interpretation is a *plausible* response to every case, not whether it is an available response. It is the view that these responses are always plausible that I will reject.

To see what it would take for my argument to succeed, it is helpful to consider what it would take for moral rationalism to be true. If moral rationalism is true, then a rationalist-friendly interpretation must be correct in every conceivable case. This is because moral rationalism just is the claim that, if an agent is morally required to Φ, then they have decisive reason, all things considered, to Φ. This claim is false – and moral anti-rationalism is true – if there is even one case where an agent is morally required to Φ but lacks decisive reason to Φ. For this reason, if we become convinced that there is at least one case where a rationalist-friendly interpretation is implausible, then we should reject moral rationalism and endorse moral anti-rationalism. In at least this sense, moral rationalism is a significantly stronger claim than moral anti-rationalism.

II

The strength of moral rationalism has sometimes itself been taken as a compelling reason to reject the view. Samuel Scheffler (1994, 56), for example, writes that moral rationalism 'is a very strong claim. Just because it is so strong, it seems to me unlikely to be true.' The thought here seems to be that, given the vastness of the conceptual space, it is highly likely that there

[54] Though this is a conceptual possibility, it would be an odd response to *The Phone Call*. This is because the best candidates for the facts that would make refusal unreasonable are the same facts that would make refusal morally impermissible. There are cases, however, where this seems like the right result. For example, this is the intuitive verdict in many cases where an agent knowingly makes her own life worse when this impacts nobody else. (This assumes that agents have what Lazar (2019) calls 'self-sacrificing options', since an agent also makes the world go impartially worse when they make their own life go worse.)

is at least one case out there where an agent lacks decisive reason to act as morality demands. Indeed, it might seem like it would be a miracle if moral rationalism turned out to be true.

This line of reasoning has force as it is, but it can be developed in a way that makes it even more persuasive. This is because, even if one is not moved by *The Phone Call* in particular, there is a good rationale for thinking that there must be some case along the same lines where a rationalist-friendly interpretation will be implausible. The reason for this is that cases like *The Phone Call* illustrate two compelling claims that, taken together, are incompatible with moral rationalism.

The first claim is that there are certain personal sacrifices that it is reasonable for an agent to refuse to make. One of the more compelling examples is that it is reasonable to refuse to sacrifice the activities, projects, relationships, and features of yourself that are necessary for you to have a desire to live – or, less dramatically, to get out of bed in the morning. As Jackson (1991, 461) puts it, these are the commitments that give our lives 'shape, meaning and value'. If we sacrificed these, then we would be little more than living corpses. Intuitively, to refuse to do this to yourself is not irrational or unreasonable. In a similar vein, it doesn't seem unreasonable to refuse to perform an action that you would despise yourself for performing, or to refuse to kill yourself when you have a strong desire to live.

The second claim is that, in certain circumstances, morality could require an agent to make personal sacrifices of these kinds. There are various ways to motivate this claim, but the cases where it seems most immediately plausible are those – like *The Phone Call* – where the consequences of failing to make the sacrifice will be catastrophic. One of the reasons that this claim is difficult to deny is that it is supported by some of our most basic moral commitments. For example, virtually everyone accepts that, from the moral point of view, everyone's interests' *matter* – indeed, many believe that, in some important sense, everyone's interests matter equally. Given this, and even supposing that it is morally permissible for an agent to favour their own interests to some degree, it seems that there must be some point at which refusing to sacrifice your own interests for the sake of other people's interests becomes incompatible with taking the interests of others seriously, let alone with treating them as moral equals.

These two claims are jointly incompatible with moral rationalism because, if they are true, then there must be a case where morality would require an agent to make a sacrifice that it is not unreasonable for them to refuse to make. Even if this argument is not decisive, the plausibility

of the two claims still provides strong support for moral anti-rationalism. For one thing, the plausibility of the two claims suggests that we have reason to prefer moral anti-rationalism over moral rationalism on the grounds of intuitive fit. These claims can only be true if moral anti-rationalism is true, and that is a reason to accept moral anti-rationalism. Given that intuitive fit is one of the most important virtues when it comes to evaluating normative theories, this consideration is significant.[55] We can put this point another way. Given that some of our deepest convictions about the content of morality and our reasons for action do not line up in the way that moral rationalism requires, and that moral anti-rationalism can vindicate these convictions, we have good reason to accept moral anti-rationalism over moral rationalism.

III

Return to *The Phone Call*. Is a rationalist-friendly interpretation plausible in this case? We can first note that this case helps to illustrate why one kind of rationalist-friendly interpretation is *not* plausible. As has been argued by rationalists and anti-rationalists alike, if we endorse agent-neutral consequentialism, then we have little hope of vindicating moral rationalism.[56] This is because, if we say that an agent is always morally required to bring about the impartially best outcome, then we will not be able to give rationalist-friendly interpretations of certain cases without making implausible claims about what agents have sufficient reason to do.

To see this, consider a variation of *The Phone Call:*

> The facts are the same as in the original case, with the following exceptions: Instead of many people dying as a result of Mary refusing to drink bleach, only one person – Xavier – will die if she refuses. Xavier is identical to Mary in every relevant respect except one. The difference is that, if he survives, Xavier will live a marginally better life – the equivalent of an extra millisecond of the mildest pleasure – than Mary would if she survived.

Since all else is equal between Mary and Xavier, any version of agent-neutral consequentialism that takes welfare to be one of the values to be maximized will be committed to claiming that Mary is morally required to drink the bleach. The problem for a rationalist-friendly

[55] See Dorsey (e.g., 2012 and 2016, Chp. 3 & 4) for some detailed arguments for moral anti-rationalism that appeal to intuitive fit.
[56] See, for example, Portmore (2011, Chp. 2), Dorsey (2015), Stroud (1998), Hurley (2006) and Sobel (2007a).

interpretation of this case is that, while it may be plausible that it is reasonable for Mary to sacrifice herself to save Xavier, it is very implausible that it is unreasonable for her to refuse. This claim is supported by two considerations. On the one hand, it seems clear that Mary must have strong non-moral reasons not to drink the bleach. After all, as discussed above, many of our deepest convictions about our reasons for action concern the reasons that agents have to pursue the projects and commitments that are central to their lives. On the other hand, the moral difference between Mary drinking and not drinking the bleach is miniscule. Put simply, given that there is barely any moral difference between the options, how could it be the case that Mary doesn't have at least sufficient reason to refuse to drink the bleach due to the strong non-moral reasons that support that option? Discussing a similar case that involves the choice between saving your partner and saving a stranger when saving the stranger will only make the world marginally better overall, Portmore (2011, 30) writes:

> On what plausible theory of practical reasons… would there not be at least sufficient reason for you to save your partner instead of this stranger? I cannot think of any.

These cases, of course, just illustrate a more general point. If agent-neutral consequentialism is true, then moral rationalism is true only if it is always unreasonable to perform any action that is not impartially best. This is not a plausible claim about our reasons for action.[57]

To be clear, I do not take this point as grounds to reject agent-neutral consequentialism. This would only be the case if moral rationalism was true. But it does show that, if one wants to vindicate moral rationalism while making plausible claims about our reasons for action, then they must embrace a different view of what morality requires.

IV

The most promising way to defend moral rationalism in the face of examples like *The Phone Call* is to endorse some version of what I will call COST. According to this view, it can be morally permissible for an agent not to perform the action that is morally or impartially best.

[57] This view of practical reason is defended by de Lazari-Radek and Singer (2014, Chp. 7). They do not defend this on the basis of its plausibility, but by arguing that our beliefs about our reasons to favour ourselves, or to favour those we love, fall prey to evolutionary debunking arguments. The principle of universal benevolence, on the other hand, survives these debunking arguments. This argument is worth further discussion, but I will not engage with it here. I would be satisfied if I could show that moral rationalism cannot be vindicated without making counterintuitive claims about either our reasons for action or our moral obligations.

This will be the case when performing that action would have sufficiently high costs for the agent. To give a familiar example, we might think that, if it would cost me my leg to save a drowning child, it would be morally permissible, due to this very fact, for me not to save the child. And this would so despite it clearly being morally better that a child is saved than that my leg is saved.[58]

There are many distinct theories that endorse COST. Among other things, these will differ in what they take to ground COST and to what extent they allow an agent to favour their own interests when these conflict with what is morally best. My plan is not to discuss these distinct theories, and nor is it to argue that we should reject COST. What I hope to show is that moral rationalism cannot be vindicated on any plausible conception of COST.

It is worth flagging one substantive assumption that I make about COST. This is that any plausible form of COST will say that there are limits to the extent to which an agent can favour their own interests. In other words, I assume that, even if COST is true, there will be cases where an agent is morally required to sacrifice their own interests in order to do what is morally best.

We can next note that, while endorsing COST may help to vindicate moral rationalism, these are logically independent propositions. The debate over COST concerns which considerations are relevant to whether an agent is morally required to perform an action in the first place. The debate over moral rationalism concerns whether, given that an agent is morally required to perform an action, they always have decisive reason to perform it. Since these two claims are independent, an anti-rationalist can accept COST and a moral rationalist can reject it.[59]

The reason that endorsing COST may nonetheless help to vindicate moral rationalism is that COST promises to provide a defender of moral rationalism with the necessary resources to explain cases like *The Phone Call* in a rationalist-friendly manner without committing them to implausible claims about what we have reason to do. This is because, when an action that is not morally best seems reasonable, it is open to a defender of COST to claim that it is morally permissible for the agent to perform the suboptimal action that seems reasonable. As we have seen, this is not something that an agent-neutral consequentialist can say.

---

[58] For some good discussions of COST, see Kagan (1989), Scheffler (1982 and 1994), Haydar (2009) and Barry & Overland (2016). For some illuminating discussions of the related concept of agent-centred options, see Lazar (2017 and 2019).

[59] Scheffler (1994, 115-133) is an example of someone who accepts COST but rejects MR. de Lazari-Radek and Singer (2014, chp. 7) accept MR but reject COST.

We can now look more closely at how this response would work when it comes to *The Phone Call.* Consider first a variation where, if Mary refuses to drink the bleach, then two strangers who are identical to Mary in all relevant respects will die in ways that are as painful as drinking bleach. In this variation, it seems reasonable for Mary to refuse. Despite this, it still seems right to say that Mary drinking the bleach is the morally best option. Assuming they agree that the action is reasonable, a defender of COST will say that, due to the costs to Mary involved in drinking the bleach, it is morally permissible for her to refuse. And this claim – that it is both morally permissible and reasonable for Mary to refuse to drink bleach – is compatible with moral rationalism.

Since we are assuming that there are limits to the extent to which an agent can permissibly favour their own interests, there will be variations of *The Phone Call* where COST says that it is morally impermissible for Mary to refuse to drink the bleach. Suppose, for instance, that one million children will be buried alive if she refuses. What a defender of COST will say about these variations is that, once we reach the point where it is implausible to claim that Mary is morally permitted to perform the morally inferior action, it will be plausible to claim that it is unreasonable for Mary to refuse to drink the bleach. And this claim – that it is both morally impermissible and unreasonable for Mary to refuse to drink the bleach – is also compatible with moral rationalism.

Now that the mechanics are clear, we can turn to whether this is a plausible response to *The Phone Call*. Two initial points are worth making. First, COST clearly does a much better job of explaining these cases in a rationalist-friendly manner than agent-neutral consequentialism. Second, there are variations of *The Phone Call* – like the Xavier variation discussed above – where COST seems to provide a plausible rationalist-friendly explanation. For these reasons, I will discuss a different kind of example in the next section that is particularly problematic for vindicating moral rationalism using COST.

With that said, it is still not obvious that COST can plausibly deal with all variations of *The Phone Call*. One reason to be sceptical is due to the arguments discussed in section II. To focus on a specific issue, suppose it is right that, if a certain number of people will die horrible deaths, then Mary is morally required to drink the bleach according to plausible versions of COST. At this point, a rationalist will have to say that it is now plausible to claim that it is unreasonable for Mary to refuse to drink the bleach. But is this plausible? When we reflect on why it seems right to say that it is reasonable for Mary to refuse to drink bleach in many variations of *The*

Phone Call, the considerations that are front and centre have to do with Mary herself – that, for instance, drinking the bleach would involve sacrificing her life and, along with it, everything that matters to her. These considerations remain in place as we move from one variation to another.

V

Suppose that cases like *The Phone Call* fail to demonstrate that plausible versions of COST cannot vindicate moral rationalism. I will now discuss a different example. I will argue that appealing to COST in this case has implications that many would be unwilling to accept. Discussing this case will also provide the resources for a general argument for moral anti-rationalism.

Consider:

> *The Voyeur:* Tom has recently moved into a new apartment. One afternoon he discovers that he can see clearly into the apartment across from him. A beautiful woman, Grace, happens to live there. He can see everything from her most mundane to her most private activities. Due to the extremely dark tinting on his windows, she is completely unaware that he is watching. Over time, Tom becomes obsessed with watching Grace; this activity becomes the centre of his life. This is not so in the compulsive way that oxycodone might become the centre of someone's life, but in the reflectively endorsed way that admiring beautiful works of art might become the centre of someone's life. Tom derives a great deal of pleasure from this activity, and, to the extent that it interferes with his other pursuits, he is always willing to make the trade-off. In short, nothing matters to Tom nearly as much as watching Grace.

We can first ask whether *The Voyeur* is an acute conflict between morality and prudence. As with the previous case, the claim that it is prudent for Tom to continue to spy on Grace seems clear. To bolster this, we can stipulate that, if he stops watching her, he will forever feel that there is a hole in his life that nothing can fill. There will never be another project, or relationship, that he truly cares about. We can also suppose that nothing else will bring him even close to the same level of pleasure.

Turning to the moral status of Tom's actions, I take it that it is intuitive that spying on someone through their window, and especially during their most private moments, is not only creepy but morally impermissible. At minimum, based on the numerous novels and films that have explored scenarios of this kind, this is the commonsense view. Other considerations support this claim. For example, it seems that it would be fitting for Grace to resent Tom if she ever learned what he was doing. It also seems fitting for Tom to feel guilty about what he is doing.

VI

I will now argue that *The Voyeur* supports a general claim about the moral relevance of prudential reasons that, if correct, makes moral rationalism virtually impossible to defend using COST. We can first suppose that, in the case as initially described, the prudential benefits that Tom derives from spying on Grace are not significant enough to make it morally permissible for him to continue to spy on her. Call this *version one*. Now suppose that we significantly ramp up the prudential benefits that Tom derives from this activity, and hence how costly it would be for him to stop spying on Grace. Call this *version two*.

We can now ask: With the stipulation that Tom benefits significantly more from spying on Grace in version two than in version one, does our conviction about the impermissibility of Tom's actions begin to diminish as we move from version one to version two? As we have understood the view so far, COST says that it should. This is because it is evidently more costly for Tom to stop spying on Grace in version two than it is in version one. The problem for COST is that this fact about Tom seems to make no difference.

In case it is not clear, it is worth saying a bit more about what I have in mind when I say that, if COST is true, then our conviction about the impermissibility of Tom's actions should begin to diminish as we move from version one to version two. The idea is this. If COST, as we have understood it so far, is true, then Tom spying on Grace is *closer to morally permissible* in version two than in version one. This is so in the following sense: Since it is more costly for Tom to stop watching Grace in version two, Tom is closer to the point where, due to the costs to the agent, it would be morally permissible for him to continue to watch Grace. To put this another way, he is closer to the point where the costs of performing the morally best action – not spying on Grace – make it morally permissible for him to perform an action that is not morally best – continuing to spy on Grace. This is so in the same sense that it is closer to morally permissible to fail to save a drowning child when it will cost you your recently

deceased mother's wedding ring than it is when it will cost you a pair of shoes that you don't particularly like. When considering this example, we take these costs to be relevant to the moral permissibility of failing to save the child, and the higher the costs, the less inclined we are to conclude that it is morally impermissible to fail to save the child. Among other things, this relevance is reflected by the fact that, all else equal, we would judge the person who won't sacrifice their shoes much more harshly than we would judge the person who won't sacrifice their recently deceased mother's wedding ring. Analogously, if COST is true, then we should be less inclined to conclude that Tom has acted impermissibly in version two than in version one, and we should judge Tom more harshly in version one than in version two. Neither of these claims seem true.

There are various things that could be said to try to explain why *The Voyeur* is not a problem for COST. For example, it might be claimed that there is not that much of a prudential difference between the cases, and that this explains why our intuitions don't change when we move from version one to version two. This particular response is mistaken, since there is a significant prudential difference. But more importantly, any response along these lines is missing the fundamental point. This is that there is something implausible about the very idea that whether it is morally permissible for Tom to spy on Grace turns on the extent to which he benefits from spying on her. This consideration does not seem to be relevant to the moral permissibility of his action. Put simply, it seems implausible to think that whether Tom enjoys watching Grace in the shower is relevant to whether it is morally permissible for him to watch her in the shower.

The explanation for why the pleasure that Tom derives from watching Grace in the shower is not relevant to whether it is morally permissible for him to do so is, I believe, something like the following: The source of Tom's pleasure is itself morally objectionable or distasteful. And when the source of an agent's pleasure is morally distasteful, then it is implausible to claim that the pleasure that is derived from this source counts against – or could defeat – a moral requirement. Assuming that something like this is correct, a plausible version of COST will have to say that certain prudential costs do not make a difference to the moral permissibility of an action; it will have to say that there are morally irrelevant prudential considerations.

Note next that, once we accept that there are morally irrelevant prudential costs, we must reject the rationalist-friendly interpretation of *The Voyeur* that claims that, due to these costs, it is morally permissible for Tom to continue to spy on Grace. On the face of it, this result may not seem so bad. After all, it is not as if a moral rationalist who accepts COST wants to say in

response to every putative counterexample that the prudent action is reasonable and morally permissible. And for all that has been said so far, it is still open to a moral rationalist to claim that it is not only morally impermissible for Tom to continue to spy on Grace, but also unreasonable. The problem is that morally irrelevant prudential considerations also make this claim difficult to defend. This is because COST is committed to the normative significance of prudential reasons – or, at least, COST is committed to the normative significance of prudential reasons if it is to have any chance of providing rationalist-friendly interpretations of cases like *The Phone Call*.

To see this, first consider a variation of *The Phone Call* where, instead of being asked to drink bleach, Mary is merely asked to click her fingers in order to prevent Xavier from dying a horrible death. Since performing this action is not costly, Mary would be morally required to perform the morally best action and click her fingers according to any plausible version of COST. For this claim to be compatible with moral rationalism, the defender of COST would also need to claim that it is unreasonable for Mary to refuse to click her fingers in this case. Now suppose that we start raising the cost to Mary of saving Xavier until we reach the variation discussed above that was particularly problematic for agent-neutral consequentialism:

> Unless Mary drinks the entire bottle of undiluted bleach that she keeps under the sink, which will result in her death, Xavier will die an equally horrible death. Xavier is identical to Mary in every relevant respect except one. The difference is that, if he survives, Xavier will live a marginally better life – the equivalent of an extra millisecond of the mildest pleasure – than Mary would if she survived.

It is implausible that Mary does not have sufficient reason to refuse to drink the bleach in this case. Since, according to agent-neutral consequentialism, Mary is morally required to drink the bleach, this verdict about Mary's reasons for action cannot be accounted for by an agent-neutral consequentialist in a way that is compatible with moral rationalism. By claiming that the costs to the agent are relevant to the permissibility of Mary's action – and more specifically that these costs make it morally permissible for Mary to refuse to perform the morally best action – COST can explain why it is reasonable for Mary to refuse to drink the bleach in a way that is compatible with moral rationalism. What, then, explains why it is unreasonable for Mary to refuse to click her fingers but reasonable for Mary to refuse to drink the bleach? It can only be that the prudential costs to Mary make a difference not only to whether the action is morally permissible, but also to whether it is reasonable. In short, for COST to have any chance of

plausibly vindicating moral rationalism, it must be the case that prudential costs make a difference not only to the moral permissibility of an action, but also to whether an agent has sufficient reason, all things considered, to perform or refuse to perform an action.

To be clear, this point is not supposed to be surprising or controversial. It amounts to nothing more than a claim that I have been assuming up to this point, which is that prudential considerations are genuine reasons for action. As I discussed in the preliminary section on prudential reasons, this claim is intuitive, and if moral rationalism could only be vindicated by rejecting this claim, then that would itself be a good reason to reject moral rationalism. The problem for COST is that, when we combine this claim with the idea that there are morally irrelevant prudential reasons, we end up with the result that an agent can have sufficient reason, all things considered, to perform an immoral action. This is because there will be cases where the prudential costs to the agent that make it reasonable for them to refrain from performing the morally best action will be morally irrelevant prudential costs. In those cases, since these costs do not make a difference to the moral status of the action – since they cannot flip the status of the action from morally impermissible to morally permissible – the agent will be morally required to perform the action that is morally best. From the moral point of view, it will be as if the agent can perform the morally best option at no cost.

This argument does not only support the claim that endorsing COST does not allow a moral rationalist to respond to every purported counterexample to moral rationalism in a plausible way. Accepting that there are morally irrelevant prudential reasons should lead anyone to reject moral rationalism. This is because moral rationalism will be implausible in any case where morally irrelevant prudential reasons conflict with a moral reason when that moral reason would fail to generate a moral requirement if the morally relevant costs to the agent were significantly lower than the morally irrelevant costs actually are. In such a case, since the irrelevant reasons don't make a difference to the moral status of the action but do make a difference to what it is reasonable for the agent to do, the moral reason will generate a moral requirement that the agent will have at least sufficient reason, all things considered, to violate.

Similar points suggest that the prospects of offering a plausible rationalist-friendly interpretation of all variations of *The Voyeur* are slim. This is because, as we keep ramping up the prudential benefits that Tom derives from spying on Grace, it must be the case that we are moving closer and closer to this action being reasonable. But, since these are morally irrelevant

prudential considerations, it cannot be the case that we are moving any closer to the action being morally permissible.

.

VII

I will now consider some responses to this argument. The first two responses that I will consider claim that, in addition to not making a difference to the moral permissibility of an action, considerations such as the pleasure that Tom will receive from watching Grace also fail to make a difference to whether an action is reasonable.

One view along these lines holds that, while prudential considerations do provide us with genuine reasons for action, pleasure derived from morally distasteful sources does not benefit the agent. We can see how implausible this view is by noting what it is committed to saying about *The Voyeur*. Recall that, if Tom continues to spy on Grace, then he will continue to derive a great deal of pleasure from this activity, and his life will have a sense of purpose. If he does not, then he will never experience anything close to the same amount of pleasure again, and his life will feel empty. On the view we are now considering, the fact that only one available course of action will lead to pleasure and fulfillment is entirely irrelevant to the question of what life is better for Tom. No plausible theory of welfare could have this implication.

There is a better view in the same ballpark. This view accepts that deriving pleasure from a morally distasteful source can makes an agent better off, but it denies that these prudential benefits provide an agent with reasons to act. Since this claim goes against an assumption that we are making – that prudential considerations are genuine non-moral reasons – I will discuss it in less detail than it perhaps deserves.

The first point to note is that, as far these things go, the idea that agents have reasons to perform actions that benefit them is about as close to self-evident as we can get. We can see this by noting that, when we are making decisions about what to do, we simply take it for granted that the fact that performing an action will be good for us counts in favour of performing that action to at least some degree. The same cannot be said of our reasons to act altruistically, or of our reasons not to perform immoral actions. Second, once we grant that morally distasteful sources of pleasure can benefit the agent, and that prudential considerations at least sometimes provide an agent with non-moral reasons to act, it is very unclear how it could be explained why these benefits do not provide Tom with reasons to act. Since these benefits are morally objectionable,

it makes sense that they wouldn't contribute to the moral permissibility of an action. But we are not talking about morality here. Finally, as with the previous response, this view has counterintuitive implications. Similar to that response, it is committed to saying that Tom has no reason at all to choose the life in which he is happy and fulfilled over the life in which he is miserable and empty. That is implausible.

VIII

A different kind of response may be tempting. This goes something like this: In *The Voyeur*, Tom acts in a way that many would consider *seriously wrong*. This may make it seem plausible to say that no matter how much pleasure Tom derives from spying on Grace – and even if this makes a difference to how he should live, all things considered – it remains unreasonable for Tom to spy on Grace due to the normative significance of the moral requirement that he must violate in order to derive this pleasure. A similar reply may be tempting, of course, in response to other examples that involve comparable or even more egregious moral violations.

There are reasons to doubt this view. But even if it is right that there are some moral reasons that are so strong that no prudential benefit – whether morally relevant or irrelevant – can make it reasonable for an agent to ignore them, this response severely underestimates the challenge that morally irrelevant reasons create for moral rationalism.

To see this, we can first note that, as I have understood them, a moral reason to Φ is a *pro tanto* moral requirement to Φ. In other words, if you have a moral reason to Φ, then, absent defeaters, you are morally required to Φ. This claim is true of our moral reasons not to spy on someone, or not to murder someone, but also true of our moral reasons to perform other actions that are much less morally significant than these. It seems plausible, for instance, that I have a moral reason to meet you for coffee at 3pm if I promised to meet you for coffee at 3pm.

The reason that this creates a problem for vindicating moral rationalism even if we accept this response to *The Voyeur* was gestured at above, but it is worth going over more carefully. To start, take whatever you consider to be the weakest moral reason that you can have to perform an action. Let's suppose, for illustration, that this is a moral reason that would fail to generate a moral requirement even if the morally relevant cost of performing the action was the equivalent of suffering a very minor and brief headache. Now suppose that the cost to the agent of performing this action is very significant. Let's say that it is the equivalent of suffering a

week-long migraine. But suppose that this is a morally irrelevant prudential cost. The result of this will be that the agent is morally required to perform an action that is, by our own lights, normatively insignificant. And this will be so despite the high cost to the agent of performing this action. The reason for this, of course, is that these prudential costs do not make a difference to the moral permissibility of the action. Now, given that it would be reasonable for the agent to fail to perform this action if it would lead to the equivalent of a minor and brief headache, it is implausible that it is unreasonable for the agent to fail to perform this action if it would lead to the equivalent of a week-long migraine.

Unlike cases like *The Voyeur,* there is no hope of explaining these cases in a rationalist-friendly manner by appealing to the normative significance of the moral requirement. We ourselves don't think that these moral requirements are all that significant. This conclusion can be bolstered by noting that, in many relevant respects, these cases are similar to the cases that led us to reject the plausibility of vindicating moral rationalism while accepting agent-neutral consequentialism. In those cases, as in these cases, it doesn't seem plausible that the agent lacks sufficient reason to refrain from performing the morally best action given that the moral reasons that support the morally best action are very weak and the non-moral reasons that support the alternative are very strong.

IX

A final point is worth discussing before concluding. Recognising that there are morally irrelevant prudential reasons allows us to see what is wrong with an increasingly popular way of attempting to vindicate moral rationalism. The clearest proponent of this view is Portmore (2011).[60] In line with what I have argued, Portmore believes that there are moral and non-moral reasons, and that prudential reasons are non-moral. He also believes that prudential reasons are normatively significant, that there are conflicts between moral and prudential reasons, and that prudential reasons can provide an agent with sufficient reason to act in ways that are not morally best. What is distinctive about Portmore's view – and what secures moral rationalism on his account – is his conception of moral deontic statuses. On this view, an action is morally permissible whenever an agent has sufficient reason, all things considered, to perform this action, and this is so whether it is moral or prudential reasons that give the agent sufficient

[60] As far as I know, this view was first defended by Portmore. In his recent work, this is also how Michael Smith secures moral rationalism. See for example (2018).

reason to act. This conception of moral permissibility guarantees moral rationalism since, in every case where an agent has sufficient reason to act in some way on the basis of prudential reasons, that action will come out as morally permissible.

This conception is not supposed to be stipulative. Portmore motivates it by considering cases that are similar to some of those that we considered when discussing *The Phone Call*. In his central case, Professor Collins has office hours at 2pm. Before he leaves his home for campus, he receives a call from a friend who lets him know that there is a rare opportunity to meet Hall-of-Fame baseball player Reggie Jackson. Jackson is Collins' hero, and meeting him would be one of the highlights of his life. As it happens, it is too late for Collins to cancel his office hours. As a result, if Collins doesn't show up, a student may end up waiting for him in vain. About this case, Portmore claims that it is intuitive that Collins has sufficient reason, all things considered, to meet Jackson. He also claims that it is intuitive that, due to the prudential benefits that Collins will derive from meeting Jackson, it is morally permissible for Collins not to show up for his office hours. Summing up his views, Portmore (2011, 53) writes:

> The explanation, then, for why Collins is not morally required to hold office hours seems to be that the self-interested and non-moral reason that he has to meet Jackson, being sufficiently strong, morally justifies his failing to fulfill his commitment to hold office hours. Thus, we seem to be implicitly committed to moral rationalism—that is, to the idea that an agent cannot be morally required to do what she has sufficient reason not to do. If instead moral rationalism were false, we would expect that if we were to take a case in which an agent has decisive reason to fulfill a moral requirement and then imagine a series of variants on that case in which the agent has ever stronger non-moral reasons to do something else, we would eventually arrive at a case in which the agent had sufficient reason to violate that moral requirement. But we never do arrive at such a case. For as soon as we are willing to say that the agent has sufficient reason to do that something else, we are no longer willing to say that she is morally required to refrain from doing that something else.

As far as it goes, Portmore's claims about Collins are plausible. Intuitively, it does seem that Collins' prudential reasons to meet Jackson give him sufficient reason to ignore his office hours, and that this is morally permissible. The problem with the argument is similar to a problem that we have already seen for COST. This is that Portmore generalises from a case

involving morally relevant prudential reasons to prudential reasons in general. This move only succeeds if all prudential reasons are morally relevant. As we have seen, this is not so. And when we consider a variation on the Collins case involving morally irrelevant prudential reasons, Portmore's claims are not plausible. Suppose, for instance, that Collins* is contemplating skipping his office hours, but the reason for this is that he derives a great deal of pleasure from the mere thought of a student showing up and waiting for him in vain. He likes to picture the student reacting to every movement in the hope that it is finally Collins, and to imagine the anxiety and annoyance that the student will feel as the minutes keep ticking by and he must repeatedly decide whether he will wait another five minutes or give up and leave with his problems unsolved. Let's suppose that, due to his love of making his students suffer, Collins* gets as much prudential benefit from this activity as Collins does from meeting Jackson. In this case, it is not plausible that the self-interested and non-moral reason that Collins* has, being sufficiently strong, morally justifies his failing to fulfil his commitment to hold office hours. Thus, *contra* Portmore, we do not seem to be implicitly committed to moral rationalism. For we would not be willing to say that as soon as this self-interested and non-moral reason makes it reasonable for Collins* to skip his office hours, it also makes it morally permissible for Collins* to skip his office hours. Given this, we should reject this conception of moral permissibility and any argument for moral rationalism that depends on it.

X

Let's take stock. I began by discussing *The Phone Call*. In this case, Mary must choose between sacrificing her own life in an excruciating way or allowing many other people to die in excruciating ways. Cases of this kind put significant pressure on moral rationalism because they suggest that some of our core commitments about reasons for action and the content of morality do not line up in the way that moral rationalism requires. These commitments can only be secured if we endorse moral anti-rationalism. Cases like *The Phone Call* serve another purpose. They allow us to see that, if agent-neutral consequentialism is true, then we should reject moral rationalism. This does not show that we should reject moral rationalism, but it does show that, if we want to vindicate moral rationalism, we will need to endorse a view about the content of morality that does not require an agent to always perform the action that is morally best. We then considered a prominent example of such a view. This is COST, which says that, if performing the morally best action would be sufficiently costly for the agent, then it is morally permissible for the agent to fail to perform this action. There are reasons to doubt

that COST can plausibly explain all cases like *The Phone Call* in a rationalist-friendly manner, but the bigger problem for this view is cases like *The Voyeur*. Through our discussion of *The Voyeur*, we saw that there are costs to the agent that don't make a difference to the moral permissibility of an action. Given that these costs do make a difference to what an agent has sufficient reason to do, a moral rationalist who endorses COST cannot plausibly explain why there are no acute conflicts like *The Voyeur* where the agent has sufficient reason, all things considered, to act immorally.

Since agent-neutral consequentialism and COST are not the only available moral theories, this result does not entail that no moral theory could be used to vindicate moral rationalism. Our discussion of *The Voyeur*, however, gave us reason to doubt that any plausible moral theory could vindicate moral rationalism. If it is implausible that the enjoyment that Tom gets from watching Grace in the shower is relevant to whether it is morally permissible for him to watch her in the shower, then no plausible moral theory will say that this is a morally relevant consideration. But once we accept that there are morally irrelevant prudential considerations, we cannot explain why there are no possible cases where these morally irrelevant considerations give an agent sufficient reason, all things considered, to act immorally. This is particularly clear when morally irrelevant prudential reasons conflict with what, by a theory's own lights, are unimportant moral requirements.

There is a more general way to put the point about morally irrelevant prudential considerations. What these illustrate is that there are considerations that are relevant to whether you should live your life in a certain way that are entirely irrelevant to whether it is morally permissible for you to live your life in that way. As I will next argue, morally distasteful prudential reasons are not the only morally irrelevant considerations. Excellence-based reasons are also morally irrelevant. This way of putting the point perhaps best reflects what I believe is the deepest reason to reject the idea that morality always provides us with decisive reason to act: There are many worthwhile but incompatible ways that we could live our lives. We cannot, at least without serious distortion, convince ourselves that all these possible lives are morally acceptable. Living a morally decent life is certainly worthwhile, but it can come at a price. It may be reasonable to refuse to pay this price given what else is on offer.

CHAPTER THREE: MORALITY AND EXCELLENCE

This chapter returns to excellence-based reasons. I will first argue that the same basic idea that allowed us to defend moral anti-rationalism in the previous chapter – that there are considerations that are relevant to whether you should live your life in a certain way that are irrelevant to whether it is morally permissible for you to live your life in that way – can also be used to defend moral anti-rationalism on the basis of excellence-based reasons. I will then discuss some unique problems that EBRs create for the plausibility of moral rationalism.

I

It is useful to first take a step back. The reason that reflecting on *The Voyeur* led to the arguments that it did for AR is that morally distasteful prudential reasons meet four conditions. They are:

> (i) *Non-Moral Reasons:* They cannot generate moral requirements.
>
> (ii) *Morally Irrelevant Reasons:* They cannot make an action that would otherwise be morally impermissible morally permissible. They cannot, in other words, prevent a moral reason from generating a moral requirement.
>
> (iii) *Normatively Significant:* They are genuine reasons for action. They make a difference to, or are relevant to, how we ought to live our lives, all things considered.
>
> (iv) *Conflict:* They can conflict with moral requirements. They can give an agent a reason to perform a morally impermissible action.

Since these prudential reasons meet these four conditions, we can make arguments like these: Take the weakest moral reason there is to Φ. Now suppose that an agent has this reason to Φ and a morally irrelevant reason to perform an incompatible action (Ψ). Since it is morally irrelevant, this reason to Ψ cannot make it the case that Ψ-ing is morally permissible, and hence the agent is morally required to Φ despite having this reason to Ψ. Now suppose that the reason to Ψ is normatively significant. To use a prudential example, suppose that failing to Ψ would lead to a life of misery. Given that a much weaker morally relevant prudential reason would clearly make it both permissible and reasonable for the agent to Ψ, it is not plausible that this

stronger prudential reason to Ψ fails to make it reasonable for the agent to Ψ. As such, the agent has sufficient reason to Ψ despite being morally required to Φ – and that is just moral anti-rationalism. More generally, whenever it would be permissible and reasonable for an agent to fail to comply with a moral reason to Φ on the basis of a morally relevant non-moral reason to Ψ, it will be reasonable for the agent to act immorally on the basis of a morally irrelevant reason to Ψ so long as the morally irrelevant reason to Ψ is on a par with, or stronger than, the morally relevant reason that would have rendered it reasonable and permissible to Ψ. Here is another way to think about this: Start with a case where an agent has a moral reason, of whatever strength, to Φ and a morally irrelevant non-moral reason to Ψ. Now suppose that we just keep ratcheting up the normative significance of the agent's non-moral reason to Ψ. Assuming that these non-moral reasons actually matter to how an agent should live, then it will eventually become implausible to deny that the agent has sufficient reason, all things considered, to comply with the non-moral reason.[61] But, since this is a morally irrelevant reason, it will not be the case that, at this point, the agent is no longer morally required to Φ. As such, at this point, the agent will have sufficient reason to act immorally.

These arguments could be made with any reason that meets the four conditions above.[62] Given this, if it can be shown that EBRs also meet these four conditions, then these same arguments – assuming that they are cogent – would also show that EBRs can provide an agent with sufficient reason to act immorally. The sections that follow will argue that EBRs do meet these conditions. I will keep this relatively brief since many of the claims that are needed have already been discussed.

II

As I argued in Chapter One, EBRs are non-moral reasons. That they are also morally irrelevant reasons can be seen by considering a variation on *The Voyeur*. For ease of reference, here is the original case:

---

[61] Unless, perhaps, it is conflicting with a particularly strong moral reason. See Chapter Two Section VIII.

[62] Though whether a particular reason, or type of reason, meets these conditions will be more convincing in some cases than others. Some may find it more plausible that morally distasteful prudential reasons meet these conditions than that EBRs do, and vice versa. For the purpose of defending AR, it just needs to be the case that *some reason or other* meets these conditions. For the purpose of this thesis being interesting to read, I hope that at least one of these two kinds of reasons strike you as a good candidate.

> *The Voyeur:* Tom has recently moved into a new apartment. One afternoon he discovers that he can see clearly into the apartment across from him. A beautiful woman, Grace, happens to live there. He can see everything from her most mundane to her most private activities. Due to the extremely dark tinting on his windows, she is completely unaware that he is watching. Over time, Tom becomes obsessed with watching Grace; this activity becomes the centre of his life. This is not so in the compulsive way that oxycodone might become the centre of someone's life, but in the reflectively endorsed way that admiring beautiful works of art might become the centre of someone's life. Tom derives a great deal of pleasure from this activity, and, to the extent that it interferes with his other pursuits, he is always willing to make the trade-off. In short, nothing matters to Tom nearly as much as watching Grace.

Tom's actions in *The Voyeur* are prudent but morally impermissible. We can next add this detail:

> *The Aesthetic Voyeur:* The case is otherwise the same as before, but, in this version, Tom does not only watch Grace, he also takes photographs of her. This does not bring him any additional prudential benefits over just watching her.

The fact that Tom is photographing Grace does not seem, by itself, to make his actions morally permissible. We can now run a similar test as in the previous chapter: We can first imagine a version of this case where Tom's photographs of Grace are terrible. Call this *version one*. We can then imagine a version of the case where Tom's photographs of Grace are aesthetically excellent. Call this *version two*.

We can next ask: With the stipulation that Tom's photographs in version two are aesthetically excellent, is it the case that Tom's actions in version two are closer to morally permissible than his actions in version one? If EBRs are morally relevant non-moral reasons, then they should be. They do not, however, seem to be. The fact that in version two Tom's photographs of Grace make good use of negative space, and that in version one they are badly out of focus, seems like a morally irrelevant fact. This conclusion is reinforced by noting that it would seem to make sense for Grace to feel the same degree of resentment in both cases, and for Tom to come to feel the same degree of guilt. The fact that the photographs he takes of Grace in version two are beautiful, then, is a morally irrelevant consideration; it is not a fact that makes a difference to the moral permissibility of his actions.

The claim that EBRs are morally irrelevant can be further supported by considering a different kind of case which is similar to one discussed earlier:

> *Two Mathematicians:* In their quests to achieve intellectual excellence, two mathematicians – Madeline and Elizabeth – perform almost identical actions throughout their lives, and they do so with the same intentions. The only difference between their acts is that, when they sit down to do mathematics, Madeline discovers a proof as powerful and elegant as Gödel's Incompleteness Theorem, whereas Elizabeth produces work of no mathematical importance.

Suppose that Madeline and Elizabeth's work has equivalent instrumental value. And suppose that, for whatever natural or supernatural reason, it was necessary for both Madeline and Elizabeth to perform morally impermissible actions to complete their work. If EBRs are morally relevant reasons, then Madeline's immoral actions are closer to permissible than Elizabeth's due solely to the intellectual value of her work. This is, after all, what provides her with an EBR that Elizabeth lacks, and there are no other differences between them. But the idea that the elegance of Madeline's proof is a morally relevant difference between her actions and Elizabeth's is implausible. Possessing mathematical talent does not provide you with a moral license to treat people worse than someone who does not possess this talent.

As with certain kinds of prudential reasons, then, excellence-based reasons are morally irrelevant reasons. Though we don't require an explanation for the purposes of the argument, it is evident that something very different explains why EBRs are morally irrelevant than what explains why morally distasteful prudential reasons are morally irrelevant. Part of the explanation may be related to the idea just expressed. Whether we have the capacity to produce excellent aesthetic or intellectual work is not up to us; it is a matter of luck. And it doesn't seem to fit very well with our conception of what morality is like to claim that, when doing so would have no instrumental benefit for ourselves or others, morality allows agents who happen to possess this capacity to perform actions that it would be impermissible for agents who happen not to possess this capacity to perform. This seems elitist and unfair. I suspect that something else is also going on here. In contrast to prudential value, aesthetic and intellectual value just don't seem like the right kinds of thing to undermine a moral requirement. This may be why it doesn't just seem implausible but odd to claim that how interesting a photographs composition is, or how elegant a mathematical proof, is a morally relevant fact. To try to fold such

considerations into morality – to claim that they are morally relevant – seems like an attempt to jam something into a place where it neither fits nor belongs.

III

Even if one accepts that EBRs are morally irrelevant reasons, moral rationalism can be defended by claiming that EBRs are not genuine reasons – or perhaps by claiming that, even if they are genuine reasons, they matter very little. This would explain why the aesthetic value of Tom's work, or the intellectual value of Madeline's work, does not seem to make their morally impermissible actions closer to morally permissible.

This response has the same problem as the analogous response to the prudential version of the argument. As we saw in Chapter One, the claim that EBRs are genuine reasons is compelling. The mere fact that Kierkegaard's decision to leave Olsen allowed him to write *Either/Or* seems to count in favour of his decision. Similarly, the fact that Hume putting words on a page led to *A Treatise of Human Nature* – or that James doing so led to *The Golden Bowl* – seems like a strong reason to put words on a page. The claim that EBRs are genuine reasons is made even more convincing when we consider comparison cases. It seems clear, for instance, that the fact that Madeline will discover an important mathematical proof, whereas Elizabeth will produce nothing of value, gives Madeline a reason to spend her time doing mathematics that Elizabeth lacks. Comparison cases also help to demonstrate that EBRs can provide agents with reasons of considerable strength. We can see this by imagining a person who, like Kierkegaard, sacrificed significant happiness and a loving relationship to become a great writer, but who, unlike Kierkegaard, failed to write anything good. Given the price he paid, it seems like it was a terrible mistake for this person to attempt to become a great writer. Even though he paid the same price, the fact that Kierkegaard succeeded – that fact that he wrote *Either/Or* – seems to go a long way towards counterbalancing these significant prudential costs. It is far from clear that his decision was a mistake.

IV

Cases like *The Aesthetic Voyeur* demonstrate that, at least in principle, EBRs can conflict with moral reasons. Given that EBRs are non-moral, morally irrelevant, and normatively significant, we now have all the components necessary to run the arguments in Section I with excellence-

based reasons. If these arguments are persuasive, we should accept that EBRs, like certain prudential reasons, can provide an agent with sufficient reason, all things considered, to act immorally.

There is a potential difference between EBRs and prudential reasons that is worth addressing. When it comes to conflicts between morality and prudence, it is widely accepted not only that such conflicts are possible, but that they are a feature of our actual normative lives. This is one reason why the 'Why be moral?' question is typically illustrated with conflicts between morality and prudence, and it is also a reason why this question has an existential, as opposed to a merely theoretical, grip on us. Some may wonder whether the same can be said of conflicts between morality and excellence. Do these only arise in idealised examples like those discussed earlier, or are our conclusions about such conflicts likely to actually matter to how someone should live? Though it is difficult to be precise, I will give some reasons for thinking that conflicts between morality and excellence are likely to occur in some people's actual lives.

To illustrate these, it is useful to start on ground that is less likely to lead to first-order disputes that could distract us from the issue at hand. This is with some reasons to think that conflicts between excellence and prudence are likely to actually occur. I will then suggest that, given the causes of these conflicts, there will almost certainly be analogous conflicts between morality and excellence regardless of what precisely we think morality demands.

Perhaps the most obvious reason that there are going to be conflicts between prudence and excellence is that achieving excellence typically takes time and energy, and sometimes a lot of time and energy. As discussed, in order to reach the heights that he did, Kierkegaard often wrote for sixteen hours a day, and others, such as Balzac, have taken on similarly fanatical schedules. Living a flourishing life also requires time and energy. Many of us have a diverse range of desires and interests which can only be satisfied by living a more relaxed life that involves indulging in a wide variety of activities. Our happiness may also depend on having a close circle of friends, a loving relationship, or an involved family life. Maintaining these relationships also requires time and energy. This is not to say that it is, for everyone, impossible to achieve excellence while pursuing a range of interests and cultivating various relationships. It is just to say that the way that some people would need to structure their lives to achieve excellence – or even just to produce the best work that they are capable of – would not allow them to perform the actions that are necessary for them to be happy. For such a person,

happiness and the achievement of excellence are incompatible. For lack of a better term, we can call these *pragmatic* reasons why excellence and prudence will sometimes conflict.

Of course, not every conflict between prudence and excellence will arise because there are literally not enough hours in the day to perform both the actions that are necessary to achieve excellence and the actions that are necessary to achieve happiness. There will often be psychological factors at play. Consider the decision to have children. About this, literary critic Cyril Connolly (1938, 116) famously wrote that 'there is no more somber enemy of good art than the pram in the hall.' If we think in purely pragmatic terms, this claim is farfetched. After all, one can always neglect or abandon one's children. Merely having children is rarely going to prevent one from producing good art. But if someone is, for whatever reason, motivated to be a *good parent*, then this could prevent them from producing good art given what this might demand. This situation will not always lead to a conflict between prudence and excellence since one may find parenthood a miserable experience, or one may be driven to be a good parent by a crushing sense of duty. But it seems that some cases will involve such a conflict. A person who is driven to be a good parent by love or affection may well find parenthood more fulfilling than they would find performing the actions that are necessary to produce excellent work. Another factor may also come into play. Certain life experiences can transform our priorities – they can cause a transformative experience[63] – and parenthood is a paradigm case. After having a child, a person may simply care much less, or not at all, about achieving excellence. In cases where a person would be significantly happier after going through this transformative experience, there will be a conflict between prudence and excellence. Such a person will have EBRs to refrain from performing the actions that will bring about this transformative experience.

None of these claims are specific to parenthood. There are many possible life experiences that could change our priorities. In any case where this change would cause us to care less about, or not at all about, achieving excellence when we have the ability to do so, we will have an excellence-based reason to avoid these experiences if we can. At least some of these experiences would be good for us. Travelling the world, or falling in love, may make our lives significantly better, but it could also make surrealist poetry, or the metaphysics of composition, seem comparatively boring and unimportant.

[63] For a detailed and fascinating discussion of this phenomena, see Paul (2014).

There are other kinds of conflict that seem particularly deep. Reflecting on his failed marriage to Vivienne Haigh-Wood, T.S. Eliot (2011, 29) wrote: 'To her the marriage brought no happiness… to me, it brought the state of mind out of which came *The Waste Land*.' In other words, if T.S. Eliot had not been in the miserable state of mind that he was, he would not have written *The Waste Land*. More generally, an agent may be unable to write a great aesthetic or intellectual work unless they are in some particular state of mind, and this state of mind could be one that is prudentially bad for them. This could be a state of mind that is intrinsically bad for them, as with Eliot, or it could be a state of mind that leads to their life falling apart in other respects. An example of the latter may be Stanisława Przybyszewska, who was so obsessed with the French Revolution – and Robespierre in particular – that she neglected virtually every other aspect of her life. One result of this obsession was that she lived in squalor and eventually became so emaciated that she could no longer hold a pen. Another result of this obsession was that she was able to write *The Danton Case*, which is often considered one of the best literary works about the French Revolution. What makes this work so admired is its level of psychological insight into the actors involved, a level of insight that she may not have been able to develop without this kind of obsession. There is another reason that an agent may be unable to achieve excellence if they are in a prudentially good state of mind: If your life is going very well, or if you are feeling particularly happy and content, you may lose the motivation to write great poetry or great philosophy. This is because you might be motivated to perform the actions that are necessary to achieve excellence by a sense of *dissatisfaction*. This could simply be the desire to get out of bad situation, or it could be that you are driven by envy, or by the sense that you need to prove something to others, or that you need to get back at someone who you perceive to have slighted or disrespected you.

These examples show that, while a person may be able to live a flourishing and happy life, and while they may be able to live a life where they achieve aesthetic or intellectual excellence, they may not be able to live both these lives. In such a case, excellence and prudence will conflict. As I hope is clear, there are many possible moral analogues to these conflicts. The exact details will depend on what we take to be morally required. But here as some suggestions.

As with living a happy life, living a morally decent life requires time and energy. This is most obvious if we suppose that there are positive duties of beneficence – whether perfect or imperfect – but we don't need to assume this. Return to the claims about parenthood. It is plausible that it is typically morally impermissible to neglect or abandon your children. But it is also plausible that, for the same pragmatic and psychological reasons noted above, the

demands of parenthood could prevent someone from achieving excellence. The plausibility of both these claims perhaps explains why one of the most prominent examples of a putative conflict between morality and excellence is of this kind. This is Williams' (1981a) example of Gauguin's decision to abandon his family and sail to Tahiti to paint. There are other examples along these same lines, some more extreme. Rousseau, for instance, left all five of his children in a Paris orphanage despite the slim prospects of survival this gave them. Again, there is no need to focus on parenthood. We seem to acquire many of our moral duties from our various associations – whether this is with friends, family, colleagues, acquaintances, or institutions – and many of these duties require time and energy to discharge.

Putting aside moral obligations that arise from associations, doing what is morally required could easily change our priorities. One way that this could happen is if doing what is morally required brings us into direct contact with the less fortunate, or in some other way make us vividly aware of the suffering in the world. This awareness may make some of our own projects seem comparatively trivial if we are unable to convince ourselves that these projects make the world a better place. Further, just as an agent may need to be in a prudentially bad state of mind to produce excellent work, an agent may need to be in a state of mind that is morally impermissible to produce excellent work. This could be the case if we are morally required to regard people as, for example, ends in themselves, or as worthy of dignity and respect. A person's ability to write a great tragedy might depend on them being in a misanthropic or nihilistic state of mind. Even if there are no moral requirements to have certain attitudes, having attitudes that would be helpful from the point of view of achieving excellence could lead someone to treat others impermissibly. Since it could help us to achieve the necessary degree of focus, we might be more likely to achieve excellence if we instrumentalise others or are indifferent to their concerns. It doesn't seem morally impermissible, for example, to be so obsessed with the French Revolution that you allow yourself to almost starve to death. But this level of obsession, which may be necessary for you to write a great play about the French Revolution, could lead you to perform immoral actions if you find yourself in certain kinds of circumstances. It could easily, for instance, lead you to ignore someone's cry for help.

There are many other parallels that could be drawn, but this is hopefully enough to demonstrate the point: Just as someone might actually face a conflict between morality and prudence, so they might actually face a conflict between morality and excellence. And just as the price of the achievement of excellence could be your happiness, or vice versa, the price of a morally decent life could be the achievement of excellence or happiness or both.

V

The previous section discussed one respect in which prudence and excellence are not that far apart. More generally, the argument that I have given for AR has the same structure in the case of both prudential reasons and excellence-based reasons. This, however, does not mean that there are no significant differences between EBRs and prudential reasons when it comes to the plausibility of moral rationalism.

The most basic and important difference between EBRs and prudential reasons is simply this: The kinds of lives that are licensed by EBRs providing agents with sufficient reason to act will often be very different from the kinds of lives that are licensed by prudential reasons providing agents with sufficient reason to act. If we suppose, as I have argued, that there are conflicts between excellence and morality just as there are conflicts between prudence and morality, then accepting moral rationalism will commit one to claiming that, in this distinct set of cases, agents act unreasonably when they achieve aesthetic or intellectual excellence. This claim may be harder to accept in cases of conflict between morality and excellence than it is in cases of conflict between morality and prudence.

Here is one general reason to think that this is so. When it comes to assessing actions from the point of view of what an agent has most reason overall to do, we make evaluative distinctions within the more general categories of 'reasonable' and 'unreasonable'. Any unreasonable action is a mistake, and a life that is composed of many mistakes will be regrettable. But mistakes come in different magnitudes. We recognise a distinction between a person who has consistently made small mistakes and a person who has made a complete mess of their lives. Even though both have failed to live as they ought to live, all things considered, we are much more inclined to say that the latter person has (e.g.) *wasted* their life. Similar points apply to the idea that an action is 'reasonable'. No reasonable action is a mistake, but there is an intuitive difference between a person who lives what we might call an *acceptable* or *merely reasonable* life and a person who lives what we might call an *esteem-worthy, admirable* or *attractive life*.[64]

We can next note that, even if we typically think that prudent action is reasonable, we do not take acting in our own interests to itself be worthy of admiration or esteem. Since prudentially

[64] These claims are not supposed to be controversial or original. The distinction between doing something that is merely reasonable and doing something that is esteem-worthy, for instance, is structurally identical to the distinction between doing something that is merely morally required and doing something that is supererogatory. A similar distinction seems to be present in other domains. It has been argued, for instance, that there are cases of *rational supererogation*. See Slote (1986) and Benn & Bales (2020) for examples of this argument.

good lives often have independently attractive components, this point may not be immediately clear. It is easiest to see when we consider someone along the lines of Rawls' grass-counter. We can imagine, for instance, a person whose only path to happiness – and only way to avoid misery – is to endlessly watch reruns of *Wheel of Fortune*. If we ourselves find happiness elusive, we may envy such a person for the ease with which they are able to attain it, but we are unlikely to admire them or to hold them in high esteem. Nor are we likely to describe their life as attractive or interesting.

The situation is different when we turn to EBRs and the successful pursuit of aesthetic and intellectual excellence. This does not merely seem to be an acceptable way to live; we think of it as an attractive and interesting life. We admire and esteem great artists and intellectuals. Many of us spend significant time studying the lives and works of great philosophers, or great musicians, and some people commit their lives, or at least careers, to this undertaking. This does not seem like a waste of time. Nor does it seem to be only as worthwhile as the happiness that it brings, which is how we are likely to evaluate a life spent endlessly watching *Wheel of Fortune*. As previously noted, the flavour of this admiration and esteem is not moral. There is no connection between admiring H.P. Lovecraft as a writer and thinking that he was a morally good, or even a morally decent, person. We admire and esteem him, when we do, because we believe that, in writing *The Rats in the Walls*, or *The Call of Cthulhu*, Lovecraft wrote excellent pieces of fiction.

This difference between EBRs and prudential reasons is relevant to the plausibility of moral rationalism for the following reason: If it is right that there are conflicts between morality and excellence, then accepting MR would commit one to the claim that it can be unreasonable for an agent to pursue an attractive and interesting life that is worthy of admiration and esteem. That is hard to believe. It is made even less plausible when we recognise that the price of an agent not pursuing this life could be the next *Golden Bowl*, the next incompleteness theorem, or the next *Treatise of Human Nature*.

Similar ideas to this are, I suspect, behind some of the remarks that Bernard Williams (1981a, 23) makes when reflecting on Gauguin. He writes that:

> … while we are sometimes guided by the notion that it would be the best of worlds in which morality were universally respected and all men were of a disposition to affirm it, we have in fact deep and persistent reasons to be grateful that that is not the world we have.

If Gauguin had not abandoned his family, he never would have painted *Where Do We Come From? What Are We? Where Are We Going?* Given the aesthetic value of this work, we feel 'gratitude that morality does not always prevail – that moral values have been treated as one value among others, not as unquestioningly supreme' (1981a, 37). If we believe that the *Confessions* has a similar or greater level of intellectual or aesthetic value, we might feel the same about Rousseau's decision to leave his children at an orphanage.[65] Some may, of course, dispute the details of these examples. For one thing, Williams' claim is imprecise and, on one interpretation, compatible with MR. Moral rationalism, after all, does not claim that moral 'values' or moral reasons are supreme. The claim only concerns moral requirements. In addition, some may believe that it was not morally impermissible for Gauguin or Rousseau to act as they did, or that their actions were not necessary for them to achieve what they did, or that the actions were unreasonable regardless of what they achieved as a result. It is not, however, these specific examples but the general idea that is compelling. Moral rationalism commits us to the claim that, when the demands of morality and excellence conflict, the pursuit and achievement of aesthetic and intellectual excellence is always a mistake. To the extent that we value great intellectual and aesthetic works – and to the extent that we are attracted to lives of this kind and admire those who live them – this will seem like an implausible implication. Perhaps the strongest proponent of this idea is Nietzsche. He writes:

> What if a symptom of regression were inherent in the "good", likewise a danger, a seduction, a poison, a narcotic, through which the present was possibly living *at the expense of the future?* Perhaps more comfortably, less dangerously, but at the same time in a meaner style, more basely? So that precisely morality would be the blame if the *highest power and splendor* possible to the type man was never in fact attained? So that precisely morality was the danger of dangers?[66]

As I interpret him, it is because acting as morality demands can undermine the achievement of creative excellence that Nietzsche believed that, if you had the ability to achieve excellence – if you were one of the 'men of great creativity, the really great men according to my

[65] I return to these points in a slightly different, and more general, context in the final chapter.
[66] This quote is from Section 6 of the Preface to *On the Genealogy of* Morality. The translation is from Brian Leiter (1997, 264).

understanding' – morality could be for you the 'danger of dangers'.[67] This claim is hyperbolic, but it gets at something important. Even if it is not the danger of dangers, morality, when coupled with MR, is a danger to the achievement of excellence. If the next Rousseau lives as he ought to live according to this view, he might not, for this very reason, write the next *Confessions*. More generally, living as you ought to live according to this view is a danger to living an attractive life; it might be to blame if someone does not live a life of the 'highest power and splendor'.

Note that these concerns are independent of the specific argument for AR that we started with. They are general reasons to think that there is something distinctively implausible about moral rationalism when we reflect on conflicts between morality and excellence that is not captured by reflecting on conflicts between morality and prudence. As we have seen, there are also general reasons to doubt the plausibility of MR when we reflect on conflicts between morality and prudence. The claim that it is reasonable for Gyges to live 'like a god among humans' has intuitive appeal, and the claim that, in a case like *The Voyeur*, it would be a mistake for an agent to live anything other than a miserable life is not easy to believe. But, as discussed in the previous chapter, the situation is a lot worse for MR if my specific argument is correct. The same is true here. If EBRs are non-moral, morally irrelevant, and normatively significant, then there will be cases where MR has the implication that it is unreasonable for an agent to achieve aesthetic or intellectual excellence even when EBRs conflict with what are, by a theories own lights, relatively trivial moral requirements. As well as amplifying the above concerns, this alone gives us good reason to reject moral rationalism.

I will conclude this chapter by echoing some remarks that I made at the end of the previous chapter. These will perhaps have greater force now that EBRs have also been discussed. There are many worthwhile but incompatible ways that we could live our lives. We cannot, at least without serious distortion, convince ourselves that all these possible lives are morally permissible. Living a morally decent life is certainly worthwhile, but it can come at a price.

---

[67] My interpretation of Nietzsche has been heavily influenced by Brian Leiter's (1997; 2014) work. He may have an issue with this claim, however. As Leiter understands him, when Nietzsche is talking about 'morality', he is talking about a social and cultural phenomenon. On this view, Nietzsche always uses the term 'moral' with scare quotes. If this interpretation is correct, it is unclear whether Nietzsche is a moral anti-rationalist as I understand this view, since I am talking about the claim that you can lack decisive reason to do what you are in fact morally required to do. I believe that Nietzsche was an anti-rationalist in my sense, but it is difficult to say. (There is a similar difficulty in trying to figure out whether Williams is a moral anti-rationalist in my sense, especially when considering his more Nietzschean work, such as *Ethics and the Limits of Philosophy* (1985).) There are also interpretations of Nietzsche that are consistent with him being a moral rationalist. For an interpretation of Nietzsche as a moral perfectionist, see Hurka (2007).

This could be your happiness, the achievement of excellence, or both. It may be reasonable to refuse to pay this price given what else is on offer.

CHAPTER FOUR: THE BLAMEWORTHINESS DEFENCE OF MORAL RATIONALISM

This and the next chapter respond to two arguments that have been given for moral rationalism. These chapters are standalone chapters both in the sense that they are each self-contained papers and in the sense that they don't rely on my previous arguments for moral anti-rationalism. This is not to say that there are no connections between the ideas in the previous chapters and the ideas in these chapters. The argument that I give in this chapter, for instance, has a similar structure to some of the arguments that I gave in the last two chapters. There may also be explanatory connections between my claims in these chapters and my claims in the previous chapters. But my central aim in these chapters is just to show that these particular arguments fail to vindicate moral rationalism. My responses to these arguments could be accepted by someone who rejects my previous arguments, or even by a moral rationalist who believes that a different argument succeeds.

I

In this chapter, I argue against a prominent defence of moral rationalism. I call this the *blameworthiness defence* (BD).[68] A clear statement of the BD is given by Portmore (2011: 43-44):

1. If S is morally required to perform *x*, then S would be blameworthy for freely and knowingly performing ~*x*.
2. S would be blameworthy for freely and knowingly φ-ing only if S does not have sufficient reason to φ.
3. So, if S is morally required to perform *x*, then S does not have sufficient reason to perform ~*x*.
4. If S does not have sufficient reason to perform ~*x*, then S has decisive reason to perform *x*.

[68] The BD is endorsed by Darwall (2006a and 2006b), Portmore (2011), Skorupski (1999), and Kiesewetter (2017).

5. Therefore, if S is morally required to perform *x*, then S has decisive reason to perform *x* – and this is just moral rationalism.

This argument is valid. Premise 2, however, is false – or so I shall argue.[69] Following Portmore, I call this claim *blameworthiness entails lack of sufficient reason* (BELS).

II

I will begin with two clarifications. First, the relevant sense of blameworthiness in BELS is *moral blameworthiness*. This is important because, if we substitute moral blameworthiness with some other notion, the BD fails to establish MR. Consider, for instance, what we might call *rational criticisability*. An agent is rationally criticisable if she freely and knowingly performs an action that she has decisive reason not to perform. This notion is distinct from – and broader than – moral blameworthiness. We can see this by noting that an agent can be rationally criticisable without being morally blameworthy. This is because a person can face a choice between two morally permissible actions but have decisive reason to perform one of these actions. If they fail to perform this action, they will be rationally criticisable but not morally blameworthy. Suppose now that we interpret 'blameworthy' in the BD as rationally criticisable. Premise 1 would then simply be a statement of moral rationalism and BELS would be a tautology.

What, then, is moral blameworthiness? Following certain defenders of the BD[70] – and many others – I shall understand this in terms of *fitting attitudes*. On this view, a person is morally blameworthy for φ-ing only if certain attitudes would be fitting, or appropriate, responses to that agent φ-ing. The paradigmatic blame emotions include *resentment*, *indignation*, and *guilt*. In short, an agent is blameworthy only if it is fitting or appropriate for us to blame them, and for them to blame themselves. Notice that this provides another distinction between moral blame and rational criticism. When a person acts (say) akratically but morally permissibly, the sorts of attitudes that make sense – that are appropriate – include regarding them as *foolish* or

[69] Other moral anti-rationalists have rejected different premises. Dorsey (2016, 54-60), for example, accepts Premise 2, but rejects Premise 1. Other anti-rationalists, such as Sobel (2007b), express significant sympathy for Premise 2, but don't officially endorse it. As far as I know, the only anti-rationalist who has explicitly rejected Premise 2 is Gert (2014: 221). He writes 'My rejection of moral rationalism entails that an appropriate response to some rationally permissible options might be guilt or indignation.' Though Gert and I agree about this, he does not say much in defence of this claim.

[70] E.g., Darwall (2006a) and Portmore (2011).

idiotic. We might also react with *incredulity* or *disdain*. It would make little sense to *resent* someone for acting akratically.

Note next that BELS cannot be the following claim:

> S would morally blameworthy for freely and knowingly φ-ing only if S does not have sufficient *moral reason* to φ.

This claim states that a person cannot be morally blameworthy for performing a morally permissible action. If we interpret BELS in this way, then the BD fails to entail Moral Rationalism. BELS instead claims:

> S would be morally blameworthy for freely and knowingly φ-ing only if S does not have sufficient reason, *all things considered*, to φ.

If we understand BELS in this way, then the BD does entail Moral Rationalism.

We can next note that BELS, so understood, can be stated in different ways. Some of these alternatives may be clearer. For instance, BELS entails that, if S is morally blameworthy for φ-ing, then S did not have sufficient reason, all things considered, to φ. It also entails that, if S had sufficient reason, all things considered, to φ, then S is not morally blameworthy for φ-ing.

In what follows, I will often drop the 'all things considered' qualification. Unless stated otherwise, when I say that an agent had – for instance – sufficient reason to φ, this should be understood as the claim that the agent had sufficient reason, all things considered, to φ. In addition, when an agent has sufficient reason to φ, I will sometimes say that φ-ing was *reasonable*. And when an agent lacks sufficient to φ, I will sometimes say that φ-ing was *unreasonable.*

Before arguing that we should reject BELS, it is worth saying something about why people endorse it. BELS is motivated by the thought that there is a 'tension' – or even 'incoherence' – in the very idea of a person being morally blameworthy for acting reasonably.[71] After all, if they have acted reasonably, then – in one very important sense – they have not made a mistake. They have not failed to live as they have sufficient reason to live. Given this, how could it be appropriate to *blame* them for doing what they did? Darwall (2006b: 292) puts the thought this way:

[71] These quotes come from Portmore (2011) and Darwall (2006b) respectively.

> It seems incoherent… to blame while allowing that the wrong action, although recommended against by some reasons, was nonetheless the sensible thing to do, all things considered…. Part of what one does in blaming is simply to say that the person shouldn't have done what he did, other reasons to the contrary notwithstanding. After all, if someone can show that he had good and sufficient reasons for acting as he did, it would seem that he *has* accounted for himself and defeated any claim that he is to blame for anything.[72]

Others – including moral anti-rationalists – make similar remarks. Sobel (2007: 155-56), for example, writes that it:

> seems quite intuitive that earnestly blaming a person for φ-ing entails the view that the agent all things considered ought not to have φ-ed…. It also seems quite intuitive to say that if … one is acting as one ought… one's action is not worthy of blame.

And Dorsey (2016: 56) claims that to deny BELS is 'very implausible' and 'puzzling'. I shall now argue that denying BELS is neither implausible nor incoherent.

III

My argument does not directly target BELS. I instead defend a claim that is incompatible with BELS. This is:

> (A) In certain cases, S would be morally blameworthy for freely and knowingly φ-ing even if S had *sufficient prudential reason to φ*.

[72] This quote from Darwall is also used to motivate the plausibility of BELS by Portmore (2011, 46). Darwall has made similar claims in other works. For example, in *The Second Person Standpoint*, he (2006a, 98) asks us to consider Williams' claim that 'blaming carries this implication, that is, that we cannot intelligibly demand that that someone act in some way without implying that she has good and sufficient reason to do so.' About this, he (2006a, 98) writes that 'this just seems straightforwardly true. Try formulating an expression with which you might address a moral demand to someone. I doubt that you can find one that does not carry the implication that she has conclusive reason to do what you are demanding or reason not to have done what you are blaming her for.' In a discussion of a weaker claim about blaming someone implying the existence of a *pro tanto* reason to act as morality requires, he (2016, 268) writes: 'When we blame someone, we presuppose that the person we are blaming cannot sufficiently answer for what he has done. It is impossible coherently to blame someone and simultaneously accept that he lacked any (nonsubscripted, nonperspective-relative) *pro tanto* normative reason to act as he was morally obligated.'

As I use this (somewhat odd) phrase, to say that S had 'sufficient prudential reason to φ' is to say (i) that S had sufficient reason, all things considered, to φ, and (ii) that what made this the case was the prudential reasons that S had. To illustrate, suppose you must choose between two morally permissible options. You know that acting in the first way will bring you pleasure, and that acting in the second way will cause you despair. All else equal, the fact that acting in the first way will bring you pleasure gives you sufficient (and perhaps decisive) prudential reason to act in this way.[73]

My argument for (A) has two stages. I first defend:

(B) In certain cases, S would be morally blameworthy for acting prudently.

As well as appealing to (B)'s intrinsic plausibility, I argue for a certain *explanation* of (B). I then argue that, once we accept that (B) is true for this reason, we ought to accept (A).

FIRST STAGE: ON (B)

According to (B), there are cases where an agent would be morally blameworthy for acting prudently. The most compelling cases of this kind are those where it is in an agent's best interests to act immorally. Before offering such a case, it will be useful to make some preliminary remarks.

To begin with, some will deny the existence of these cases because they believe that morality and prudence never conflict. I will simply assume that this claim is false. I make some other assumptions. For one, I assume that prudence is *normative*. This is so in the sense that, if performing some action would be in an agent's best interests, then – due to this very fact – the agent has a reason to perform that action. I also assume that prudential reasons are *non-moral reasons*. This is not to say that we don't have moral reasons to act prudently. But it is to say that, even if we didn't have moral reasons to act prudently, we would still have reasons to act prudently.

Note next that one can accept (B) without denying that there are *also* cases in which an agent is not blameworthy for performing a prudent but immoral action. Plausible examples of such cases include those where the agent either doesn't *know* the action is immoral or is not *free* to

[73] For simplicity, I focus solely on prudential reasons in this piece. These are common-ground between defenders of MR and AR. For what it's worth, I believe the same argument could be run with other types of non-moral reasons, including excellence-based reasons.

perform a morally permissible action. To avoid these complications, it should be assumed – unless stated otherwise – that the agents I discuss act freely and knowingly.

There is a further complication I shall avoid. It is plausible that prudential considerations can affect *whether* an agent is morally required to perform an action in the first place. This may be so, for instance, when helping others would ruin their own life. As these are not cases of prudent immorality, I am not interested in them here. I am interested only in cases where an action is both morally impermissible and prudent.

We can next note that, as it is stated, (B) is perfectly compatible with BELS. This is because it can be accepted that an agent would sometimes be blameworthy for performing a prudent action but denied that this will ever be so when her prudential reasons give her *sufficient reason* to perform that action. This is why my argument requires a second stage. This second stage rests, not on the mere truth of (B), but on what I argue is the best explanation of (B).

The final point to note is that, on its face, (B) is exceedingly plausible. Consider the ring of Gyges. Given that the function of this ring is to remove prudential costs, Gyges' discovery that, when he turns the setting of this ring towards himself, he becomes invisible radically alters the prudential status of various available actions by removing most of the personal costs of performing them. Despite this, it seems clear that he is blameworthy for seducing the king's wife, attacking and killing the king, and then taking over the kingdom. After all, we are not only blameworthy for performing actions for which we are – or are likely to be – caught, punished, or actually blamed.[74] In short, even though discovering the ring of Gyges makes various immoral acts prudent, it does not seem to make a person impervious to being blameworthy for performing these acts. It seems clear then that – as (B) states – a person can be blameworthy for performing a prudent action.

That (B) is intuitively compelling is crucial to my argument. But, since the step from (B) to (A) relies on claims about the explanation of (B), it is insufficient. To draw these explanatory claims out, I shall now present another example. As I will refer to elements of this example

[74] Of course, a large part of the explanation for why the ring removes prudential costs is because you will not be caught, punished or blamed for your actions. Being blamed is often bad for you, but merely being *worthy* of blame is not.

throughout the paper, it is quite detailed. This will also allow us to avoid getting side-tracked by ultimately irrelevant objections that are likely to arise in response to a simpler case.[75]

EXAMPLE

When she was younger, Rebecca was stunningly beautiful. She took great pride in this fact. It is what she valued about herself above all else, and the central source of her happiness.

Everything changed for Rebecca when, after drinking too much at a party, she decided to drive home. On the way, she fell asleep at the wheel and crashed into a tree. This resulted in serious injuries. Most pertinently, Rebecca's face was severely disfigured. Following the crash, Rebecca became deeply depressed. Her doctors eventually declared that the damage was irreversible. Rebecca would never be beautiful again.

A number of years have now passed. Rebecca remains miserable. In part due to the change in her attitude since the crash, her friends have faded away. She is now completely alone. Though Rebecca has no mirrors in her apartment, when she happens to catch her reflection, she is overcome by self-loathing. And when she happens to see an attractive person, she is overcome by envy. This adds substantially to her misery.

One night, Mephistopheles appears to Rebecca. He offers her the following deal: If she is willing to torture and irreversibly mutilate five beautiful strangers, her own beauty will be restored. And this time it will never be taken from her, nor will it ever fade. There are conditions. Rebeca must select the victims herself, and they must be both innocent and kind.

Rebecca is torn. On the one hand, due to her envy, misery and deep desire to return to who she once was, mutilating these beautiful people is appealing. On the other hand, she wonders whether she would be able to live with herself. She also wonders whether the restoration of her beauty will really bring her happiness. When she asks Mephistopheles, he confirms that her beauty will restore her happiness, and that, once this has happened, she will flourish in numerous other ways. If she doesn't perform these actions, then she will never escape from the current depths of her despair. To further sweeten the deal, Mephistopheles makes the following guarantee: Rebecca will never be caught for these crimes, nor will she ever be suspected. He also guarantees that Rebecca will never feel guilt. Reflecting on her actions will be like remembering a horror film she saw long ago. He further promises that the performance of these

[75] This last point also explains why the example is as extreme as it is. Everyday cases tend to contain many confounding factors.

actions will bring Rebecca the most intense pleasure she has ever felt, as well as a deep sense of satisfaction and achievement.

Though she believes that performing these actions would be immoral, and though she knows that she is free to turn down the deal, Rebecca ultimately accepts. As promised, she thoroughly enjoys torturing and mutilating her victims. Her beauty and happiness are restored, and she goes on to live a flourishing life. Her victims, on the other, are never the same again. Aside from the permanent physical damage, they are permanently psychologically scarred. As a result, their lives fall apart. One victim – Charlotte – is unable to live with this trauma, and she eventually takes her own life.

CLAIMS

Two claims about this case are clear. First, Rebecca's belief about the moral status of her actions is *correct*. Torturing these innocent people was patently – and gravely – immoral.[76] Second, it seems clear that, if we somehow did find out about Rebecca's actions, we would – at least as a default – consider her to be morally blameworthy for performing them. To deny this has seemingly absurd implications. It implies, for instance, that her victims could not appropriately *resent* Rebecca for wrongly mutilating them. It also implies that, if Rebecca *did* come to feel guilt over her actions, this emotion would be unfitting.

As some may be suspicious of this second assumption – at least in this context – it is worth making clear just how weak it is intended to be. To say that someone is *default blameworthy* – as I shall use this term – is not to say that they *are* blameworthy. It is instead to say that there are considerations that support their blameworthiness, and hence that, if they are *not* blameworthy, there must be some other consideration that explains why this is so. Absent such a consideration, we should conclude that they are in fact blameworthy. This is because, if a person is default blameworthy, then we do not need *further* considerations in support of their blameworthiness. What makes them default blameworthy *just is* that we already have such considerations.

To illustrate, Rebecca is default blameworthy because she freely and knowingly performed a gravely immoral action. This strongly supports her blameworthiness. Given this, if she is not blameworthy, there must be some consideration that explains this. As I shall sometimes put

[76] I shall say not say more to defend this claim, even though rejecting it is one way to save BELS from my argument. I assume that if defending either BELS or its denial required rejecting this claim, then that would give us decisive reason to reject BELS or its denial.

this point, for it to be the case that Rebecca is not blameworthy, there must be some consideration that *undermines* the appropriateness of blaming Rebecca.

There is a third claim that I need to make about this case, which may be more controversial. This is that Rebecca's decision to accept Mephistopheles' deal – and her subsequent performance of these actions – was in her best interests. Here is the basic rationale for this claim: We can first note that there are actually two independent claims here. The first is that Rebecca had prudential reasons to perform these actions, and the second is the comparative claim that, of all the actions available to her, no other action was better supported by prudential reasons.

The case for the first claim largely rests on the sheer plausibility of:

> *The Prudential Value of Happiness*: Happiness may not be all that matters to how well a person's life is going for them, but it *does* matter. A person will always be better off in one important respect – and, all else equal, will always be better off overall – if they are happy rather than miserable.

Since a person's prudential reasons are grounded in their welfare, it can be inferred from this claim that, if performing some action will make a person happy, then there is a prudential reason for them to perform it. And if this is right, then Rebecca had prudential reasons to perform the immoral actions she did.

Unless hedonism is true, this does not entail that Rebecca's choice was *prudent*. It is possible that other options were better supported by non-hedonic prudential reasons. Some of these other (possible) reasons can be stipulated away without altering the important features of the case. But others cannot. For instance, it might be claimed that it is intrinsically bad for an agent to freely and knowingly act immorally. This claim is plausible. What is far less plausible, however, is the thought that this purported prudential consideration against her actions is stronger than the overwhelming (and uncontroversial) prudential considerations in favour. To claim that it is worse for an agent to act immorally than it is for them to suffer lifelong despair, self-hatred and isolation commits one to an absurdly moralistic conception of welfare.

Assuming these claims are correct, I have described a genuine case of prudent immorality. In this case, Rebecca is what I have called default blameworthy. As such, if there are no considerations that undermine her default blameworthiness, (B) is true.

I shall make one further assumption about this case. This is that the *only* non-moral reasons that support Rebecca's immoral actions are her prudential reasons. This assumption makes my argument simpler to present. Unlike my other assumptions, I believe it may be false. But, as far as I can tell, the assumption does not affect my argument.

Let's now return to Rebecca. Imagine that, after somehow finding out about her actions, you come face-to-face with her.[77] As expected, she feels no guilt. But she also believes that any resentment her victims feel – or any indignation you feel – is inappropriate. You ask her why. To put it in the above terminology, you ask her to point to the consideration that undermines her default blameworthiness for wrongly mutilating her victims. She says:

> 'I recognise that my actions were morally indefensible. Your blame is not inappropriate due to a mistake about this fact. Your blame is inappropriate because these actions were in my best interests. It is this consideration that undermines the fittingness of blame.'

I shall refer to this as *Rebecca's rationale.* On its face, it is far from convincing. The idea, for instance, that her *enjoyment* of mutilating her victims mitigates the appropriateness of their resentment – or our indignation – seems deeply implausible. After all, if this was true, then it would be correct for us to blame a killer less once we discovered that they were a sadist. But this discovery, if anything, tends to intensify our indignation.

It is intuitive, then, that – as with Gyges – this is a case where (B) holds true. But what exactly *explains* this. In other words, *why* is it correct to think that, even though Rebecca had good prudential reasons to torture her victims, she is nonetheless morally blameworthy for doing so? If we can answer this question – which I shall refer to as (Q) – then we will have explained and vindicated (B).

[77] To be clear, this is a hypothetical version of Rebecca. The actual Rebecca was never suspected.

MY ANSWER TO (Q)

The correct answer to (Q), I shall argue, is:

> *Wrong Kind of Reason* (WKR): When it comes to S's blameworthiness for freely and knowingly performing an immoral action, S's prudential reasons to perform that action that are due to the benefits that she would derive from performing the immoral action are the wrong kind of reason to undermine the appropriateness of moral blame.[78]

On this view, Rebecca's rationale fails because she cites reasons in favour of mutilating her victims that *do not bear on* whether she is morally blameworthy for mutilating her victims. In other words, her rationale fails because she cites reasons that do nothing to undermine the fittingness of blame.

WKR requires clarification and defence. It is important to first distinguish it from a distinct claim with which it is easily confused. There are countless considerations that Rebecca could cite in favour of her actions that would fail to undermine her blameworthiness. She could claim, for instance, that her actions pleased Moloch – a god associated with human sacrifice. This would *not* be a case of citing a reason of the wrong kind, as it would not be a case of citing a (normative) reason at all. Unlike this consideration, reasons of the wrong kind genuinely favour a response.

Some might wonder whether the failure of Rebecca's rationale is due to a more *general* mistake than WKR. In attempting to demonstrate that she is not blameworthy, Rebecca cites facts about the *deontic status* of her actions. She concedes that they were *morally impermissible* but claims they were *prudentially required*. Though such claims are clearly relevant to whether she ought to have *done it*, we might doubt their relevance to the conceptually distinct question of whether

[78] This claim is narrower than it may initially appear. As noted above, I am not assuming that prudential considerations make no difference to whether an agent is morally required to perform an action in the first place. I am also not assuming that prudential considerations make no difference to whether an agent has acted freely. At least for the purposes of this argument, WKR only concerns cases – like (I hope) Rebecca's – where the claim that the agent has freely and knowingly acted immorally is not in dispute. Note also that WKR is only concerned with the prudential benefits that an agent receives as a result of acting immorally. Again, this is all the argument requires. In Rebecca's case, this would include things like the pleasure that she derives from the mutilation itself, and the satisfaction and happiness that she later derives from having her beauty restored. For ease of expression, I will often speak simply of Rebecca's 'prudential reasons', but this is shorthand for prudential reasons of this kind.

she is *blameworthy for* doing it.[79] Though facts other than the deontic status of an action matter to whether a person is blameworthy for performing that action, it is a mistake to think that the deontic status doesn't also matter. If Rebecca could convince us, for instance, that her actions were morally required, it is difficult to see how we could appropriately blame her for performing these morally required actions.[80] For this reason, Rebecca is not making a general mistake in offering us deontic considerations. Her mistake, according to WKR, consists in offering us the *wrong considerations*.

Note next that, if WKR is correct, it is an instance of a more general – and much discussed[81] – phenomenon. There are various contexts in which there seem to be reasons of the wrong (and right) kind. It is highly controversial what *makes* a reason a reason of the wrong kind across these contexts – that is, what the correct general account of a wrong kind of reason is. This is not a question we need to settle. My argument ultimately rests on the claim that Rebecca's prudential reasons have a certain feature that explains why they fail – and must fail – to undermine the fittingness of moral blame. As I explain below, this feature can be discussed without reference to the idea of a reason of the wrong kind. Nonetheless, since many – and perhaps all – purported reasons of the wrong kind share this feature, it will be useful to begin by discussing some common examples.

TWO EXAMPLES

I shall begin with what some consider to be the paradigm example of a wrong kind of reason. These are *pragmatic reasons for belief*.[82] These are considerations that favour believing *p*, but which do *not* speak in favour of the truth of *p*. For instance, believing that your life is meaningful can have various psychological benefits. When this is so, this fact may count in favour of you having this belief. This fact, however, is independent of the truth of this belief. It would speak equally in favour of believing any proposition – true or false – that had the same psychological benefits.

[79] That these are distinct questions can be seen by noting that two people can coherently agree that some action is required, but disagree about whether a person is blameworthy for failing to performing it.
[80] This is to agree with Portmore (2011), who considers – and rejects – purported counterexamples to BELS with this structure. That said, I do think – in a fairly complicated way – that considerations of moral worth count against BELS. To save space, I have not focused on this here.
[81] For a general overview of this literature, see Gertken & Kiesewetter (2017).
[82] For good discussions of pragmatic reasons for belief, see Reisner (2018) and Schroeder (2012). For a defence of the claim that there are pragmatic reasons for belief – and that these are of the wrong kind – see Reisner (2009).

Pragmatic reasons for belief can be contrasted with what are often called *epistemic* or *alethic* reasons for belief. These are the *right kind* of reason for belief. An epistemic reason is a consideration in favour of believing *p* that does speak in favour of the truth of *p*. For instance, the fact that, under normal conditions, you perceive a chair in the room is an epistemic reason to believe that there is a chair in the room. This is because your perception of the chair is evidence that there is in fact a chair.

To better illustrate this distinction – and to get at the sense in which pragmatic reasons are reasons of the wrong kind – return to Rebecca. Consider her belief – held before the crash – that she is beautiful. This belief had numerous psychological benefits. It was the basis of her sense of self-worth and the foundation of her happiness. In short, Rebecca's belief about her own beauty was good for her. We can next note that some of these benefits were a result of Rebecca *believing* that she is beautiful, not of her *being* beautiful. As such, if these benefits count in favour her having this belief, then they equally count in favour her having this belief after the crash.

The same is not true of Rebecca's alethic reasons to believe she is beautiful. Before the crash, Rebecca would have had strong – and perhaps conclusive – evidence that she is in fact beautiful. Among other things, other people's reactions to her, and her own perceptions, would speak in favour of this proposition. After the crash, the same considerations would provide Rebecca with strong – and perhaps conclusive – evidence that she is not beautiful.

If this is right, then Rebecca had conclusive epistemic reason to believe that she was beautiful before the crash, but not beautiful after the crash. On the other hand, she had conclusive pragmatic reason to believe that she was beautiful before and after the crash. And if this is right, then, after her disfigurement, she has conclusive pragmatic reason to believe that she is beautiful, and conclusive epistemic reason to believe that she is not beautiful.

In our original case, Rebecca believes, after the crash, that she is not beautiful. We can also imagine a version of the case in which, after the crash, she continues to believe – against the evidence – that she is beautiful. Various causal mechanisms could account for this, but let's suppose that, just after waking, Rebecca was hypnotised to continue to hold her positive belief about her beauty, and to be unable to recognise evidence to the contrary as evidence to the contrary. Call this version of Rebecca *Rebecca**. I shall also refer the proposition that Rebecca believes – that she is not beautiful – as *p*, and the proposition that Rebecca* believes as *p**.

There are various things we may want to say about Rebecca and Rebecca*. We might wonder, for instance, whether Rebecca or Rebecca* – or both – have sufficient reason, all things considered, to have the belief they have. This is the most difficult question we can ask. But there are other important questions we can ask, and it is in many of these contexts that pragmatic reasons are the wrong kind of reasons.

For example: Does Rebecca *know that p*, and does Rebecca* know that *p**? Given that both already believe these propositions, this question turns on at least two further considerations: Is *p*(*) true, and is Rebecca(*) justified in believing that *p*(*)? The first of these considerations has nothing to do with reasons for belief, so we can focus on the second. When we ask whether someone in justified in believing *p* – in the sense relevant to whether they know that *p* – we are asking, very roughly, whether they hold the belief for reasons that speak in favour of the truth of the belief. Pragmatic reasons for belief simply do not bear on this question. Whether Rebecca(*) is justified in believing *p*(*) – or whether she knows that *p*(*) – does not turn, to any degree, on whether this belief is good, neutral or terrible for her. Given this, if Rebecca* cited the prudential benefits of her belief when arguing that her belief was justified, she would be making a mistake. She would be citing reasons of the wrong kind. Similar claims apply to other familiar – and central – epistemic notions. These include whether a belief is *credible*, *warranted*, or *theoretically* (or *epistemically*) *rational*.

Since not everyone accepts that there are pragmatic reasons for belief, it is worth giving a further example. This will hopefully be accepted by those who reject the first. I shall call these *pragmatic reasons against envy*.[83] After her disfigurement, Rebecca feels intense envy whenever she sees a beautiful person. This emotion causes her significant despair and disillusionment. For ease of expression, I shall focus on Rebecca's envy of Charlotte.

As with belief, there are various things we can say about this emotion. For example, it seems clear that feeling envy is bad for Rebecca. As such, she has prudential reasons not to envy Charlotte.

There is another kind of evaluation we can make of Rebecca's envy. We can ask whether it is fitting. Whether envy is fitting turns on whether the envied person is *enviable* – or worthy of being envied. In our case, this question turns on whether Charlotte is enviable. Though it is

[83] For a discussion of the wrong kinds of reasons to feel envy, see D'Arms and Jacobson (2000a and 2000b). For more general discussions of envy, see D'Arms (2017) and Thomason (2015).

difficult to give a complete account of the fittingness conditions of envy, some conditions are relatively clear. For instance, A's envy of B is fitting only if (i) the feature of B that A envies is *worth wanting*; (ii) A does not already possess the feature of B that they envy; and (iii) if A did possess this feature, they would continue to want it. Each of these conditions is intuitive. It makes no sense to envy someone for having something that is not worth having, or for possessing something that you already possess, or for having something that you don't actually want. If we focus on these conditions, it is plausible that Rebecca's envy of Charlotte is fitting. It is plausible, for instance, that beauty is worth wanting.

Just as some of the fittingness conditions of envy are relatively clear, it is relatively clear that certain considerations are not among these conditions. It is clearly false, for instance, that A's envy of B is fitting only if B was born on the 15th of April. It is similarly clear that prudential considerations – on the part of the envier – are not among these conditions. For one, a person can clearly be enviable regardless of how terrible it is for us to envy them. If you marry the girl of my dreams, and achieve all of the worthwhile goals that I have tried but failed to achieve, then it clearly makes sense for me to envy you. That this emotion will cause me despair is irrelevant to whether what you have is worth having, or whether I actually want it. Similarly, that envying someone will make you ecstatically happy makes no difference to whether that person is enviable. After all, we could easily concoct a case in which it would be prudent for someone to envy Charlotte for being mutilated, but this would do nothing to make her situation enviable.

Given this, if we want to show Rebecca that her envy of Charlotte's beauty is unfitting, we need to show that her envy does not meet conditions like (i)-(iii). We could try to show, for instance, that Charlotte's beauty is not worth wanting. We would be making a mistake, however, if we tried to show that her envy is unfitting because it is bad for her. This may give her good reason not to envy Charlotte, but it would be a good reason of the wrong kind.

SOME REMARKS

That both examples involve reasons of the wrong kind is intuitive. When we are trying to determine whether some belief is credible, justified, or epistemically rational, the fact that it makes a person happy seems entirely orthogonal. Similarly, when we are trying to determine whether someone is enviable, or whether what they have is worth wanting, the fact that having the attitude of envy is bad for the envier seems beside the point.

As is probably clear, I have focused my discussion on one feature of these examples. That these two examples share this feature is unsurprising, since it is – to use Schroeder's (2012) term – an 'earmark' of reasons of the wrong kind. We can call this feature *irrelevance*. Using pragmatic reasons for belief as illustration, Schroeder (2012: 459-60) puts the point this way:

> So there seems to be a distinctive dimension of rational assessment of beliefs – sometimes called *epistemic* rationality – that is affected by the epistemic reasons of which the subject is aware but not affected by the pragmatic reasons of which the subject is aware. The same observation goes, whether we are talking about the rationality *of* believing or the rationality *in* believing – the distinction that epistemologists sometimes call the distinction between *propositional* and *doxastic* justification. Pascal's reasons no more affect whether someone who believes in God can be said to do so rationally than they affect whether it would be rational for someone to believe in God… Focusing on epistemic rationality – this distinct dimension of the rational assessment of beliefs – allows us to elide the question of whether there is also some sense in which Pascal showed belief in God to be rational, perhaps a more global or practical sense less central to epistemology…. What is important is that there is some central dimension of rational assessment that is not affected by Pascalian considerations.

As well as holding in the case of pragmatic reasons for belief, irrelevance holds for pragmatic reasons against envy. There is some central dimension of assessment – an emotion's fittingness – that is not affected by pragmatic considerations.

It is important to emphasise that the explanation for these claims is not that the pragmatic reasons are insufficiently strong. However happy believing that she is beautiful makes Rebecca*, and however miserable envying Charlotte makes Rebecca, it is still the case that Rebecca*'s beliefs are unjustified, and that Rebecca's envy is fitting. The explanation is that these pragmatic reasons do not bear on whether a belief is justified or whether envy is fitting.

My remarks thus far may raise a question. I have been assuming that these pragmatic reasons are genuine reasons to have the response. Suppose there is a case where these pragmatic reasons give an agent sufficient or decisive reason, all things considered, to have one of these responses. What does WKR imply about such a case? Start with envy. Suppose it is true that envying Charlotte is terrible for Rebecca, and that this fact gives her decisive reason not to envy Charlotte. Since these are reasons of the wrong kind, what we would learn from such a case is

that Rebecca has decisive reason to feel an *unfitting emotion*. We would *not* learn that prudential reasons are among the fittingness conditions of envy after all. This conclusion isn't particularly surprising. There seem to be various cases where it is reasonable to feel an unfitting emotion. If someone holds a gun to my head and tells me to admire him or he will kill me, then I have sufficient reason to admire him. But this would not be fitting since this person is not *admirable*. We do not even need such farfetched cases to illustrate this point. Given that positive emotions are plausibly good for us, and that negative emotions are often the fitting responses to our lives and the world around us, it is probably often true that feeling unfitting positive emotions is reasonable. Similar claims apply to pragmatic reasons for belief. If some pragmatic reason gave us decisive reason to believe that *p*, this would not teach us a surprising new fact about (say) epistemic rationality. If we know anything about this, we know that believing *p* merely because it makes you happy is epistemically irrational. What we would learn instead is that it is sometimes reasonable to have epistemically irrational beliefs.

It is worth also noting that our discussion suggests a feature that is shared by reasons of the *right kind*. This is just the converse of irrelevance, so we can call it *relevance*. A consideration is relevant if it *does* make a difference to the assessment. Hence, epistemic reasons are relevant because they make a difference to whether you are justified in believing that *p*.

A TEST

Before returning to our case, it will be useful to have a general test for relevance and irrelevance. We can then apply this to Rebecca. I suggest the following. We start with some state of affairs, and then make the pertinent normative judgement. The *pertinent* normative judgement being a judgement about whatever we are testing. Suppose we want to know what sorts of reasons are relevant to epistemic justification. We can first consider the following state of affairs: *S* has strong evidence for *p*, and only weak evidence against *p*. It is also true that *S's* belief that *p* has various psychological benefits for *S*. My initial judgement is that, in this case, *S*'s belief that *p* is epistemically justified.

The next step is to alter the state of affairs, and check whether this makes any difference to the initial judgement. If it does, then relevance is true, and if not, then irrelevance is true. We can first suppose that, holding the psychological benefits fixed, *S* has only weak evidence for *p*, and strong evidence for not-*p*. Intuitively, this makes a difference to the initial judgement. An agent is epistemically unjustified in believing *p* when she has much stronger evidence that not-*p*. Unsurprisingly, then, evidential considerations come out as relevant on this test.

We can next suppose instead that, holding the initial evidential considerations fixed, believing *p* would lead *S* to despair. Intuitively, this does not make a difference to the initial judgement. It is epistemically justified to believe that *p* when you have strong evidence that *p*, and this is so even when believing *p* would lead to despair. Again unsurprisingly, pragmatic considerations come out as irrelevant on this test.

It should be noted that this is the simplest kind of application of this test. It is one in which, when it comes to the relevant reasons, our initial judgement reverses. Other cases are more complicated, since, in these cases, varying the relevant reasons does not reverse our initial judgement. This is easiest to illustrate with envy. Suppose we think that, in our original case, Rebecca's envy of Charlotte's beauty is fitting. Now suppose we vary the facts that are intuitively relevant to this judgement, such as by making Charlotte somewhat less beautiful. This adjustment is unlikely to lead us to reverse our initial judgement, but Charlotte's beauty is surely relevant to the appropriateness of Rebecca's envy of her beauty.

It would be a mistake to conclude from this that the test fails to track relevance and irrelevance. This is because varying Charlotte's beauty does make a difference to the appropriateness of Rebecca's envy. We can see this by noting that the appropriateness of Rebecca's envy turns not only on whether she can appropriately feel the emotion at all, but also on whether the *intensity* of her envy is fitting. Suppose, in the initial case, Rebecca can appropriately feel envy to degree *n*. Now suppose we make Charlotte's beauty less worth wanting. It may still be true that Rebecca can appropriately envy Charlotte, but it will no longer be true that she can appropriately envy Charlotte to degree *n*. As such, the extent to which Charlotte's beauty is worth wanting is relevant to the appropriateness of Rebecca's envy. These same points do *not* hold of Rebecca's pragmatic reasons not to envy Charlotte. Whether it would be bad for Rebecca to envy Charlotte makes no difference to whether Rebecca can appropriately envy Charlotte at all, or to the degree of envy that she can fittingly feel.

I shall avoid these complications by defending the claim that Rebecca's prudential reasons make no difference to either our initial verdict about her blameworthiness *or* the intensity of blame that can appropriately be felt in response to her actions.

BACK TO REBECCA

Drawing on our discussion of pragmatic reasons for belief and against envy, I shall now offer considerations in favour of WKR. These attempt to establish that Rebecca's prudential reasons to mutilate her victims do not bear on whether she is blameworthy for mutilating her victims.

First, a dialectical point. Since I only need to establish irrelevance, we can now drop all talk of 'wrong kinds of reasons'. Understood without reference to this idea, my claim becomes:

> WKR*: When it comes to S's blameworthiness for freely and knowingly performing an immoral action, S's prudential reasons that are due to the benefits that she would derive from performing the immoral action are *irrelevant* to whether S is morally blameworthy for performing that action.

WKR* would answer (Q) just as well as – and may simply be equivalent to – WKR. Though I shall continue to use the phrase 'wrong kind of reason', if someone does not like my use of this phrase – or the idea of a wrong kind of reason – this can be understood as *irrelevant reason*.

RUNNING THE TEST

I shall begin with the test outlined above. We can start with the case as originally described. Rebecca freely and knowingly mutilates her victims. These actions were immoral, but she had prudential reasons to perform them. Our initial judgement about this case, I shall suppose, is that Rebecca is blameworthy for mutilating her victims. We also require a judgement about the *intensity* of the blame that can appropriately be felt. To run the test, I do not need to defend a correct intensity. Whatever you believe the correct intensity is, I shall call this *n*. For ease of reference, call this *version one* of the case.

We can now test Rebecca's prudential reasons for irrelevance by altering the prudentially relevant facts. In *version two*, Rebecca performs the same actions as in version one. The difference is that, when she mutilates her victims, she feels *nothing*. In *version three*, Rebecca again performs the exact same actions as in version one. In this case, however, performing these actions causes Rebecca severe psychological distress.

If Rebecca's prudential reasons are relevant to her blameworthiness, then these changes to the prudential facts of the case will make a difference either to whether Rebecca is blameworthy at all, or to the intensity of the blame that can be appropriately felt towards Rebecca. In other words, Rebecca would be *less* blameworthy in version one than in version three. More

carefully, as we move from version one to version three, *n* will change from fitting to unfitting. This will either be because the blame emotions become entirely inappropriate, or because *n* becomes inappropriately intense.

I submit that, when we reflect on these different cases, these changes do not take place. If *n* is fitting in version one, then it is fitting in version three. The idea that, as Charlotte is being impermissibly tortured by Rebecca, the appropriateness of her resentment rests, to any degree, on whether Rebecca is enjoying the experience of torturing her – or is otherwise benefitting from these actions – is entirely implausible. The same is true of the idea that the appropriate intensity of guilt that Rebecca can feel about her impermissible actions – if she ever came to feel guilt – depends on whether, and to what extent, she got off on mutilating her victims, or how much she enjoys her new life afterwards. She can appropriately feel equally guilty for what she did to these innocent people in all three cases. Similar points hold of *our* emotions. The idea that, in version one, Rebecca is less deserving of indignation for impermissibly mutilating her victims because she enjoyed it defies belief. As noted earlier, we do not regard sadistic killers as less deserving of indignation than non-sadistic killers.

If these claims are correct, then Rebecca's prudential reasons to mutilate her victims are not relevant to whether she is blameworthy for mutilating her victims. And, if this is right, then it follows that her prudential reasons are the wrong kind of reason to undermine Rebecca's blameworthiness.

To see that this test is not rigged, note that varying other features of version one does make a difference to Rebecca's blameworthiness. This is most obvious when we alter the morally relevant facts. Holding the prudential benefits fixed, suppose that all Mephistopheles' asks of Rebecca is that she make five beautiful people's lives temporarily worse by poisoning them with a virus that causes an aesthetically unappealing skin rash. This rash last for a year but leaves no permanent trace. In this case, it would make sense for Rebecca's victims to resent her, but it would be unfitting for them to feel as resentful as those in version one can fittingly feel. This is because their resentment would be failing to accurately track the badness of Rebecca's actions. It would be the same kind of mistake as feeling as resentful of someone for stealing your coat as for stealing your child. Moral reasons, then, come out as relevant.

It is not only moral facts that make a difference. Suppose that we alter how *free* Rebecca was to perform these actions. Consider a version of the case where Mephistopheles, wanting to make it less likely that Rebecca backs out, coerces her into mutilating her victims – which will

still have all the same prudential benefits – by threatening to torture *her* if she does not do so. This fact, I take it, would make Rebecca less blameworthy. As a rule, that someone was coerced into performing an action counts against their blameworthiness for performing that action.

In sum, moral reasons are relevant to whether a person is morally blameworthy for performing an action. The same holds of considerations concerning a person's freedom to perform the action. If Rebecca's rationale had included considerations of these kinds, she would at least be citing considerations that bear on her blameworthiness. Prudential reasons, on the other hand, are irrelevant. Since Rebecca's rationale consisted solely of these irrelevant considerations, it could not – even in principle – undermine her blameworthiness. This is what explains why she is blameworthy even though she had good prudential reasons to mutilate her victims.

INTUITIVENESS

We can next note that this idea – that an agent's prudential reasons to perform an immoral action do not bear on whether they are blameworthy for performing that action – is just straightforwardly plausible.

To get at this, we can generalise the point made above. There is something deeply implausible about the idea that the blame that a victim of a wrongdoing can fittingly feel rests, to any degree, on whether the wrongdoer benefited from wronging them. As a defence against being blamed, claims like 'but I enjoyed wronging you', 'but it was good for me to wrong you', and 'but I wanted to wrong you' seem to have no undercutting force at all.

In case some are concerned about drawing conclusions from these farfetched cases, it is worth making two points. First, these fantastic cases have real-life analogues. After all, there really are people who derive significant pleasure from torture and murder. A number of well-known serial killers fall into this category. For example, in his 1969 cipher, the Zodiac Killer wrote:

> I like killing people because it is so much fun. It is more fun than killing wild game in the forest, because man is the most dangerous animal of all. To kill something is the most thrilling experience. It is even better than getting your rocks off with a girl.

If we should accept that experiencing pleasure is a prudential benefit, then we should accept that the Zodiac Killer had prudential reasons to perform the actions they did. He is, in this respect, no different than Rebecca. Further, in some cases – including in the case of the Zodiac killer – these people were never caught, and they seemed to lack the capacity for guilt. For all

practical purposes, they were wearing the ring of Gyges. Given the prudential value that such killers derived from these acts, I doubt we can consistently hold that Rebecca's actions were prudent, but that these killer's actions were not.

We can then make the same points regarding these real-life cases. It is deeply implausible that the appropriateness of the resentment that the Zodiac's victims can feel – or of our indignation upon learning of the Zodiac's acts – rests, to any degree, on whether the Zodiac found committing the murder 'so much fun', 'a thrilling experience' or 'better than getting your rocks off with a girl.'

Second, WKR is intuitively compelling even when applied to the more banal immoralities of everyday life. People often lie, act selfishly and let each other down. In some instances, there are good prudential reasons to perform these actions. It would be incredible to claim that it is never in a person's best interests to lie to someone, or to betray them. Nonetheless, it seems equally incredible to claim that the benefit someone derives from wrongly betraying you bears on the resentment you can appropriately feel. On the contrary, it seems entirely natural to say 'I don't care whether you benefitted from betraying me.' Similarly, when someone close to us acts selfishly, there seems to be no undermining force at all in the claim that they *wanted* to act selfishly. Finally, when we fail to perform minimal duties of beneficence, it just seems absurd to think that those who have suffered as a result should feel any better about us when we inform them that impermissibly allowing them to suffer was genuinely good for us.

The fact that these claims generalise across these different sorts of cases – realistic and unrealistic, unusual and mundane – strongly supports the implausibility of the very idea that an agent's prudential reasons to perform an immoral action can bear on whether they are blameworthy for performing that action.

In addition to passing the test above, I conclude that WKR has significant intuitive appeal. It is about as clear as the claim that, when it comes to whether a person is enviable, the enviers prudential reasons not to feel the emotion are irrelevant.

IV

Before moving to the second stage of my argument, I shall summarise my claims so far. To regain the beauty she lost, Rebecca mutilates five beautiful strangers. These actions are immoral. Despite their moral status, Rebecca had strong reasons to perform them. These reasons derive from the significant prudential benefits that their performance brings to Rebecca. This case, I have claimed, is an example of *prudent immorality*.

Intuitively, Rebecca is *morally blameworthy* for mutilating these innocent strangers. This is so despite her having good prudential reasons to perform these actions. I have defended a claim that explains and justifies this intuition. This is:

> *WKR*: When it comes to S's blameworthiness for freely and knowingly performing an immoral action, S's prudential reasons to perform that action that are due to the benefits that she would derive from performing the immoral action are the wrong kind of reason to undermine the appropriateness of moral blame.

If Rebecca is *not* blameworthy for mutilating these strangers, there must be a consideration that explains why this is so. As I have put this point, there must be a consideration that *undermines* the default appropriateness of moral blame. If WKR is true, Rebecca's prudential reasons to mutilate her victims cannot play this undermining role, since they do not bear on whether she is morally blameworthy. This is so no matter how good these reasons are. Assuming there are no other non-moral non-prudential considerations that undermine her blameworthiness, it follows that Rebecca *is* morally blameworthy. And this is just (B).

We can now generalise this claim. S will be morally blameworthy for performing a prudent action if (i) that action was immoral; (ii) the immoral action was performed freely and knowingly; and (iii) there are no other non-moral non-prudential reasons that undermine the appropriateness of moral blame.

V

I will now defend:

> (C) If we should accept (B), then we should accept (A).

Recall that, according to (A), there are cases where a person would be morally blameworthy for φ-ing even if she had sufficient prudential reason to φ. My argument is easiest to illustrate if we make (A) narrower. From here on, I shall understand it as:

> Rebecca would be morally blameworthy for freely and knowingly acting immorally even if her prudential reasons to mutilate her victims gave her sufficient reason, all things considered, to perform these actions.

As this claim entails the more general version of (A), and entails that BELS is false, I only need to defend this narrow version of (A).

Recall next that (B) alone does *not* entail (A), and nor does it entail this narrow version of (A). It is coherent to claim – as defenders of BELS must if they accept (B) – that a person can be morally blameworthy for acting prudently, but that they would *not* be morally blameworthy if their prudential reasons gave them sufficient reason, all things considered, to act prudently. For this reason, my argument is more complicated than (C) may suggest. (C) should be understood as the claim that:

> If we accept that there are cases – like Rebecca's – where WKR explains (B), then we should accept that, in those same cases, it would be true that, even if S's prudential reasons gave S sufficient reason, all things considered, to act prudently, S would still be morally blameworthy for acting prudently.

For the purposes of illustrating (C) – and since I have just defended it – I shall here assume that WKR explains Rebecca's case.

To get at the rationale behind (C), it may be easiest to start comparatively. We can first ask: *If* Rebecca's prudential reasons to mutilate her victims gave her sufficient reason to mutilate her victims, would she still be morally blameworthy? As just noted, a defender of BELS must answer 'No' to this question. Some may find this implication of BELS counterintuitive – and perhaps sufficiently so to reject the view. I think this is right, but it is not the point I want to make. Notice instead that this answer commits one to the view that Rebecca's prudential reasons to perform these immoral actions are reasons of the *right kind* to undermine moral

blameworthiness. This is because, according to BELS, any reason that helps determine what an agent ought to do, all things considered, is *relevant* to whether that agent is morally blameworthy for performing an action.

If WKR is true, this is a fundamentally misguided way to think about Rebecca's blameworthiness. Though the prudential benefits she derives from mutilating her victims do count in favour of her mutilating her victims, they are irrelevant to the resentment that her victims can appropriately feel. Rebecca's rationale goes wrong, *not* because she cites considerations in favour of her actions that do not give her sufficient reason, all things considered, to perform these actions – as BELS must claim – but because she cites considerations that do not bear on the fittingness of her mutilated victims resentment. That endorsing BELS forces one to reject WKR is itself a good reason to reject BELS.

Notice next that, if WKR is true, then answering 'No' to the above question makes little sense. This is because the reasons that would give Rebecca sufficient reason, all things considered, to mutilate her victims, are the *exact reasons* that are irrelevant to whether she is blameworthy for performing these actions. They are the exact reasons, that is, that are unable to undermine the fittingness of blame. From this, it seems to follow that Rebecca *would* be morally blameworthy for mutilating her victims even if her prudential reasons gave her sufficient reason, all things considered, to do so. Once again – and perhaps to belabour the point – this is because the reasons in virtue of which her actions would be reasonable *just are* the reasons that make no difference to the fittingness of her mutilated victims' resentment, or of our indignation, over her performance of these acts.

Here is another way to get at this. We can first ask: Would Rebecca be blameworthy for mutilating her victims if she *didn't* have any prudential reasons to perform these actions? I doubt that anyone would deny the answer is 'Yes'. We can next note that, if WKR is true, then adding these reasons cannot make a difference to Rebecca's blameworthiness. And this is so no matter how good these reasons are. As such, if Rebecca is blameworthy absent these prudential reasons, then she will remain blameworthy when these reasons are added. We can next suppose – at least for the sake of argument – that we add these irrelevant reasons and that these irrelevant reasons give Rebecca sufficient reason, all things considered, to mutilate her victims. Would this make a difference to her blameworthiness? The answer seems to be 'No'. Reasons that do not bear on the fittingness of blame can make a difference to whether an agent ought to perform an action. But reasons that do not bear on the fittingness of blame cannot

make a difference to the fittingness of blame. From this, it seems to follow that *even if* these irrelevant reasons gave Rebecca sufficient reason to act, they would not bear on whether she is morally blameworthy. And given that, absent these reasons, Rebecca would be morally blameworthy for mutilating her victims, it follows that she *would* be morally blameworthy even if these irrelevant reasons gave her sufficient reason, all things considered, to act as she does. And this is just (A).

To further clarify this claim, it may help to return to the fittingness of envy. This will also allow us to see some of the mistakes that defenders of BELS make if WKR is true. Suppose that Rebecca's envy of Charlotte is fitting – say, because Charlotte's beauty is worth wanting – and that her envy of Charlotte is bad for her. We can now ask:

> (Q*) Why is it correct to think that, though Rebecca has strong prudential reasons not to envy Charlotte, her envy of Charlotte is nonetheless fitting?

The answer is that Rebecca's prudential reasons not to feel this emotion do not bear on whether it is fitting for her to feel this emotion – that is, on whether Charlotte is worthy of being envied. Suppose next that, after seeing the severe despair that Rebecca's envy is causing her, one of her old friends – Adam – tries to convince Rebecca that her envy is unfitting. He offers the equivalent of Rebecca's rationale:

> 'It is true that Charlotte's beauty is worth wanting. Your envy is not inappropriate due to a mistake about this fact. Your envy is inappropriate because feeling this emotion is not in your best interests. It is this consideration that undermines the fittingness of your envy.'

Adam's mistake here is that he is falsely assuming that prudential facts about the envier are relevant to – that they bear on – the fittingness of envy. Since this is not so – since these are reasons of the wrong kind – this rationale does nothing to show that Rebecca's envy is inappropriate. And this is so regardless of how good these prudential reasons are. Defenders of BELS make the same mistake as Adam. They falsely assume that an agent's prudential reasons to perform an immoral action are relevant to – that they bear on – the fittingness of blame; on whether an agent is worthy of blame.

We can next ask: *If* Rebecca's prudential reasons not to envy Charlotte gave her sufficient reason, all things considered, not to envy Charlotte, would her envy still be fitting? The answer is 'Yes'. And the rationale for this is identical to the rationale for the claim that Rebecca would

still be blameworthy for mutilating her victims. The reasons against Rebecca's envy are the *exact reasons* that do not bear on the fittingness of envy. Hence, given that there are considerations in favour of Charlotte's enviableness – that her beauty is worth wanting – it will be the case that Rebecca's envy is fitting even if these prudential reasons give her sufficient prudential reason not to envy Charlotte.

This example allows us to see other mistakes that defenders of BELS make. As previously noted, those who advocate BELS claim that the very idea of someone being blameworthy for performing a reasonable action is 'incoherent', 'puzzling', and 'very implausible'. It is worth making two points about these claims.

First, even on their face, it is not clear how compelling these claims are. After all, many accept that we can appropriately *regret* performing actions that we had sufficient reason to perform.[84] For example, if someone can either live a happy life or produce excellent philosophical works – but cannot do both – then it is very plausible that, whatever choice they make, they can appropriately feel some regret. In either case, they will have sacrificed a life that was well worth living. Plausibly, however, either choice is reasonable. If this claim is not incoherent or implausible, then the burden is on defenders of BELS to explain why it is incoherent or implausible to feel some degree of guilt when we perform a reasonable action.

Second, and whatever we make of the above claim, there is nothing incoherent or implausible about denying BELS if WKR is true. If S's prudential reasons to commit an immoral action are irrelevant to the appropriateness of blaming S for performing that action, then it is clear why S could be a fitting target of blame even when these irrelevant reasons make the action reasonable. At the least, there is nothing more implausible or incoherent about this idea than the idea that, when reasons that are irrelevant to the appropriateness of envying S make not envying S reasonable, S could still be a fitting target of envy.

WKR also allows us to see more clearly where exactly defenders of BELS go wrong. To get at this, first note that there *are* attitudes which have fittingness conditions which include any reasons that feed into what an agent ought to do, or feel, all things considered. An example is the flavour of criticism that I above called *rational criticism.* This attitude is fitting only if a person acts unreasonably. It makes no sense to criticise someone for acting unreasonably when

[84] For a nice discussion of rational regret, see Hurka (1996).

they have not acted unreasonably. It would make no sense, for example, to regard Rebecca as foolish for mutilating her victims if she had sufficient reason to do so.

The mistake that defenders of BELS make is in assuming that moral blame is like rational criticism in having fittingness conditions that include any reason that feeds into what an agent has sufficient reason, all things considered, to do. This is evidently false if WKR is true. This is because one kind of reason that does feed into what an agent has sufficient reason to do – her prudential reasons to act immorally – is not among the fittingness conditions of moral blame. To illustrate, if WKR is true, the enjoyment that Rebecca takes in wrongly mutilating her victims does bear on whether her victims can appropriately consider her foolish for performing these actions, but not on whether they can appropriately resent her for performing these actions.

This assumption is evident in claims that defenders of BELS make. Return to Darwall's quote above:

> 'It seems incoherent… to blame while allowing that the wrong action, although recommended against by some reasons, was nonetheless the sensible thing to do, all things considered…. After all, if someone can show that he had good and sufficient reasons for acting as he did, it would seem that he *has* accounted for himself and defeated any claim that he is to blame for anything.'

Darwall here asserts that, if you can show that your action was 'sensible', then you will have 'defeated' any claim that you are to blame for performing that action. This clearly assumes that any reasons that bear on whether you had 'good and sufficient' reasons for acting as you did bear on whether you can be appropriately blamed. This is a mistake. As just noted, Rebecca's prudential reasons to mutilate her victims *can* show that her actions were sensible, but not that one of her mutilated victim's resentment is unfitting.

As with everything I have said in this section, these claims are only true if WKR is true. But if WKR *is* true, then it seems that we should accept (A) and reject BELS. This conditional has some intrinsic interest. It shows that BELS can be coherently rejected. More importantly, however, WKR is very plausible. This shows that defenders of BELS are committed to rejecting a very plausible claim. Since we should accept WKR – or so I have argued – we should accept (A) and reject BELS. I conclude that the blameworthiness argument fails to establish moral rationalism, or to show that moral anti-rationalism is false.

Now the groundwork has been laid, my argument can be restated as a direct argument against BELS. Here is one simple form that argument could take:

> If BELS is true, then Rebecca would not be morally blameworthy for acting immorally if her prudential reasons gave her sufficient reason to perform the immoral action.
>
> Rebecca would be morally blameworthy for acting immorally if her prudential reasons gave her sufficient reason to perform the immoral action.
>
> Therefore, BELS is false.

This argument is valid. The first premise cannot be denied. The claim that an agent would be blameworthy for φ-ing only if they lack sufficient reason to φ straightforwardly entails that, if Rebecca had sufficient reason to φ, then she would not be blameworthy for φ-ing. The second premise can be denied, but I have argued that it is a mistake to do so. Since both premises are sound and the argument is valid, we should reject BELS.

CHAPTER FIVE: DO WE WANT MORAL RATIONALISM TO BE TRUE?

This chapter will discuss a particular way of motivating moral rationalism. This is to argue that it would be bad, or undesirable, if moral anti-rationalism were true. I will first explain this argument – and how I will understand it – and then I will argue that it is unsuccessful.

I

To appreciate the force of this argument for moral rationalism, the most important thing to emphasise is that the debate between MR and AR concerns the *normative significance* of morality – it concerns how much the demands of morality really matter to how we ought to live our lives. Moral rationalism is the view that morality is, in one important sense, the *most important* normative domain. Its demands are unique in that, unlike other normative standards, they are never counterbalanced by competing considerations. At the end of the day, morality always wins. If moral anti-rationalism is true, then morality lacks this kind of importance; it does not matter this much. On any form of anti-rationalism, the answer to the question 'Should I be moral?' is sometimes 'no'.

It is *this* idea – that morality lacks this kind of importance – that we are supposed to find unappealing. In other words, we *want* moral demands to matter more than anything else. We want it to be true that, whenever an agent wonders whether they should comply with a moral requirement, the answer is always 'yes'. We want immoral agents to be making a mistake. Despite not being straightforwardly epistemic, the idea that moral anti-rationalism is unattractive has been taken to provide good reason to reject this view and endorse moral rationalism.

Let me give some examples of this style of argument. In her paper *Moral Overridingness and Moral Theory* (1998), Sarah Stroud argues that we have good reason to reject any moral theory that cannot be squared with moral rationalism. Any theory, that is, that issues verdicts that we don't plausibly have decisive reason to comply with. One of her main examples is consequentialism. As she (1998, 186) puts it, 'the apparent incompatibility of consequentialism with overridingness [moral rationalism] constitutes a count against consequentialism.' Indeed, she suggests, at one point, that this incompatibility may constitute a 'reductio' of consequentialism (1998, 187). Her argument for this claim is *not* that we should reject any such

theory because moral rationalism is true.[85] Rather, she claims that we have reason to reject any such theory because 'it would be a *desirable* result were we able to show that morality is overriding' (1998, 171). We would, she claims, '*prefer* to hold onto the idea of moral overridingness if we can' (1998, 176). In a similar vein, she also writes that our 'aspirations for morality's status' include vindicating its overridingness, and that 'we have reason to hope that overridingness can be sustained' (1998, 171).

Paul Hurley has made similar remarks concerning the apparent incompatibility of consequentialism and moral rationalism. He argues that addressing this tension by rejecting moral rationalism is not a good strategy for a consequentialist. In defending this claim, Hurley also never argues that there are good reasons to believe that moral rationalism is true. Rather, he argues that a consequentialist should not reject moral rationalism because the 'consequence is the threatened marginalization of morality' (2009, 26), and, similarly, that embracing anti-rationalism 'threatens the marginalization of her moral standards for rational agents' (2009, 33). This would be bad because it is in tension with a 'central aim of many consequentialists', namely 'to demonstrate that we should [all things considered] be doing more than moderate morality requires of us' (2009, 60). Given that what many consequentialists want is for their moral standards to determine what we ought to do, all things considered, vindicating consequentialism at the cost of moral rationalism will seem like a pyrrhic victory to a consequentialist because it entails that these standards *don't* determine what we ought to do, all things considered. In short, rejecting moral rationalism would vindicate consequentialism only by minimizing the normative significance of morality. This victory, Hurley supposes, will ring hollow. The claim that consequentialists *want* to vindicate is that, when consequentialism demands that we φ, we should *actually φ*. They want moral demands to be decisive.

Others have also claimed that moral anti-rationalism is unattractive. Thomas Nagel, for example, writes that as 'a matter of moral conviction, I myself am inclined against this possibility [anti-rationalism] … I am inclined strongly to hope, and less strongly to believe, that the correct moral theory will always have the preponderance of reasons on its side' (1986, 106). Reflecting on why many people reject anti-rationalism, he writes:

[85] This is Portmore's (2011) argument against agent-neutral forms of consequentialism.

> Since we don't want life to be like that, it is natural to hope that such theories are false… It may seem impossible that living as we have decisive reason to live should constitute a bad life, or a life that is less than optimal given our circumstances. It may seem impossible that an immoral life should be better than a moral one, or that a moral life should be a bad one. But I believe that what lies behind these impossibility claims is not ethics or logic but the conviction that things should not be that way – that it would be bad if they were. So it would. (1986, 109)

To give a final example, in their discussion of Parfit's version of the dualism of practical reason, de Lazari-Radek and Singer claim that 'any form of the dualism of practical reason undermines morality' (2014, 163). This is because, on any form of the dualism, an agent will sometimes have sufficient reason, all things considered, to act immorally. According to Singer and de Lazari-Radek, morality would only *not* be undermined – it would only be 'truly important' – if it was the case that 'when I choose to act wrongly, I am always acting contrary to a decisive reason' (2014, 163). That is, if moral rationalism was true. To conclude their discussion of the dualism, they write: 'If we want morality to be truly important, we need to be able to do better in overcoming the dualism' (2014, 163)). Given that, following this conditional, they attempt to show that we can overcome the dualism, it seems clear that they want, and assume that others want, for morality to be 'truly important'.

There are some important differences between these remarks. The most notable is that some are descriptive, while others are normative. Stroud, for instance, states both that we aspire to vindicate moral rationalism, and that we have reason to hope that moral rationalism is vindicated. Nonetheless, these remarks are all motivated by the same fundamental concern. This is that there is something unattractive about the idea that morality is less important to how we should ultimately live our lives than moral rationalism claims.

To focus the discussion, I will primarily examine descriptive versions of this idea. In particular, I will focus on the claim that we *want* moral rationalism to be true.[86] Of course, since this is ultimately an empirical claim, nothing that I say here will be decisive. Nonetheless, I believe that a strong case can be made that this claim is very unlikely to be true.

[86] Though what I say can easily be extended to other descriptive claims. For instance, that it seems bad if moral anti-rationalism is true, or that we hope that moral rationalism is true.

Though I mainly focus on this descriptive claim, it is important to note that my argument has implications for the normative claims as well. For one thing, as with Stroud, the normative claims are often defended by appealing to these putative psychological facts. As a result, if there is good reason to doubt these psychological claims, then there is good reason to doubt any argument that rests on them. In addition, if it is right, as I will argue, that most of us, if we thought it through carefully, would not want moral rationalism to be true, then that is good evidence that it makes sense not to want moral rationalism to be true. And I assume that, if it makes sense not to want moral rationalism to be true, then it makes sense not to hope that moral rationalism is true.

II

I first want to flag a potential problem. For this argument for moral rationalism to get off the ground, it needs to be the case that wanting P to be true gives us reason to endorse P. This is questionable. Although rejecting this claim seems like a promising way to respond to this argument, I will not pursue this strategy, for three reasons. First, it is not obvious that the claim is illegitimate. It doesn't seem that uncommon, for instance, to go beyond paradigmatically epistemic reasons for belief when choosing between competing theories, and, in any case, there may be pragmatic reasons for belief. Second, even if the claim is illegitimate, it has psychological force. As the above quotes indicate, people are in fact moved to reject moral anti-rationalism because it seems unattractive or bad to them. This makes this issue important to address if we want to make a persuasive case for moral anti-rationalism. Finally, since I will argue that we don't want moral rationalism to be true – and, indeed, that we want moral anti-rationalism to be true – I am happy to grant that, if the truth of some view is undesirable, then that is good reason to reject that view.

III

Let's assume, then, that we would have reason to reject moral anti-rationalism, and endorse moral rationalism, if it turned out that we wanted moral rationalism to be true. The next question is: *Do* we want rationalism to be true? A good place to start is with an (obvious) observation. Whether a person would want moral rationalism to be true depends entirely on their *attitude towards morality*. It depends, for instance, on how much they *care about* acting

morally. Given this, there is little reason to doubt that some people would want moral rationalism to be true. As a simple illustration, consider what I will call:

> *The Morally Motivated Agent (MA)*: This is a person who (a) has some set of first-order moral beliefs, (b) cares exclusively about morality, and (c) believes – and wants to believe (or hopes) – that we always have decisive reason to act as morality demands.

The MA is certainly conceivable. And, for a person like this, the truth of moral anti-rationalism is going to seem deeply unappealing. The reasons for this are all linked to the fact that, if she rejects moral rationalism, then she must accept that morality matters less than she had supposed, and less than she wants.

For instance, the MA is likely to derive some comfort from the thought that, when a person acts immorally, they are making a *mistake*. If the immoral agent thinks otherwise, they are at least wrong about this. Once the MA rejects moral rationalism, she will no longer be able to rely on this thought. There will be cases where she cannot honestly say that those who violate a moral demand had decisive reason to act differently. To get a sense of how depressing this could be, we only need to imagine what it would be like to learn that someone who performed an action that we consider monstrous was not in fact failing to live as they ought to live, all things considered. Discovering that the normative landscape was composed in this way would make the world seem like a much darker place. And although one can accept moral anti-rationalism without accepting that horrific actions are reasonable, it still seems plausible that, given her psychology, coming to accept that various immoral actions are reasonable would make the world seem like a much darker place to the MA.

The flipside of this is also true. The MA is likely to derive some comfort – and perhaps some degree of self-worth – from the thought that, whatever anyone else is doing, she is at least living as she has decisive reason to live. This may be false if moral anti-rationalism is true. Depending on the details of the view she comes to believe, she may come to believe that she has made various mistakes in how she has lived her life. Or, even if she has not made mistakes, she may still have denied herself various pleasures, or not pursued various desired ends, because she believed at the time that she lacked sufficient reason to indulge in these pleasures and undertake these pursuits due to them being morally impermissible. These revelations may lead to a significant amount of regret.

A final reason is that, if your values are constructed like the MA's, then coming to accept moral anti-rationalism seems apt to cause an existential crisis. After all, what you would be coming to accept is that what you cared about more than anything else was not *worth* caring about to the extent that you did. Having your values disintegrate in this way can be the source of considerable despair.

We can next ask: Does the concession that there are conceivable agents who would not want moral anti-rationalism to be true spell trouble for moral anti-rationalism? I take it that, for this to be so, attitudes of this sort would need to be relatively widespread. This is because it is difficult to see why this would be a significant problem if this was not the case. After all, some actual individual is likely to find – and some conceivable individual certainly will find – virtually any non-trivial normative proposition unattractive. It seems, then, that a stronger claim is required. This is that many or most people wouldn't want moral anti-rationalism to be true. The real issue, then, is whether the conceivability of such agents provides support for the stronger claim.

The answer to this seems to be 'no'. This is because we can also imagine people for whom the truth of moral anti-rationalism wouldn't seem unattractive. I shall discuss two examples. Consider first:

> *The Morally Indifferent Agent* (IA): Like the MA, the IA has some set of first order moral beliefs. Unlike the MA, the IA does not care, or cares very little, about living as he morally ought to live.

There are two versions of the IA. The first, who may not even be conceivable, is someone who doesn't care about *any* normative standard. Such a person simply doesn't care about living how they ought to live in any sense. The second, which I will focus on here, is someone who either uniquely doesn't care about morality or who includes morality among the subset of normative standards he doesn't care about. In contrast to acting morally, this person may be deeply

committed to acting in ways that make his own life go better, or to being the best dressed person in any room. [87]

The IA would clearly *not* want moral rationalism to be true. It would be very odd to want or hope that a normative standard that you don't care about is overriding in every case. Indeed, given that the IA *does* care about certain normative considerations, he is going to hope that, when morality conflicts with these other considerations, morality *loses*. He is going to hope that AR is true.

Consider next:

> *The Morally Alienated Agent* (AA): As with the other agents, the AA has some set of first order moral beliefs. She also believes that moral demands always provide us with decisive reason to act. Unlike the MA, however, the AA hope that morality matters less than she believes it does.

We may initially wonder why someone would have the AA's attitudes. But, on reflection, this is no mystery. On many conceptions – and on any plausible conception – morality can, in the right (or wrong) circumstances, demand that we make significant sacrifices. That we, for instance, not live as we most strongly desire to live, or that we perform actions that are detrimental to our own happiness or to the happiness of those we care about most. More fundamentally, morality is generally taken to be *inescapable*. Moral requirements apply to an agent *regardless* of her desires. For these reasons, morality can feel like an oppressive and alienating external force that crushes and constrains us.

It should be noted that one can believe that morality makes inescapable demands without experiencing these feelings. The IA, for instance, would not feel like the AA even if he believed that morality demanded some significant sacrifice. The relevant difference is that the AA feels *bound* to act morally – she feels that she *must* act as morality demands, all things considered. In other words, the reason that morality feels oppressive to the AA is precisely *because* she

[87] It is simplest to imagine the IA as an *amoralist*, and I will speak as if this is so. But the existence of the IA is also possible on plausible forms of motivational judgement internalism that say that, if an agent sincerely judges that she morally ought to Φ, then she will be motivated to Φ to at least some degree. For the purposes of my discussion, all that needs to be possible is that an agent can care more about non-moral reasons and requirements than she cares about moral reasons and requirements such that, when these conflict, the agent is always moved to perform the action that the non-moral reasons recommend. See Rosati (2016, esp. Section 3.2) for a helpful discussion of motivational judgement internalism.

believes that morality has the kind of normative significance that moral rationalism claims. As I hope is clear, the truth of moral anti-rationalism would come as a great relief to the AA.

To sum up, it is worth emphasising some connections between the MA and these other characters. The closest comparison is between the MA and the AA, since they both believe that moral rationalism is true. The difference between them is their other attitudes towards this proposition. While the binding nature of moral demands is a source of comfort to the MA, the AA feels imprisoned by the normative significance of moral demands.

The IA and the MA have less in common, but they do mirror each other in one important respect. Part of what explains why the MA wants moral rationalism to be true is that she cares about morality and doesn't care about other normative standards. Given this, it makes sense that she would want the considerations she cares about to always override the considerations that she doesn't care about. And part of what explains why the IA *doesn't* want moral rationalism to be true is that he cares about other normative standards and doesn't care about morality. Given this, it makes sense that he would want the considerations he cares about to override the considerations that he doesn't care about.

***

The conceivability of the IA and the AA is significant for various reasons. One of these has already been mentioned. Since we can easily imagine people for whom the falsity of moral rationalism would not seem undesirable, the mere fact that we can imagine people for whom it would does little to support the claim that the truth of moral rationalism would be desirable. At minimum, the issue at this level is a wash.

A stronger claim also seems defensible. Suppose it is correct that, for the descriptive version of the claim to be problematic for moral anti-rationalism, it must be the case that many or most people would prefer moral rationalism to be true. It then seems plausible to claim that the fact that we can easily conceive of people for whom this would not be so places the burden of proof on those who want to defend the descriptive claim. They must offer positive reasons for thinking that many people actually have the attitudes described, or at least that a sufficient number of people do not have the attitudes that would undermine this claim. This is because, if we can just as easily imagine people who don't have these attitudes as people who do, then the natural default would be to assume that while some people want moral rationalism to be true, others do not. And this is obviously not strong enough to support the view that many or

most of us want moral rationalism to be true. If this is right, then, all else equal, we should reject the descriptive claim.

Of course, all else may not be equal. To bolster my case, I shall now offer a direct argument against the claim that many or most of us want moral rationalism to be true. I will argue, in particular, that certain features that we are likely share with the IA and the AA give us very good reason to doubt that many or most people would, if they thought it through carefully, want moral rationalism to be true.

IV

We can begin by considering one feature that many of us are unlikely to share with the IA and the AA. This is the *intensity* and *pervasiveness* of their attitudes. It is doubtful that many of us feel every perceived moral demand as a crushing weight, or that most of us are entirely indifferent to every perceived moral demand. A similar point, of course, applies to the MA. It is farfetched to suppose that many of us care exclusively about living as morality demands, or even that we care about this significantly more than we care about anything else. Many of us, for instance, care at least as much about our own welfare, or the welfare of those we love. The attitudes of all these characters are, in this respect, unusual.

We can next note a respect in which the IA and the AA do not seem unusual. Given that the IA and the AA experience moral demands as they do, it seems natural – it makes sense – that they would not want moral rationalism to be true. It seems very plausible that, if we suddenly stopped caring about morality, then we would also not want moral demands to be decisive. And the same is true if – perhaps due to a transformation of our desires or life circumstances – morality suddenly conflicted with the things that we care about as much as, or more than, we care about morality. The explanation for this, in both cases, is the same as the explanation for why the IA and the AA don't want morality to be overriding. Why, for instance, would anyone want it to be true that, when there is a conflict between considerations one cares about more and considerations one cares about less, the considerations that one cares about less are decisive? That would be an odd attitude.

Consider next that it is very plausible that many or most of us would feel like the IA or the AA in response to certain moral demands that we could conceivably face. The assumption required to vindicate this claim is minimal. All that needs to be true is that, in some conceivable case,

morality could demand that you sacrifice something that you care about at least as much as you care about complying with that demand. This could be something very significant, such as your happiness or ambitions, or it could be something less significant, such as giving up some of your leisure time to help the less fortunate.

These two claims spell trouble for the idea that we want moral rationalism to be true. To see this, note what this idea entails. Moral rationalism, recall, is the view that an agent always has decisive reason, all things considered, to act as morality demands. As such, to want moral rationalism to be true *just is* to want it to be true that, in every possible case, you have decisive reason to act as morality demands. Given this, if there are any possible cases in which many or most of us would not want morality to be overriding, then many or most of us do not want moral rationalism to be true. If the above claim is correct, then there are possible cases where many or most of us would not want morality to be overriding. Indeed, if these claims are correct, then we want moral anti-rationalism to be true. We want it to be true that, in at least certain cases, we have sufficient reason, all things considered, to act immorally. Assuming that wanting P to be true gives us reason to endorse P, it seems that our attitudes do not give us reason to accept moral rationalism and reject moral anti-rationalism, but reason to reject moral rationalism and accept moral anti-rationalism.

V

I will now discuss two potential problems with this argument. The first is that the argument is misleading in the following way: I began by claiming that, in order not to want moral rationalism to be true, you do not need to have the extreme attitudes of the IA or the AA. What I didn't mention is that, in order to want moral rationalism to be true, you also don't need to have the extreme attitude of the MA. This omission may give the impression that I am assuming that, in order to want moral rationalism to be true, it is necessary to have very bizarre attitudes when this is not the case.

This is partly correct. It is not necessary to be exactly like the MA for it to make sense to want moral rationalism to be true. This is because it is not necessary to care exclusively about morality. It is enough to care about morality more than anything else. After all, if you care about morality more than anything else, then it seems clear that, when moral demands conflicts with something you care about less, you will want morality to be decisive. This point, however, does little to blunt the force of the argument. The psychological claim that needs to be true

remains implausibly strong. Given how many of us are constituted, it seems highly unlikely, for instance, that most of us would want morality to be decisive if it demanded that we harm those we love most.

Another potential problem with my argument is that it rests on a substantive assumption about the content of morality. This is that morality can demand that we sacrifice things that we care about as much as, or more than, we care about complying with moral demands. This assumption could be denied. It could be argued that morality does not make the kinds of demands that we don't want to be decisive.

As above, there is something to this. For one thing, it does seem right that whether moral rationalism strikes us as an attractive proposition depends in large part on how often we believe that acting morally is incompatible with pursuing other things that we care about. For another, it is certainly plausible to deny that conflicts of this kind are a common occurrence.

Once again, however, these claims do little to undermine the argument. The first point to make is that the plausible idea that morality infrequently conflicts with other things we care about is compatible with everything I have said. All the argument assumes is that there are conceivable circumstances – however unusual – in which there would be such conflicts.

Of course, even this very weak assumption could be denied. It could be claimed that morality never conflicts with anything we care about as much as, or more than, we care about acting as morality demands. If this is true, then, at minimum, it is difficult to see why we wouldn't want morality to be overriding. This is because, on this view, morality would never require us to do anything other than what we most want to do.

There are, however, numerous problems with making this move. The most obvious is that it is implausible that there are no conceivable cases where morality conflicts with other things we care about. As such, defending the descriptive claim in this way commits one to an implausible first-order moral theory.

Another problem is worth noting. This is that vindicating the desirability of moral rationalism in this way comes at a price that those who defend the descriptive claim are unlikely to be willing to pay. To see this, consider first that, on the face of it, whether moral rationalism is true seems to be an existentially significant question. If we are trying to decide how we ought to live, then it seems to matter whether moral demands are always decisive. After all, we often face conflicts between what appear to be moral and non-moral considerations. We may, for

instance, believe that we are morally required to sacrifice some of our resources to help the less fortunate, but also believe that we would be happier if we instead put these resources towards an endeavour that is significant only to us. When one is in the grip of this kind of situation, questions about the normative significance of morality loom large – this is the exact kind of situation that is apt to raise the 'why be moral?' question.

That the truth of moral rationalism is an important issue is a view that seems to be shared by the writers quoted above. Consider, for example, the opening passage of Stroud's (1998, 170) article:

> This article takes up a traditional question about the place of morality in lives animated by and mostly taken up with other concerns. Is morality simply one perspective among many in terms of its rational authority? Or does it take priority over our other commitments? Someone who violates a moral demand is, trivially, making a mistake from the point of view of morality. But is she necessarily making a mistake from any other, more general point of view? That is one way of putting the issue of whether morality is *overriding*.

As this passage makes clear, Stroud treats the question of whether morality is overriding as a significant one. What is on the line is whether we must give priority to morality over the other concerns and commitments that take up most of our lives. Nagel expresses a similar view. He writes that 'the impersonal element in any objective morality will be significant and depending on circumstances may become very demanding; it may overshadow everything else. In this chapter I want to discuss the tension… that results when these demands of impersonal morality are addressed to individuals who have their own lives to lead' (1986, 101), and that a moral theory with 'any significant requirements of impartiality can pose a serious threat to the kind of personal life that many of us take to be desirable' (1986, 102). Again, the issue, for Nagel, is how we should treat morality when its demands are in tension with other things that we care about.

Note next that, if you vindicate the desirability of moral rationalism by endorsing a moral theory that doesn't allow for the sorts of conflicts that my argument assumes, then what you have done, in effect, is vindicate this claim at the expense of the importance of moral rationalism. This is because it doesn't matter whether moral requirements are always decisive if they never conflict with other things we care about. Morality, on this view, would not have priority over our other concerns even if moral rationalism were true; its truth would make

virtually no difference at all to how we ought to live. This cannot be an appealing implication for those who think that whether moral rationalism is true is an important issue.

VI

To conclude, I shall discuss a final character. This is partly for completeness, but also because it may help us to diagnose what is going wrong when people object to moral anti-rationalism in this way. We can call this character:

> *The Reluctant Anti-Rationalist* (RAR): This is a person who (a) has some set of first-order moral beliefs, (b) cares exclusively about morality (or at least cares about this more than anything else), but (c) believes that moral anti-rationalism is true. They want morality to be overriding but cannot accept that it is.

Like the MA, the RAR cares about morality more than anything else. Unlike the MA, he has reluctantly come to believe that morality is not as significant as he wants it to be. For the same reasons as the MA, it is clear that the truth of moral anti-rationalism would be depressing to the RAR. Also for the same reasons as the MA, it is unlikely that many or most of us are psychologically similar to the RAR.

It is interesting to note, however, that some actual anti-rationalists have expressed negative attitudes towards moral anti-rationalism. One instance is related to an example already discussed. Though de Lazari-Radek and Singer make a general claim about the desirability of moral rationalism, the focus of their discussion is Parfit's version of the dualism of practical reason. Roughly put, this claims that, when morality and self-interest conflict, an agent will often have sufficient reason to act immorally. In other words, it will rarely be a mistake to ignore moral demands when it is in your interests to do so. Parfit (2016, 182) calls some of the implications of his view 'disturbing'. Henry Sidgwick, the most famous defender of the dualism, was also disturbed by his inability to reject the rationality of egoism, which he called 'a dubious guidance to an ignoble end' (1907/1962, 200). It is not only defenders of the dualism who feel this way. David Sobel (2007a, 14-15), who believes that moral anti-rationalism is an implication of the subjectivist theory of reasons that he endorses, writes that:

> One might say that large considerations of self-interest can defeat moral demands on the scale of what it makes most sense to do overall, yet continue to say that the overridden demands truly were the demands of morality… To someone who

> believes that the best account of practical reason will vindicate moral reasons as necessarily overriding for all agents, the above thoughts will undoubtedly seem disappointing and problematic. However, to someone persuaded that the best account of practical reason will not have this upshot – and there are a number of us – something like the above will seem to be a disappointment we must learn to live with.

Of course, the mere fact that some anti-rationalists are disturbed or disappointed by certain aspects of their view does not show that we want moral rationalism to be true. But reflecting on this issue does seem to bring out where the objection is going wrong. de Lazari-Radek and Singer are instructive here. They start from the claim that we don't want morality to only be as significant as the dualism implies. This may be true. From this claim, however, they move straight to the claim that we want morality to be decisive in every case – that 'when I choose to act wrongly, I am *always* acting contrary to a decisive reason' (2014, 163; my italics). This is a huge leap. There are countless anti-rationalist views between these extremes. These range from the view that you always have decisive reason to act immorally to the view that, in one particular case, you have sufficient reason to act immorally.[88] Since moral anti-rationalism is such a broad category, the fact that we don't want one specific form of anti-rationalism to be true provides almost no support for the claim that we want moral rationalism to be true. Others may be making the same mistake. They may be assuming that, if there are versions of moral anti-rationalism that we don't want to be true, then we must want moral rationalism to be true. Or perhaps they are assuming that, if there are various cases where we want morality to be overriding, then we must want morality to be overriding in every case. Both lines of reasoning are clearly flawed.

It is worth noting that not all actual anti-rationalists have been reluctant anti-rationalists. Nietzsche, for example, considered full commitment to morality the 'danger of dangers'. In large part, this is because he believed that complying with moral demands – and, more generally, possessing, developing and maintaining the dispositions necessary to be a morally decent person – could prevent a potential genius from producing great creative works. It would lead to them living more comfortably, but 'less dangerously and more basely'.[89] In a similar

[88] The fact that anti-rationalist theories come in weaker and stronger forms, and that this can make a significant difference to their plausibility, is a point that Dorsey has emphasised in other contexts (e.g. 2012, 7-11).
[89] These quotes come from Section 6 of the Preface of *On the Genealogy of Morality*. The translation is from Leiter (1997, 264).

vein, reflecting on Gauguin's decision to abandon his family and sail to Tahiti to paint, Bernard Williams (1981a, 23) writes that this case '…serves to remind us…that while we are sometimes guided by the notion that it would be the best of worlds in which morality were universally respected and all men were of a disposition to affirm it, we have in fact deep and persistent reasons to be grateful that that is not the world we have.' More generally, Williams (1985, 181-182) worries about moralities 'natural' tendency to rule out important non-moral considerations and how it 'can come to dominate a life altogether'. Susan Wolf makes similar claims. She writes, for instance, that:

> …it would be unrealistic and perhaps even undesirable to expect people to be committed to morality unconditionally. Even if, as one hopes, moral values reach to the very core of a person's identity, they are not, nor do we want them to be, the only values or attributes that comprise that core. (1992, 256)

And that 'our values cannot be fully comprehended on the model of a hierarchical system with morality on top' (1982, 438). More generally, if we believe that living as morality demands can be incompatible with living various other attractive lives, then we are unlikely to want moral requirements to always be decisive.

These points, however, go beyond the much simpler point that I want to make in this chapter. This is that, even if it is true that there are cases where we want morality to be decisive – and even if it is true that there are undesirable versions of moral anti-rationalism – this does not show that we don't want moral anti-rationalism to be true. On the other hand, the fact that there are cases where we don't want moral demands to be decisive *does* show that we don't want moral rationalism to be true, since to want moral rationalism to be true just is to want moral demands to be decisive in every case.

CONCLUSION

This thesis has defended moral anti-rationalism. This mainly involved making positive arguments for moral anti-rationalism and offering criticisms of arguments that have been given for moral rationalism. To conclude, I will briefly summarise my claims and make a few remarks about what I hope to have achieved.

I will start with the critical material. This is mainly found in Chapter Four and Chapter Five. In these chapters, I argued that two promising ways of defending moral rationalism fail. Chapter Four was a discussion of the blameworthiness defence of moral rationalism. I rejected a key premise of this defence, which claims that a person can only be morally blameworthy for freely and knowingly performing an action if they lacked sufficient reason, all things considered, to perform that action (BELS). We should reject BELS because there are reasons that make a difference to whether a person has sufficient reason to perform an action, but which are the wrong kinds of reasons to make it unfitting to blame the person for performing that action. These reasons do not bear on the fittingness of moral blame. The upshot of this is that, even if these reasons of the wrong kind gave someone sufficient reason to act immorally, they would still be morally blameworthy for freely and knowingly performing the immoral action. Chapter Five discussed a different style of argument for moral rationalism. This is that we have reason to accept moral rationalism because we want moral requirements to provide us with decisive reason to act. I argued that this defence of moral rationalism fails because it is highly unlikely that many or most of us want moral rationalism to be true. This becomes apparent when we recognise what wanting moral rationalism to be true amounts to. It is wanting it to be true that, in every conceivable case, it is a mistake not to act as morality demands. This will include cases, as rare or unlikely as they may be, where morality requires us to sacrifice the things that we care about. Few of us are likely to want morality to be decisive when it demands that we sacrifice our happiness, or that we abandon a project that is important to us.

These are clearly not the only arguments that have been given for moral rationalism. And, for all that I have said here, some other argument may be more successful.[90] But this does not show that the failure of these two arguments is inconsequential for the prospects of vindicating moral rationalism. Both arguments have force, and both have been offered by moral rationalists as

[90] There are certainly other arguments that I would like to spend more time thinking about. These include, among others, certain internalist/subjectivist arguments for MR, such as Julia Markovits' (2014) Kantian defence of moral rationalism. There are also arguments on related issues that I would like to further explore, such as Alison Hills' (2010) epistemic argument against rational egoism.

good reasons to accept MR. The blameworthiness argument, in particular, is the best argument for moral rationalism that I have come across. If some of the best arguments for moral rationalism fail, then that gives us reason to doubt that any argument will vindicate moral rationalism. This failure also provides at least some support for moral anti-rationalism. This is strengthened if it is also the case that there are compelling positive arguments for moral anti-rationalism. This is what I attempted to provide in Chapter Two and Chapter Three.

Before discussing the earlier chapters, it is perhaps worth nothing that the division between positive and critical material in this thesis is not as sharp as the above discussion may suggest. The critical chapters also included some positive material, and the positive chapters also included some critical material. In arguing against the claim that we want moral rationalism to be true, for example, I tried to motivate the idea that we in fact want moral anti-rationalism to be true. To the extent that this kind of reasoning can support a view, this gives us reason to accept moral anti-rationalism. And in arguing that we should accept that certain prudential reasons can give an agent sufficient reason to act immorally, I also tried to show that recognising that there are morally irrelevant non-moral reasons allows us to see what is wrong with arguments for moral rationalism that rely on a conception of deontic statuses that holds that an action is morally permissible whenever an agent has sufficient reason to perform that action.

Turning to the positive chapters, I will start by saying something about Chapter One. This was a discussion of excellence-based reasons. The claims that I defended in this chapter can be accepted by moral anti-rationalists and moral rationalists alike. I first argued that agents have normatively significant reasons to perform actions that lead to the achievement of aesthetic or intellectual excellence. I then argued that these reasons are neither moral nor prudential reasons. While accepting these claims is compatible with both MR and AR, recognising these reasons does help to motivate the plausibility of moral anti-rationalism. One reason for this is that it helps us to avoid an incorrect picture of what is on the line in the debate between these views. When we think of questions about the normative significance of morality only in the context of conflicts between morality and prudence, we may be led to believe that moral rationalism is both more attractive and easier to defend than it in fact is. It may seem more attractive because it may appear that all that AR does is provide a justification for morally

excessive selfishness.[91] Recognising that there are also excellence-based reasons helps us to see that our non-moral reasons for action are rich and diverse. There is much more at stake when moral demands and non-moral reasons conflict than just our self-interest. Thinking only of conflicts between morality and prudence may make MR seem easier to defend than it is because it is a feature of many moral theories – and perhaps of our everyday moral thinking – that morality already gives us significant leeway to pursue our own good. This may suggest that moral rationalists already have a plausible way of explaining why it is sometimes reasonable to act on non-moral reasons even when there are good moral reasons to do otherwise. Once we accept that there are non-moral reasons that are not tied to an agent's own good, we can see that this familiar idea is not going to be enough on its own to vindicate moral rationalism. In short, recognising these additional non-moral reasons increases the difficulty, and diminishes the appeal, of defending moral rationalism.

Chapter Two and Chapter Three defended moral anti-rationalism. I argued that both prudential and excellence-based reasons can provide an agent with sufficient reason, all things considered, to act immorally. Chapter Two discussed prudential reasons. I began by arguing – following others – that there is little hope of providing a plausible defence of moral rationalism if we accept agent-neutral consequentialism. The best chance is by accepting a version of COST. As well as being plausible, COST promises to provide the resources to give compelling rationalist-friendly explanations of cases where it seems reasonable for an agent to act in her own interests when this is not morally best. I then argued that COST cannot provide plausible rationalist-friendly interpretations of all such cases, and hence cannot vindicate MR. This is because cases like *The Voyeur* demonstrate that there are prudential reasons that are *morally irrelevant*. These prudential reasons make a difference to how an agent should live, all things considered, but do not make a difference to the moral permissibility of an action. As well as being a problem for vindicating MR using COST, the existence of these prudential reasons gives us a general reason to reject moral rationalism. This is because there doesn't seem to be any plausible way of explaining why, when these kinds of prudential reasons conflict with moral requirements – and especially with weak moral requirements – they never provide an agent with at least sufficient reason to act.

[91] See Catherine Wilson (1993) for a forceful argument along these lines. More specifically, she argues that anti-rationalists and defenders of COST are both guilty of trying to justify their own privileged and comfortable ways of life.

In Chapter Three, I argued that this same line of reasoning extends to excellence-based reasons. These are also morally irrelevant reasons, and they can also conflict with moral requirements. This makes it hard to see how it could be plausibly explained why, when these reasons conflict with moral requirements – and especially with weak moral requirements – they never provide an agent with at least sufficient reason to act. I then argued that EBRs raise a distinct problem for the plausibility of moral rationalism. This is that, if there are such conflicts between morality and excellence, then accepting MR would have the implication that it is unreasonable to live certain lives that we find attractive and admirable. This gives the argument for AR from excellence-based reasons a kind of intuitive force that the argument from prudential reasons may lack. As noted, it might seem appealing that one of the implications of moral rationalism is that morally excessive self-interest is unreasonable. But the idea that pursuing and achieving aesthetic or intellectual excellence can be similarly unreasonable is much less appealing.

Throughout this thesis, I have gestured at what I believe to be the deepest reason to accept moral anti-rationalism and reject moral rationalism. This is that we cannot square moral rationalism with the idea that there are many worthwhile but incompatible ways that we could live our lives. It is not plausible that all these worthwhile lives are morally permissible, and nor is it plausible, given the significance of the non-moral reasons that favour them, that each of these impermissible lives is unreasonable. This does not suggest that living a moral life is itself a mistake. Indeed, living a moral life seems to be one of the worthwhile lives that we could live. What it does suggest is that it is not always a mistake to refuse to live a moral life given what else can be on offer. While I find this line of reasoning compelling – and I believe that at least aspects of this idea are behind a number of arguments that have been given for moral anti-rationalism – it is, when expressed in this way, just the sketch of a convincing argument. One thing that I hope my positive arguments have achieved is to have filled in some of the necessary details. I hope to have shown, for instance, that the price of living a moral life could be your happiness, the achievement of aesthetic or intellectual excellence, or both. To put this another way, it may be impossible for someone to live a happy or satisfying life, or to create excellent aesthetic or intellectual works, if they choose to live a moral life. When I vividly imagine being faced with such a choice, it seems far from implausible that it is reasonable to refuse to live a miserable life, or reasonable to pursue aesthetic or intellectual excellence.

It should perhaps be stressed, as a final point, that my arguments give us reason to accept moral anti-rationalism whatever we think of this general picture of the normative landscape. There are, I have argued, two kinds of reasons that make a difference to whether an agent has

sufficient reason, all things considered, to perform an action, but that do not make a difference to whether performing this action is morally permissible or morally required. These morally irrelevant non-moral reasons can conflict with moral requirements. It is perhaps plausible that, in every such conflict, an agent has sufficient reason, all things considered, to act as morality demands. But, given their normative significance, it is not plausible that, in every such conflict, these morally irrelevant reasons never provide an agent with even sufficient reason, all things considered, to act immorally.

BIBLIOGRAPHY

- Baker, Derek. “The Varieties of Normativity.” In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plumkett. New York: Routledge, 2018a.

- Baker, Derek. “Skepticism about Ought Simpliciter.” In *Oxford Studies in Metaethics,* Vol. 13, edited by Russ Schafer-Landau. Oxford, New York: Oxford University Press, 2018b.

- Barry, Christian, and Gerhard Overland. *Responding to Global Poverty: Harm, Responsibility, and Agency*. Cambridge: Cambridge University Press, 2016.

- Benn, Claire, and Adam Bales. “The Rationally Supererogatory.” *Mind*, Vol. 129, No. 515 (2020).

- Bradford, Gwen. “The Value of Achievements.” *Pacific Philosophical Quarterly*, Vol. 94, No. 2 (2013).

- Bradford, Gwen. *Achievement*. Oxford: Oxford University Press, 2015.

- Brink, David O. *Moral Realism and the foundations of Ethics*. Cambridge; New York: Cambridge University Press, 1989.

- Butler, Joseph. *Fifteen Sermons Preached at the Rolls Chapel and Other Writings on Ethics*. Edited by David McNaughton. Oxford: Oxford University Press, 2017.

- Carbonell, Vanessa. “What Moral Saints Look Like.” *Canadian Journal of Philosophy*, Vol. 63, No. 3 (2009).

- Carbonell, Vanessa. “De Dicto Desires and Morality as a Fetish.” *Philosophical Studies*, Vol. 163, Issue 2 (2013).

- Connolly, Cyril. *Enemies of Promise*. London: Routledge & Kegan Paul, 1938.

- Copp, David. “Does Moral Theory Need the Concept of Society?” *Analyse & Kritik*, Vol. 19, No. 2 (1997).

- Copp, David. “The Ring of Gyges: Overridingness and the Unity of Reason.” In *Morality in a Natural World*. Cambridge; New York: Cambridge University Press, 2007.

- Cullity, Garrett. “Weighing Reasons.” In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star. Oxford; New York: Oxford University Press, 2018.

- D'Arms, Justin. "Envy." In *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), edited by Edward N. Zalta.

- D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophical and Phenomenological Research,* Vol. 61, No 1 (2000a).

- D'Arms, Justin, and Daniel Jacobson. "Sentiment and Value." *Ethics*, Vol. 110, No. 4 (2000b).

- Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, Mass.: Harvard University Press, 2006a.

- Darwall, Stephen. "Morality and Practical Reason: A Kantian Approach." In *The Oxford Handbook of Ethical Theory*, edited by David Copp. Oxford: Oxford University Press, 2006b.

- Darwall, Stephen. "Making the 'Hard' Problem of Moral Normativity Easier." In *Weighing Reasons*, edited by Errol Lord and Barry Maguire. Oxford: Oxford University Press, 2016.

- de Lazari-Radek, Katarzyna, and Peter Singer. *The Point of View of the Universe: Sidgwick and Contemporary Ethics.* New York: Oxford University Press, 2014.

- Dorsey, Dale. "Three Arguments for Perfectionism." *Nous*, Vol. 44, No. 1 (2010).

- Dorsey, Dale. "Weak Anti-Rationalism and the Demands of Morality." *Nous*, Vol. 46, No. 1 (2012)

- Dorsey, Dale. "Two Dualisms of Practical Reason." In *Oxford Studies in Metaethics*, Vol. 8, edited by Russ Shafer-Landau. Oxford: Oxford University Press, 2013.

- Dorsey, Dale. "How *Not* to Argue Against Consequentialism." *Philosophy and Phenomenological Research*, Vol. 90, No. 1 (2015).

- Dorsey, Dale. *The Limits of Moral Authority*. Oxford; New York: Oxford University Press, 2016.

- Eliot, T.S. *The Letters of T.S. Eliot, Volume 1: 1898-1922*. Edited by Valerie Eliot and Hugh Haughton. London: Faber and Faber, 2011.

- Finlay, Stephen, and Mark Schroeder. "Reasons for Action: Internal vs. External." In *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), edited by Edward N. Zalta.

- Fletcher, Guy. *Dear Prudence: The Nature and Normativity of Prudential Discourse*. Oxford: Oxford University Press, 2021.

- Foot, Philippa. "Morality and Art." *Proceedings of the British Academy*, Vol. 56 (1970).

- Foot, Philippa. "Morality as a System of Hypothetical Imperatives." *The Philosophical Review*, Vol. 81, No. 3 (1972).

- Foot, Phillipa. *Natural Goodness*. Oxford; New York: Oxford University Press, 2001.

- Gardner, Sebastian. "Tragedy, Morality and Metaphysics." In *Art and Morality*, edited by Jose Luis Bermudez and Sebastian Gardner. London: Routledge.

- Gert, Joshua. *Brute Rationality: Normativity and Human Action*. Cambridge; New York: Cambridge University Press, 2004.

- Gert, Joshua. "Moral Rationalism and Commonsense Consequentialism." *Philosophy and Phenomenological Research*, Vol. 88, No. 1 (2014).

-Gertken, Jan, and Benjamin Kiesewetter. "The Right and the Wrong Kind of Reasons." *Philosophy Compass*, Vol. 12, No. 5 (2017).

- Graham, Peter A. "In Defence of Objectivism about Moral Obligation." *Ethics*, Vol. 121, No. 1 (2010).

- Graham, Peter A. "A Sketch of a Theory of Moral Blameworthiness." *Philosophy and Phenomenological Research*, Vol. 88, No. 2 (2014).

- Graham, Peter A. "Two Arguments for Objectivism about Moral Permissibility." *Australasian Journal of Philosophy*, Vol. 99, No. 1 (2021).

- Hannay, Alastair. *Kierkegaard*. London; Boston: Routledge & Kegan Paul, 1982.

- Haybron, Dan. "Well-Being and Virtue." *Journal of Ethics and Social Philosophy*, Vol. 2, No. 2 (2007).

- Haydar, Bashshar. "Special Responsibility and the Appeal to Cost." *The Journal of Political Philosophy*, Vol. 17, No. 2 (2009).

- Heyd, David. *Supererogation: Its Status in Ethical Theory*. Cambridge: Cambridge University Press, 1982.

- Heyd, David. “Supererogation.” In *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), edited by Edward N. Zalta.

- Hills, Alison. *The Beloved Self: Morality and the Challenge from Egoism*. Oxford: Oxford University Press, 2010.

- Hurka, Thomas. *Perfectionism*. New York: Oxford University Press, 1993.

- Hurka, Thomas. “Monism, Pluralism, and Rational Regret.” *Ethics*, Vol. 106, No. 3 (1996).

- Hurka, Thomas. “Nietzsche: Perfectionist.” In *Nietzsche and Morality*, edited by Brian Leiter and Neil Sinhababu. New York: Oxford University Press, 2007.

- Hurley, Paul E. “Does Consequentialism Make Too Many Demands, or None at All?” *Ethics*, Vol. 116, No. 4 (2006).

- Hurley, Paul. *Beyond Consequentialism*. Oxford: Oxford University Press, 2009.

- Jackson, Frank. “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.” *Ethics*, Vol. 101, No. 3 (1991).

- Joyce, Richard. *The Myth of Morality.* Cambridge; New York: Cambridge University Press, 2001.

- Joyce, Richard. “Morality, Schmorality.” In *Morality and Self-Interest*, edited by Paul Bloomfield. Oxford; New York: Oxford University Press, 2007.

- Kagan, Shelly. *The Limits of Morality*. Oxford; New York: Oxford University Press, 1989.

- Kagan, Shelly. “Rethinking Intrinsic Value.” *The Journal of Ethics*, Vol. 2, No. 4 (1998).

- Kiesewetter, Benjamin. *The Normativity of Rationality*. Oxford; New York: Oxford University Press, 2017.

- King, Alex. “Reasons, Normativity, and Value in Aesthetics.” *Philosophy Compass*, Vol. 17, No. 1 (2022).

- Korsgaard, Christine M. “Two Distinctions in Goodness.” *Philosophical Review*, Vol. 92, No. 2 (1983).

- Langton, Rae. “Objective and Unconditional Value.” *The Philosophical Review*, Vol. 116, No. 2 (2007).

- Lazar, Seth. “Deontological Decision Theory and Agent-Centred Options.” *Ethics*, Vol. 127, No. 3 (2017).

- Lazar, Seth. “Moral Status and Agent-Centred Options.” *Utilitas*, Vol. 31, No. 1 (2019).

- Leiter, Brian. “Nietzsche and the Morality Critics.” *Ethics*, Vol. 107, No. 2 (1997).

- Leiter, Brian. *Nietzsche on Morality*. New York: Routledge, 2014.

- Look, Brandon C. “Gottfried Wilhelm Leibniz.” In *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), edited by Edward N. Zalta.

- Lowrie, Walter. *A Short Life of Kierkegaard*. Princeton: Princeton University Press, 2013.

- Markovits, Julia. *Moral Reason*. Oxford: Oxford University Press, 2014.

- McGonigal, Andrew. “Aesthetic Reasons.” In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star. Oxford; New York: Oxford University Press, 2018.

- McLeod, Owen. “Just Plain ‘Ought’.” *The Journal of Ethics*, No. 5, Vol. 4, 2001.

- Nagel, Thomas. “The Fragmentation of Value.” In *Mortal Questions*. New York: Cambridge University Press, 1979.

- Nagel, Thomas. “Living Right and Living Well.” In *The View from Nowhere*. New York: Oxford University Press, 1986.

- Parfit, Derek. *Reasons and Persons*. Oxford; New York: Oxford University Press, 1984.

- Parfit, Derek. *On What Matters*, Vol. 1. Oxford; New York: Oxford University Press, 2011.

- Parfit, Derek. “Conflicting Reasons.” *Etica & Politica/Ethics & Politics*, Vol. 18, No. 1 (2016).

- Paul, L. A. *Transformative Experience*. Oxford: Oxford University Press, 2014.

- Pettit, Philip. *The Birth of Ethics*. Oxford: Oxford University Press, 2018.

- Pettit, Philip. *When Minds Converse*. Oxford: Oxford University Press, Forthcoming.

- Plato. *The Republic*. Translated by G.M.A. Grube, Revised by C.D.C. Reeve. Indianapolis; Cambridge: Hackett Publishing Company, 1992.

- Poe, Edgar Allan. “The Philosophy of Composition.” In *The Portable Edgar Allan Poe*, edited by J. Gerald Kennedy. New York: Penguin Books, 2006.

- Portmore, Douglas W. *Commonsense Consequentialism: Wherein Morality Meets Rationality.* New York: Oxford University Press, 2011.

- Portmore, Douglas W. "Replies to Gert, Hurley, and Tenenbaum." *Philosophy and Phenomenological Research,* Vol. 88, No. 1 (2014).

- Railton, Peter. "Moral Realism". *The Philosophical Review*, Vol. 95, No. 2 (1986a).

- Railton, Peter. "Facts and Values." *Philosophical Topics*, Vol. 14, No. 2 (1986b).

- Railton, Peter. "Some Questions About the Justification of Morality." *Philosophical Perspectives,* Vol. 6 (1992).

- Railton, Peter. "Humean Theory of Practical Rationality". In *Oxford Handbook of Ethical Theory,* edited by David Copp. Oxford: Oxford University Press, 2006.

- Rawls, John. *A Theory of Justice*. Cambridge, Mass.: Belknap Press, 1971.

- Reisner, Andrew. "The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem." *Philosophical Studies,* Vol. 145, No. 2 (2009).

- Reisner, Andrew. "Pragmatic Reasons for Belief." In In *The Oxford Handbook of Reasons and Normativity,* edited by Daniel Star. Oxford; New York: Oxford University Press, 2018.

- Rosati, Connie S. "Moral Motivation." In *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), edited by Edward N. Zalta.

- Rule, Anne. *Green River, Running Red.* New York: Pocket Books, 2004.

- Scanlon, T.M. *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press, 1998.

- Scheffler, Samuel. *The Rejection of Consequentialism.* Oxford; New York: Oxford University Press, 1982.

- Scheffler, Samuel. *Human Morality*. Oxford: Oxford University Press, 1994.

- Schroeder, Mark. *Slaves of the Passions*. Oxford; New York: Oxford University Press, 2007.

- Schroeder, Mark. "The Ubiquity of State-Given Reasons." *Ethics,* Vol. 122, No. 3 (2012).

- Sepielli, Andrew. "Subjective and Objective Reasons." In *The Oxford Handbook of Reasons and Normativity,* edited by Daniel Star. Oxford; New York: Oxford University Press, 2018.

- Shafer-Landau, Russ. *Moral Realism*. Oxford; New York: Oxford University Press, 2003.

- Sidgwick, Henry. *The Methods of Ethics: Seventh Edition*. London: Palgrave Macmillan, 1907/1962.

- Skorupski, John. *Ethical Explorations*. Oxford: Oxford University Press, 1999.

- Shaver, Robert. *Rational Egoism: A Selective and Critical History*. Cambridge: Cambridge University Press, 1998.

- Shaver, Robert. "Egoism". In *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), edited by Edward N. Zalta.

- Slote, Michael. "Rational Dilemmas and Rational Supererogation." *Philosophical Topics*, Vol. 14, No. 2 (1986).

- Smith, Michael. *The Moral Problem.* Oxford: Blackwell, 1994.

- Smith, Michael. "Humean Rationality." In *The Oxford Handbook of Rationality*, edited by Piers Rawling & Alfred R. Mele. Oxford: Oxford University Press, 2004.

- Smith, Michael. "Three Kinds of Moral Rationalism." In *The Many Moral Rationalisms*, edited by Karen Jones & Francois Schroeter. Oxford: Oxford University Press, 2018.

- Sobel, David. "The Impotence of the Demandingness Objection." *Philosophers' Imprint*, Vol. 7 (2007a).

- Sobel, David. "Subjectivism and Blame." *Canadian Journal of Philosophy Supplementary Volume*, Vol. 33 (2007b).

- Sobel, David. *From Valuing to Value: A Defence of Subjectivism*. Oxford: Oxford University Press, 2016.

- Strawson, Peter. "Freedom and Resentment." *Proceedings of the British Academy*, Vol. 48 (1962).

- Stroud, Sarah. "Moral Overridingness and Moral Theory." *Pacific Philosophical Quarterly*, Vol. 79, No. 2 (1998).

- Thomason, Krista K. "The Moral Value of Envy." *The Southern Journal of Philosophy*, Vol. 53, No. 1 (2015).

- Tiffany, Evan. “Deflationary Normative Pluralism.” *Canadian Journal of Philosophy*, Vol. 37, No. 5 (2007).

- Wall, Steven. “Perfectionism in Moral and Political Philosophy.” In *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), edited by Edward N. Zalta.

- Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, Mass.: Harvard University Press, 1994.

- Williams, Bernard. “A Critique of Utilitarianism.” In *Utilitarianism: For and Against*, with J.J.C Smart. Cambridge: Cambridge University Press, 1973.

- Williams, Bernard. “Moral Luck”. In *Moral Luck*. Cambridge; New York: Cambridge University Press, 1981a.

- Williams, Bernard. “Internal and External Reasons.” In *Moral Luck*. Cambridge; New York: Cambridge University Press, 1981b.

- Williams, Bernard. “Persons, Character and Morality.” In *Moral Luck*. Cambridge; New York: Cambridge University Press, 1981c.

- Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge, Mass.: Harvard University Press, 1985.

- Williams, Bernard. “Internal Reasons and the Obscurity of Blame.” In *Making Sense of Humanity*. Cambridge; New York: Cambridge University Press, 1995.

- Wilson, Catherine. “On Some Alleged Limits to Moral Endeavour.” *The Journal of Philosophy*, Vol. 90, No. 6 (1993).

- Wolf, Susan. “Moral Saints.” *The Journal of Philosophy*, Vol. 79, No. 8 (1982).

- Wolf, Susan. “Morality and Partiality.” *Philosophical Perspectives*, Vol. 6 (1992).

- Wolff, Jonathan. *Why Read Marx Today?* New York: Oxford University Press, 2002.

- Worsnip, Alex. “Eliminating Prudential Reasons.” In *Oxford Studies in Normative Ethics*, Vol. 8, edited by Mark C. Timmons. Oxford: Oxford University Press, 2018.