

THESIS

MORALITY AND REASONS TO ACT

Submitted by

Joseph Gavronsky

Department of Philosophy

In partial fulfillment of the requirements

For the Degree of Master of Arts

Colorado State University

Fort Collins, Colorado

Fall 2012

Master's Committee:

Advisor: Elizabeth Tropman

Michael Losonsky

Kelly Martin

## ABSTRACT

### MORALITY AND REASONS TO ACT

The main goal of this thesis is to shed some light on the nature of reasons to act, and the nature of the relationship between morality and reasons to act, through a defense of rational egoism. Rational egoism holds that an agent's reasons to act are grounded by his self-interest, which is conceived as something different from, and not relativized to, his desires. In other words, for a rational egoist, an agent is rational if he does what is, in fact, good for him. I develop a version of rational egoism, and then argue that my version of rational egoism is at least as viable as, if not preferable to, other theories about the nature of reasons to act, and the nature of the relationship between morality and reasons to act. I claim that rational egoism provides a uniquely compelling account of the nature of reasons to act, and the nature of the relationship between morality and reasons to act.

The rational egoism that I endorse treads a sort of theoretical middle ground between its most compelling competitor theories, capturing their theoretical merits and yet avoiding the problems that they are vulnerable to. One of rational egoism's most compelling competitor theories holds that it is morality itself that necessarily provides agents with reasons to act. This theory, known as intrinsic moral rationalism, and advanced by Russ Shafer-Landau, is most compelling for its compatibility with moral convention, but is troubled in that it seems to confer upon morality a mysterious force that allows it to impose upon agents. I will argue that my rational egoism is also compatible with moral convention, and yet not mysterious in the troubling way that intrinsic moral rationalism is.

The other of rational egoism's most compelling competitor theories, which I refer to as *the desire-satisfaction view*, holds that it is an agent's desires, in some sense, that necessarily provide agents with reasons to act. In its most basic form, the desire-satisfaction view holds that it is an agent's actual desires that necessarily provide him with reasons to act. More sophisticated versions of the desire-satisfaction view, like Bernard Williams' view, for example, hold that an agent's desires, qualified in some way, necessarily provide him with reasons to act. The desire satisfaction view, in general, is most compelling for the prominence it gives desires, yet troubled by its commitment to the proposition that even an agent's desires for things that are cruel, self-destructive, or meaningless, nonetheless provide the agent with reasons to act. And although Williams' view and other sophisticated versions of the desire-satisfaction view may be able to get out of this commitment, it would take too much unmotivated theoretical machinery for them to do so. I will argue that my rational egoism also gives due regard to desires, yet avoids the troubling commitment of the desire-satisfaction view, and does so without appeal to any unmotivated theoretical machinery.

My rational egoism is not, itself, necessarily committed to one particular view regarding the nature of the relationship between morality and reasons to act. However, I strongly suspect that if rational egoism is true, then there is always, or almost always, a reason to do the moral thing. On one normative moral theory, namely moral egoism, morality just requires agents to do that which is in their self-interest. If moral egoism is true, and there are indeed grounds for believing that it is, and rational egoism is true, then there will always be a reason to do the moral thing. But even if moral egoism is not true, I argue that there are solid grounds for believing that morality usually indicates that agents should do that which is in their self-interest. In this case, if rational egoism is true, there will usually be a reason to do the moral thing.

## TABLE OF CONTENTS

Introduction.....	1
Chapter 1 Against Intrinsic Moral Rationalism .....	5
1.1 Introducing Moral Rationalism, and Distinguishing Our Main Concern from Related Moral and Metamoral Concerns.....	6
1.2 Moral Rationalism as a Solution to the Metamoral Question .....	16
1.3 A Presumptive Case in Favor of Moral Rationalism.....	18
1.4 A Strike Against Intrinsic Moral Rationalism .....	23
1.5 Two Intrinsic Moral Rationalist Replies .....	27
1.6 Conclusion .....	30
Chapter 2 Against the Desire-Satisfaction View .....	32
2.1 The Actual Desire-Satisfaction View .....	32
2.2 Problems for the ADS View .....	34
2.3 ADS View Reconsidered.....	36
2.4 Idealized Desire-Satisfaction Views.....	40
2.5 Williams' IDS View .....	41
2.6 Williams' View and the ADS Problem .....	43
2.7 Williams' View and the IDS Problem.....	46
2.8 Independent Support for Idealization.....	50
2.9 Conclusion .....	54
Chapter 3 Self-Interest and Reasons to Act .....	56
3.1 What is Self-Interest?.....	57
3.2 Rational Egoism .....	61
3.3 Rational Egoism and the ADS Problem .....	63
3.4 Rational Egoism and the Theoretical Virtue of the Desire-Satisfaction View .....	68
3.5 Rational Egoism and the IDS Problem .....	70
3.6 Rational Egoism and the Problem with Intrinsic Moral Rationalism .....	72
3.7 Rational Egoism, Moral Egoism, and Extrinsic Moral Rationalism.....	75
3.8 Conclusion .....	89

## Introduction

Some action may be the morally required action, but is there necessarily any reason to do it? Does morality, itself, necessarily provide us with reasons to do the moral thing? Assuming it would be the moral thing to rescue a homeless puppy, is it the rational thing to do, for the man who cares nothing about morality or puppies? If not morality, what sorts of considerations ground our reasons to act? What reasons do we have to do anything at all? These are the sorts of concerns that I will worry about in the following thesis.

I suspect that many people feel compelled to do the moral thing, whatever they think the moral thing to do is, and that they feel some sort of guilt or remorse when they do not do what they take to be the moral thing. I imagine that morality feels to them like some powerful force hanging over them, and imposing upon them reasons to act. Others, however, may have let go of this compulsion, given up on morality, or just do not experience it as a powerful, imposing, influence upon their lives and practical reasoning. Perhaps, when it comes to making choices, this latter class gives more weight to whatever it is they desire, or what would be best, in a non-moral sense, for them to do, or what would be the most prudent thing to do. In what follows, I will attempt to thoroughly and rigorously evaluate what I take to be the most plausible theories about the relationship between morality and reasons to act. The main theories that I will consider go, roughly, as follows:

(1) Intrinsic moral rationalism is the thesis that morality *itself* necessarily provides agents with reasons to act. That is, there is necessarily a reason to do the moral thing, and the reason is just that it is the moral thing. The view is that morality itself, by its very nature, necessarily provides agents with a reason to do the moral thing. Supposing lying is morally wrong, the

intrinsic moral rationalist would hold that this fact, itself, necessarily provides agents with a reason to refrain from lying.<sup>1</sup>

(2) Extrinsic moral rationalism is the thesis that morality necessarily provides agents with a reason to act in virtue of its necessary connection with some other intrinsically reason-giving kind of consideration such as desire or self-interest. That is, there is necessarily a reason to do the moral thing because there is necessarily a reason to do what one desires, or what is in one's self-interest, and there is a necessary connection between morality and desire, or self-interest. Extrinsic moral rationalists can be classified according to which other consideration (desire or self-interest or some other consideration) they believe intrinsically and necessarily provides agents with a reason to act.

(3) Moral antirationalism is the thesis that morality does not necessarily provide agents with a reason to do the moral thing.<sup>2</sup> Moral antirationalists may hold that morality usually, sometimes, or never provides agents with a reason to do the moral thing.

---

<sup>1</sup>The word, *fact*, may here be misleading. I do not mean to suggest that intrinsic moral rationalists are committed to moral realism, the thesis that there are objective moral facts.

<sup>2</sup>I follow Shafer-Landau's use of the terms: *intrinsic moral rationalism*, *extrinsic moral rationalism* and *moral antirationalism*. Russ Shafer-Landau, *Moral Realism: a Defence*. Oxford: Clarendon Press, 2003, 193-204. My language should not be confused with the language of internalism/externalism which might be more familiar to some readers. David Brink calls the theory that morality itself necessarily provides agents with reasons to act *internalism about reasons*, and he calls the denial of internalism about reasons *externalism about reasons*. Brink's *internalism about reasons* matches up quite well with Shafer-Landau's *intrinsic moral rationalism*. And Brink's *externalism about reasons* is compatible with both what Shafer-Landau calls *extrinsic moral rationalism* and what Shafer-Landau calls *moral antirationalism*. *Externalism about reasons*, for Brink, is the view that it is not the case that morality itself necessarily provides agents with reasons to act. *Externalism about reasons*, then, could allow that morality does not necessarily provide agents with reasons to act, or that morality does necessarily provide agents with reasons to act, but not in virtue of morality itself, rather in virtue of some connection between morality and something else that is itself necessarily reason-providing. David Brink, *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press, 1989, 42. Bernard Williams also uses some similar language. For Williams, roughly speaking, the *internal interpretation* holds that facts about an agent's reasons to act are determined by the agent's desires. Whereas, the *external interpretation* holds that facts about an agent's reasons to act hold independently of the agent's desires. Bernard Williams, "Internal and External Reasons," *Moral Luck* (New York: Cambridge University Press, 1981) 101-113. Throughout this thesis, I follow the language of Shafer-Landau as it seems most appropriate given the way I proceed.

As I assess (1), (2), and (3), I also consider the following more fundamental question: is there necessarily a reason for doing anything at all, and if so, what is that reason? As a response to this question, I endorse rational egoism, the thesis that self-interest, itself, necessarily provides agents with a reason to act. That is, there is necessarily a reason to do that which is in the agent's self-interest, and the reason is just that it is in the agent's self-interest. The view is that self-interest itself, by its very nature, provides agents with a reason to do that which is in their self-interest. Supposing it is in an agent's self-interest to drink green tea, the rational egoist would hold that this fact, itself, provides the agent with a reason to drink green tea.

I will make a case for rational egoism, and the corresponding version of extrinsic moral rationalism, the version which holds that there is necessarily a reason to do the moral thing because there is necessarily a reason to do that which is in the agent's self-interest, and there is a necessary connection between morality and self-interest. Of course this endeavor must include both a defense of rational egoism, and a defense of the necessary connection between morality and self-interest. Although I will make some points in favor of the necessary connection between morality and self-interest in chapter section 3.7 of this thesis, I will focus primarily on making a case for rational egoism. I will only be able to make a preliminary case for the sort of extrinsic moral rationalism I want to defend, but I consider this to be solid progress, because I think just getting this theory on the map is a significant endeavor.

Here is how I will proceed. In chapter 1, I will make a preliminary argument against moral antirationalism that appeals to the way agents conventionally think and talk about morality, but then I will argue against intrinsic moral rationalism. This will amount to a preliminary case in favor of some version of extrinsic moral rationalism. If readers find themselves on board with the case I make for moral rationalism in general and, at the same time,

on board with the case I make against intrinsic moral rationalism, then they will likely feel compelled to accept some version of extrinsic moral rationalism.

In chapter 2, I argue against the desire-satisfaction view, which is the theory that, roughly, all of our reasons to act are grounded by our desires. In this way, I hope to rule out any versions of extrinsic moral rationalism, and moral antirationalism for that matter, that are connected with the desire-satisfaction view. Since the most plausible versions of extrinsic moral rationalism are most likely the versions that connect moral rationalism with either the desire-satisfaction view or rational egoism, and we have a case in favor of extrinsic moral rationalism but a case against the desire-satisfaction view, we have a case in favor of the version of extrinsic moral rationalism that is connected with rational egoism. This is just a preliminary case, though.

In chapter 3, I develop a version of rational egoism and I make the case that, at best, my version of rational egoism, as a theory, fares better than the other candidate theories, namely intrinsic moral rationalism and the desire-satisfaction view, but at worst, fares just as well. The conclusion I reach is that there are grounds for believing that rational egoism is true, and grounds, though perhaps less solid at this point, for believing that the corresponding version of extrinsic moral rationalism is true. At the very least, I am confident that I can show that rational egoism and the corresponding version of extrinsic moral rationalism are as deserving of consideration as the other theories on the map.

## Chapter 1

### Against Intrinsic Moral Rationalism

A main question that I consider in this thesis goes as follows: is there necessarily a reason to do the moral thing, and if so, what is that reason? Just for ease of reference, I will refer back to this question as *the metamoral question*.<sup>3</sup> In chapter 1, I will survey various theories that are supposed to provide answers to the metamoral question, and I will present a preliminary argument that, of these theories, moral rationalism is the most tenable. However, there are different versions of moral rationalism; I will argue that one version of moral rationalism, namely the thesis that it is morality *itself* that necessarily provides agents with reasons to act, is untenable. So, all in all, I will be suggesting that we ought to accept another version of moral rationalism: the version which holds that morality necessarily provides agents with reasons to act in virtue of a necessary connection between morality and some other intrinsically reason-giving sort of consideration.

I will proceed as follows. In section 1.1, I will clarify much of the technical language that I will be using throughout this thesis, and distinguishing the question on the table from other related moral and metamoral concerns. In section 1.2, I will introduce the different qualified versions of moral rationalism, and the answers that they provide to the metamoral question. I will also introduce the opposing view, moral antirationalism, and the answer that it provides to the metamoral question. In section 1.3, I will make a preliminary argument in favor of moral

---

<sup>3</sup>I should explain briefly that it is a metamoral question rather than a normative moral question because, generally, normative morality is about what morality requires that agents do, and so questions about the relationship between morality and rationality would fall under the heading of metamorality, rather than normative morality. Also, I should say that I do not mean to suggest this is the only metamoral question. Indeed, there are many other metamoral questions, for example, the sorts of questions that have to do with the metaphysics, epistemology, and semantics of morality. I just choose to call this the *metamoral question* so that I can easily refer back to it throughout the thesis.

rationalism that appeals to the way agents conventionally think and talk about morality. In sections 1.4 and 1.5, I will argue that although we have a preliminary case in favor of moral rationalism, there are grounds for rejecting the thesis that it is morality *itself* that necessarily provides agents with reasons to act.

### 1.1 Introducing Moral Rationalism, and Distinguishing Our Main Concern from Related Moral and Metamoral Concerns

Moral rationalism holds that morality necessarily provides agents with a reason to do the moral thing.<sup>4</sup> As formulated, moral rationalism leaves open whether it is morality itself that intrinsically provides agents with a reason to do the moral thing, or it is in virtue of a necessary connection between morality and some other intrinsically reason-giving sort of consideration that morality necessarily provides agents with a reason to do the moral thing. The general, unqualified thesis of moral rationalism serves as an answer to the first part of the metamoral question; moral rationalism holds that there is indeed necessarily a reason to do the moral thing. Supposing it is a fact that lying is morally wrong, the moral rationalist would hold that this fact provides agents with a reason to refrain from lying. Or supposing it is a fact that keeping promises is morally good or valuable, the moral rationalist would hold that this fact provides agents with a reason to keep their promises. The general, unqualified thesis of moral rationalism,

---

<sup>4</sup>I follow Shafer-Landau in my use of the term *moral rationalism*. Shafer-Landau, 2003. Later I will distinguish between two versions of moral rationalism, *intrinsic moral rationalism* and *extrinsic moral rationalism*, and I will follow Shafer-Landau in my use of these terms as well. My language should not be confused with the language of internalism/externalism which might be more familiar to some readers. Brink, for example, uses the term *internalism about reasons* for the thesis that morality itself necessarily provides agents with reasons to act. Brink, 42. I am choosing to follow Shafer-Landau's terminology as it is more appropriate given how I proceed in this thesis. For more details about how other writers have used similar terms in different ways, readers might find it helpful to refer back to footnote 2.

as formulated, does not serve as an answer to the second part of the metamoral question, namely: *what* reason is there to do the moral thing?

Several points of clarification are in order before I move forward with a discussion of the different qualified versions of moral rationalism and how they provide answers to the second part of the metamoral question. Moral rationalism holds that morality *necessarily* provides agents with a reason to do the morally required thing; the first point of clarification has to do with the sort of necessity that I will be discussing throughout this thesis. The sort of necessity I will be discussing is not metaphysical necessity; I will be following a more restricted sense of necessity. In my understanding, a relationship of metaphysical necessity holds in all metaphysically possible worlds. So, if it were true that morality necessarily, in the metaphysical sense, provides agents with a reason to do the moral required action, then it would follow that morality always provides agents with a reason to do the morally required action in all metaphysically possible worlds. I do not use this sense of necessity because there could conceivably be possible worlds whose laws of nature are so different from those of the actual world that different moral truths would hold, and different truths about the relationship between morality and rationality would hold. Another sort of necessity that I choose not to follow is conceptual necessity, and here I diverge from Brink.<sup>5</sup> In my understanding, conceptual necessity, roughly, is the sort of necessity which holds in virtue of some concept. So if it were true that morality necessarily, in a conceptual sense, provides agents with a reason to do the morally required thing, then agents always have a reason to do the morally required action in virtue of the very *concept* of morality. Since it is the concept of morality, then, that determines its reason-giving power, the reason-giving power of morality cannot depend on any other substantive considerations such as what the content of morality turns out to be, or what is the correct theory of practical reasons. If I were to

---

<sup>5</sup>Brink, 42.

use the conceptual sense of necessity, this would possibly beg the question against, or quickly rule out, extrinsic moral rationalism. Extrinsic moral rationalism holds that morality always provides agents with reasons to act, not in virtue of the concept of morality, but in virtue of a necessary connection between morality and some other intrinsically and necessarily reason giving sort of consideration. Since I want to leave room for extrinsic moral rationalism, this is why I choose to diverge from Brink, and not keep with the conceptual sense of necessity. Throughout the rest of this thesis, I will be using *necessity* in the following restricted sense: a relationship of necessity holds in the actual world and all *relevantly close* possible worlds. A possible world is relevantly close to the actual world just in case its laws of nature are the same as those of the actual world, and the same relevant facts obtain about agents, their desires, and their interests. Continuing with this sense of necessity, moral rationalism holds that morality always provides agents with reasons to do the morally required action in the actual world and all relevantly close possible worlds. One benefit of keeping with this weaker, more restricted sense of necessity is that it does not rule out metaphysical or conceptual necessity. In other words if something is necessarily the case, using the restricted sense of necessity I have chosen to follow, it could also turn out to be necessarily the case, in a metaphysical, or conceptual sense.

The second point of clarification is that moral rationalism is neutral with respect to normative moral theory.<sup>6</sup> Normative moral theory has to do with the content of morality, in contrast with metamoral theory which has to do with the ontological, epistemological, and semantic status of morality, and how morality relates to rationality. The content of morality includes claims about what agents should do, ought to do, or more specifically what morality

---

<sup>6</sup>Although moral rationalism as such is neutral with respect to normative moral theory, I believe that moral rationalism is most defensible, all things considered, when paired with particular normative moral theories. I will have more to say about this in chapter section 3.7.

requires, or recommends, that agents do. Moral rationalism, as such, is neutral with respect to all claims about what morality *specifically* requires. Moral rationalism just holds that agents necessarily have a reason to do what morality requires, whatever it turns out that morality requires.

The third point of clarification has to do with the sort of reasons that I am concerned with in this thesis, and the distinction between the sort of reasons I am concerned with, and another sort of reasons. In this thesis, I am primarily concerned with good, or *justifying reasons*, and I define *justifying reasons* as genuine considerations in favor of doing something or other. Justifying reasons, unsurprisingly, are the sort of reasons that can justify an agent's taking some action, and importantly, justifying reasons are something different from explanatory reasons. *Explanatory reasons* are considerations, provided by the elements in an agent's subjective motivational set, that can serve in an explanation as to why an agent takes some action or another. I follow Williams in my understanding of an agent's subjective motivational set, and what elements an agent's subject motivational set may include.<sup>7</sup> As possible elements of an agent's subjective motivational set, Williams mentions dispositions of evaluation, patterns of emotional reaction, personal loyalties, projects, and commitments, but does not purport to give an exhaustive account of all the elements that may be included in an agent's subjective motivational set.<sup>8</sup> Hereafter, for ease of reference, I will use the term *desire* to refer to all elements of an agent's subjective motivational set, granting that the elements of an agent's subjective motivational set may include desires among a number of other things.

---

<sup>7</sup>Williams, "Internal and External Reasons," 101-113.

<sup>8</sup>Ibid., 105.

So, if agent S took some action A because S had some desire D, then S's having D would count as an explanatory reason for why S did A; indeed, S's having the desire D could serve in an explanation for why S did A.<sup>9</sup> For example, if Chuck desires to eat light bulbs and then eats some light bulbs, his having the desire to eat light bulbs would count as an explanatory reason for why he ate some light bulbs; indeed, his having the desire to eat light bulbs could serve in an explanation as to why he ate some light bulbs.<sup>10</sup> Although importantly, it is another matter whether or not Chuck has a justifying reason for eating light bulbs. In other words, although we have an explanation for Chuck's light-bulb-eating, it is another matter whether or not Chuck is justified in eating light bulbs. Presumably, Chuck would have no justifying reason to eat light bulbs, but just what sort of consideration it is that provides agents with justifying reasons to act is the primary problem of this thesis, so we are not warranted here in affirming or denying whether or not Chuck has a justifying reason to eat light bulbs. The verdict on this matter depends on the relationship between desires and reasons to act. If an agent's desires provide him with a justifying reason to fulfill them, then Chuck would have a justifying reason to eat light bulbs. But I take it that, conventionally, we would say that Chuck would not have a justifying reason to eat light bulbs; that is, we would tend to think that Chuck would not be justified in eating light bulbs. Although his desire to do so could serve in an explanation as to why he did.

At this point, it is appropriate to say something about the term *motive* as it may be familiar to some, and is not to be confused with the term, *justifying reason*. Indeed, justifying reasons, and not motives, are what I am primarily concerned about in this thesis.<sup>11</sup> But I will

---

<sup>9</sup>Note that the desire, itself, is not the explanatory reason why S did A. S's having the desire is the explanatory reason why S did A.

<sup>10</sup>This example is inspired by Brink, 39.

<sup>11</sup>Correspondingly, I am not concerned about the debate between motivational internalism and motivational externalism. Motivational internalism is the thesis that morality necessarily provides agents with motives to do the

make some attempt to provide an account of what motives are, in order to distinguish them from justifying reasons. I think that most agents, or perhaps normal agents, feel compelled to act in accordance with their desires, and recall that I use the term *desire* to refer to all elements of an agent's subjective motivational set keeping in mind that the an agent's subjective motivational set may include desires among a number of other things, like beliefs, for instance.<sup>12</sup> So, when I say I think that most agents, or normal agents, feel compelled to act in accordance with their desires, what I mean is that I think that most agents, or normal agents, feel compelled to take actions that will serve as a means to the fulfillment of their desires. But, as an important point of clarification, I am not committed to the claim that agents can *only* feel compelled to do that which they desire, in the conventional sense of the word desire. In other words, I do not believe that agents can only feel compelled to do what they want to do, or to do what will serve as a means to meeting their wants. I believe that agents can feel compelled to act in accordance with any of the elements of their subjective motivational set. For example, an agent may not *want* to take some action A, and yet the agent may nonetheless believe that he should do A, and then he may feel compelled to do A, in accordance with his belief that he should do A. I think this *feeling* is what is meant by the term *motive*. So I define *motive* roughly as follows: the feeling, or compulsion, that agents have to take actions in accordance with the elements of their subjective motivational set. For example, if an agent has a desire to torture cats, it is likely that he would have a motive to torture them, for I assume that most agents feel compelled to take actions that will serve as means to the fulfillment of their desires. Similarly, an agent may have

---

moral thing, and motivation externalism is the thesis that morality does not necessarily provide agents with motives to do the moral thing. I think motivational externalism is true, but that is not what I am concerned about in this paper.

<sup>12</sup> By my lights, I am keeping with Williams here, in holding that an agents' beliefs count as elements of the agent's subjective motivational set. Williams, "Internal and External Reasons," 101-113.

the desire to torture cats, and yet believe that he should not torture cats, and then from his belief that he should not torture cats, he could become motivated to not torture cats.

Interestingly, a motive may provide an agent with an explanatory reason to act; indeed, an agent's having a motive to take some action could serve in an explanation as to why the agent took that action. It is another matter whether or not motives provide agents with justifying reasons to act. Returning to our light-bulb-eater now, although he may be motivated to eat light bulbs, presumably he would not be justified in doing so.

The fourth point of clarification is just that moral rationalists are not necessarily committed to the stronger thesis that morality necessarily provides agents with *absolute* justifying reasons to act; moral rationalists are just committed to the thesis that morality necessarily provides agents with *prima facie* justifying reason to act. *Prima facie justifying reasons* are the sort of justifying reasons that can be overridden by additional considerations, whereas *absolute justifying reasons* cannot be overridden; *absolute justifying reasons* are the ultimate, best, all things considered, justifying reasons. To say that an agent has an absolute reason to take some action or other is to say that the overall force of rationality indicates that he ought to take that action. Unless I say otherwise, going forward, whenever I use the term *reason*, I will be referring to *prima facie* justifying reasons.

Since I am discussing *prima facie* justifying reasons, this means that the thesis of moral rationalism is that agents necessarily have *some* reason to do the morally required action; it is a separate matter whether or not agents have an absolute reason to do the morally required action. Whether or not agents have an absolute reason to do the morally required action will depend upon whether or not moral rationalism is true, and whether or not there are any other sorts of considerations that necessarily provide agents with reasons to act. And if moral rationalism is

true and there are other considerations that also necessarily provide agents with reasons to act, then an additional theory would be needed to explain how the different sorts of reason-giving considerations are to be weighed against each other. For example, if moral rationalism is true, and it also turned out to be true that agents necessarily have a reason to do that which they desire, then an additional theory would be needed to explain whether it would be morality or desires that have rational priority. That is, if an agent desired to torture cats, and yet morality forbade cat-torturing, we would need another theory that could indicate what it would be all-things-considered rational for the agent to do, or in other words, what the agent would have an absolute reason to do.

The fifth point of clarification is that, by referencing moral facts, I am not suggesting that moral rationalists are necessarily committed to the view that there are objective moral facts, that is, states of affairs referred to by true moral claims whose truth value holds independently of the subjective states of actual or hypothetical beings or groups of beings. The view that there are objective moral facts is a commitment of a theory called moral realism which holds that (1) moral claims can be either true or false, (2) at least some moral claims are true, and (3) the truth value of moral claims holds, in some sense, independently of the subjective states of actual or hypothetical beings or groups of beings. Moral rationalists, as such, can consistently deny or affirm any of the three theses of moral realism.<sup>13</sup>

This is to say that the moral rationalism debate is, to a significant extent, different from the moral realism debate. The moral realism debate has to do with the ontological, epistemological, and semantic status of morality, while the moral rationalism debate has to do with the reason-giving character of morality whatever its ontological, epistemological, and

---

<sup>13</sup>However, it might be the case that moral rationalism is most defensible when paired with a defense of moral realism.

semantic status turns out to be. I am here primarily concerned with the reason-giving character of morality. It will be useful, however, to briefly review just how moral rationalism is neutral with regard to the three theses of moral realism.

Those that affirm (1), the thesis that moral claims can be either true or false, are called cognitivists. Cognitivists hold that moral claims can be true or false; that is, moral claims are propositional. For example, the claim that “lying is wrong” can be either true or false: true if lying is, after all, wrong, and false if lying is not, after all, wrong. Cognitivist moral rationalists would likely hold that the facts referred to by true moral claims necessarily provide agents with a reason to do the moral thing. In contrast, those that deny (1) are called noncognitivists. Noncognitivists hold that there are no moral facts, and that moral claims are non-propositional. That is, moral claims are not the sort of thing that can be true or false. Rather, moral claims are just expressions of emotion or commands. For example, a noncognitivist would hold that the moral claim “lying is wrong” really just expresses the emotion “lying, boo!” or the command “don’t lie.” Up until now, I have referred to moral *facts* as reason-giving, but of course, noncognitivists do not believe that there are moral facts; for noncognitivists, morality is just a system of expressions of emotion or commands. So noncognitivist moral rationalists would likely believe that moral claims, which are really just expressions of emotion or commands, necessarily provide agents with reasons to act.

Those that affirm (2), the thesis that at least some moral claims are true, are called success theorists, whereas those that deny (2) are called error theorists. Both success theorists and error theorists grant (1). Error theorists just hold that, although (1) is true, there are no moral facts and so all moral claims are false. Moral rationalist success theorists would likely hold that the facts expressed by true moral claims necessarily provide agents with a reason to do the moral

thing. Moral rationalist error theorists would likely hold that moral claims (although all false) necessarily provide agents with a reason to do the moral thing.<sup>14</sup>

Those that affirm (3), the thesis that the truth value of moral claims holds, in some sense, independently of the subjective states of actual or hypothetical beings or groups of beings, are called objectivists, whereas those that deny (3) are called subjectivists. Both objectivists and subjectivists grant (1) and (2). Subjectivists just hold that the truth value of moral claims depends, in some sense, on the subjective states of actual or hypothetical beings or groups of beings. Both objectivist and subjectivist moral rationalists would likely hold that the facts expressed by true moral claims necessarily provide agents with reasons to act. Objectivists and subjectivists just disagree about what determines the truth value of a moral claim, or what constitutes a moral fact. Objectivists believe that moral facts hold independently of the subjective states of actual or hypothetical beings or groups of beings, while subjectivists believe that moral facts are determined by the subjective states of actual or hypothetical beings or groups of beings.

---

<sup>14</sup>Whether or not moral rationalist error theory is defensible is another matter. The moral rationalist error theorist would likely hold that, although no moral claims are true, moral claims nonetheless necessarily provide agents with reasons to act. This view seems difficult to defend for it is difficult to see how a false claim could necessarily provide an agent with a reason to act. I will illustrate this point with an example. I am choosing a non-moral example just because I think it is most helpful here. This example is largely inspired by Williams, "Internal and External Reasons," 102. Suppose Jim wants a gin and tonic, and just suppose that, in this case, Jim has a reason to do what he wants. On the counter in front of him is a bottle of petrol. Smith walks in and claims that the bottle of petrol is actually a bottle of gin. Jim believes Smith's claim. Still, it really does not seem like Jim has a reason to mix the petrol with tonic and ice, and drink it. This is because Jim wants a gin and tonic, not a petrol and tonic. So, Smith's false claim does not seem to provide Jim with a reason to mix the petrol with tonic and ice, and drink it, even though Jim believes Smith's claim. In contrast, suppose that the bottle is actually filled with gin, and Smith claims that the bottle is filled with gin. It seems like this true claim would provide Jim with a reason to mix what is in the bottle with tonic and ice, and drink it. This is because Jim wants a gin and tonic and we have supposed that Jim has a reason to do what he wants. So the conclusion I reach is that a moral rationalist could consistently be an error theorist, but that this view would be difficult to defend since it is difficult to see how a false claim could provide an agent with a reason to act. The claim that (2) is false, then, could serve as an objection to moral rationalism. But on the other hand, perhaps a moral rationalist error theorist could give a plausible imaginative story about how a false claim could provide an agent with a reason to act. Error theorists must, after all, be imaginative, for they hold that conventional morality is in error, and so they must provide some imaginative story about how we came to make this error, and what morality really *is*, if it is anything at all.

By rejecting (1), (2), or (3), every moral antirealist denies that there are objective moral facts in some way or another. Then they provide some revisionary account to explain what realists think are objective moral facts *really* are. The noncognitivist revision goes as follows: morality is really just a system of expressions of emotion or commands. So what a realist believes are true claims referring to objective moral facts are really just expressions of emotions or commands. The error theorist revision goes as follows: there are no moral facts and thus no true moral claims. So what a realist believes are true moral claims referring to objective moral facts are really all false. The subjectivist revision goes as follows: there are moral facts and thus true moral claims, but the truth value of moral claims is determined by the subjective states of actual or hypothetical beings or groups of beings. So what a realist believes are objective moral facts are really subjective moral facts. The bottom line here is that realists and antirealists have drastically different conceptions of what is going on with morality in general, but realists and all the different kinds of antirealists can still be moral rationalists just in case they believe that morality, whatever it *really* is, necessarily provides agents with a reason to act.

## 1.2 Moral Rationalism as a Solution to the Metamoral Question

These points being made, I will now move forward with a discussion of the different qualified versions of moral rationalism and the answers they provide to the metamoral question: is there necessarily a reason to do the moral thing, and if so, what is that reason? There are two different qualified versions of moral rationalism: intrinsic moral rationalism and extrinsic moral rationalism. Both provide answers to the second part of the metamoral question, but, in different ways. Intrinsic moral rationalism holds that morality *itself* provides agents with reasons to act. That is, there is necessarily a reason to do the moral thing, and that reason is just that it is the

moral thing. The view is that morality itself, by its very nature, provides agents with a reason to do the moral thing. Still supposing that lying is morally wrong, the intrinsic moral rationalist would hold that this fact, itself, provides agents with a reason to refrain from lying. And still supposing that keeping promises is morally good, the intrinsic moral rationalist would hold that this fact, itself, provides agents with a reason to keep their promises.

Extrinsic moral rationalism, in contrast, holds that morality necessarily provides agents with reasons to act in virtue of its necessary connection with some other intrinsically reason-giving kind of consideration such as desire or self-interest. That is, there is necessarily a reason to do the moral thing because there is necessarily a reason to do what one desires or what is in one's self interest, and there is a necessary connection between morality and desire or self-interest. Extrinsic moral rationalists can be classified according to which other consideration (desire or self-interest) they believe intrinsically provides agents with a reason to act.

The point of divergence between intrinsic and extrinsic moral rationalism concerns just what it is that intrinsically provides agents with reasons to act. Correspondingly, intrinsic moral rationalists and extrinsic moral rationalists give different answers to the second part of the metamoral question. Intrinsic moral rationalism holds that morality intrinsically provides agents with reasons to act, whereas extrinsic moral rationalism holds that some other consideration intrinsically provides agents with reasons to act. Correspondingly, intrinsic moral rationalism holds that the reason to do the moral thing is just that it is the moral thing, whereas extrinsic moral rationalism holds that the reason to do the moral thing is that it is necessarily in one's self-interest to do the moral thing, or it fulfills one's desires to do the moral thing. In that extrinsic moral rationalists believe that there is a necessary connection between morality and one of these

other respective considerations, extrinsic moral rationalists agree with intrinsic moral rationalists that morality is necessarily reason-giving.

The opposing thesis, namely, moral antirationalism holds that morality does not necessarily provide agents with a reason to do the moral thing. So the moral antirationalist answer to the metamoral question is: there is not necessarily a reason to do the moral thing. However, there might be contingent reasons for doing the moral thing.

In thinking about the various ways of answering the metamoral question, it is helpful to analyze some conventional moral phenomena. In the following section, I interpret two moral phenomena as evidence in favor of moral rationalism.<sup>15</sup>

### 1.3 A Presumptive Case in Favor of Moral Rationalism

Morality, in some way, *seems* to necessarily provide agents with a reason to do the moral thing. That is, according to the way agents conventionally think and talk about morality, it seems that moral rationalism is true. Another way of making this point is that pre-theoretical moral convention suggests that moral rationalism is true. I will support this claim by describing two conventional moral phenomena.

The first phenomenon: when an agent S is asked what justified her acting in some way, it is not uncommon, strange, or surprising for her to say something like the following: the reason I had for acting that way is that it was the moral way, or, I was justified in doing what I did because it was the morally required thing. And it would not be uncommon, strange, or surprising for other agents to accept her response. By my lights, it would seem uncommon, strange, or

---

<sup>15</sup>My argument in the following section is largely influenced by Shafer-Landau's presumptive argument for moral rationalism. Shafer-Landau, 192-193.

surprising for the other agents to ask her what reason she had to do the moral thing. Just that it was the moral thing is enough to satisfy conventional moral agents when they ask what justifies taking some action or another.

To be clear, I will describe this phenomenon with a more concrete example. Suppose Jones is on his way to defend his master's thesis and he sees a child drowning in a lake. He stops to rescue the child and arrives late to his thesis defense. Also suppose that what Jones did was in fact the moral thing to do. Finally Jones arrives at his thesis defense and attempts to provide a justification for his tardiness. In addition, suppose that all the members of Jones' committee are conventional moral agents. Jones might claim something like the following, "I am late because I stopped to rescue a child drowning in the lake; this was the moral thing to do. And so I am justified in doing what I did."<sup>16</sup> I think the committee members would accept Jones' claims without further questions asked. If my example is an accurate portrayal of a conventional moral phenomenon, it seems to suggest that morality necessarily provides agents with a reason to act. This seems to be the case because the committee members do not doubt that there is a reason to do the moral thing, or ask, "but what reason did you have to do the moral thing, or, what justifies your doing the moral thing?" They seem to accept that there is necessarily a reason to do the moral thing.

I will say more about the necessity here. I have attempted to describe a conventional moral phenomenon, and I have claimed that it seems to suggest that there is *necessarily* a reason to do the moral thing. Importantly, I do not purport to have here established with certainty that conventional moral agents believe that there is necessarily a reason to do the moral thing; my claim is just that when we look at this phenomenon, conventional moral agents *seem* to accept

---

<sup>16</sup>Just to be clear, I am referring to *justifying reasons* and not *explanatory reasons* in this example. By the way, Jones would have an explanatory reason for rescuing the child and arriving late to his defense. But justifying reasons are the sort of reasons I am interested in here.

that there is necessarily a reason to do the moral thing. By my lights, that conventional moral agents would not ask Jones any further questions after he claimed that he stopped to rescue the child for the reason that it was the moral thing to do suggests that they believe that morality necessarily provide agents with reasons to act. If conventional moral agents believed that morality only contingently provided agents with reasons to act, we would probably expect them to ask more questions of Smith.

To illustrate this more clearly, let us draw an analogy between morality and something that conventional agents do not accept to be necessarily reason-giving. Humor will work. Suppose Jones shows up to his thesis defense wearing a funny hat and he says to his committee, “I am wearing this funny hat because *it is the humorous thing to do.*” It would not seem uncommon, strange, or surprising for his committee members to ask him, “but what reason did you have to do the humorous thing?” Note the contrast with morality. The committee members would ask Jones what reason he had to do the humorous thing because they do not accept that humor necessarily provides one with reasons to do the humorous thing. If humor provides agents with reasons to do the humorous thing, it would seem to be in virtue of some further necessarily reason-giving sort of consideration. The point is that conventional agents seem to accept morality to be necessarily reason-giving in contrast with something else like humor, which they seem to accept to be only contingently reason-giving.

At this point, one might object to my use of the term *conventional moral agent* on the grounds that it is overly technical, abstract, difficult to define, or that I am smuggling in just the concepts I need to support moral rationalism, or that there is no such thing as a conventional moral agent. To which I would respond by emphasizing that, by *conventional moral agent*, all I mean is what I presume to be an ordinary or normal moral actor, that is, a person in the maturity

of his faculties who, with respect to morality, behaves in a common, unsurprising way. I am here, after all, entitled to make presumptions because this is just a presumptive argument. All things considered, not much really turns on the success or failure of my presumptive argument anyway. If it is taken as a success, all I have provided is a preliminary argument in favor of moral rationalism which might serve as one mark, however small, in its favor. I would still need to provide an all things considered defense of moral rationalism. On the other hand, if my preliminary argument is not taken as a success, then the antirationalist would still need to provide an all things considered defense of moral antirationalism.

A second phenomenon that I interpret as evidence in favor of moral of rationalism is the phenomenon of moral assessment: conventional moral agents blame, condemn, or hold each other accountable for doing the immoral thing, and they praise each other for doing the moral thing. I take this as evidence in favor of moral rationalism because only moral rationalism gives a sense of fairness to the practice of moral assessment and the notions of blame, condemnation, accountability, and praise. For if moral rationalism is true, when moral agents blame each other for doing the immoral thing, they are blaming each other for doing something that they necessarily had a reason to refrain from doing. This behavior seems to merit blame. Correspondingly, when moral agents praise each other, they are praising each other for doing something that they necessarily had a reason to do, and this seems to merit praise. Essentially, agents are being assessed according to their attentiveness to reasons. This seems like a fair basis for assessment.

But to really see the way in which only moral rationalism gives a sense of fairness to moral assessment, it is useful to look at the way in which the opposing thesis fails to make sense of moral assessment. The opposing thesis is the thesis of moral antirationalism which, recall,

holds that morality does not necessarily provide agents with a reason to do the moral thing. If moral antirationalism is true, it is quite difficult, by my lights, to make sense of moral assessment and notions like blame, condemnation, accountability, and praise. Supposing moral antirationalism is true and agents have no reason to do the moral thing, then when moral agents blame each other for doing the immoral thing, they are blaming each other for doing something that they did not have a reason to refrain from doing, and this behavior does not seem to merit blame. Correspondingly, when moral agents praise each other, they are praising each other for doing something that they did not necessarily have a reason to do, and this behavior does not seem to merit praise. Agents are not being assessed according to their attentiveness to reasons.

The bottom line here is that it is difficult to understand what could count as a fair basis for moral assessment other than attentiveness to reasons. Moral antirationalists might claim that desires, rather than reasons, are what counts as the basis of moral assessment. That is, one is moral if one's desires are.<sup>17</sup> But it is still difficult to understand how desires could count as a fair basis of moral assessment if there is not necessarily a reason to cultivate moral desires.

I should make it clear that I am assuming that moral assessment should be understood as a fair practice. I make this assumption because most conventional moral theories count fairness as a moral virtue. This assumption is warranted because all I am doing here is making a presumptive argument with appeal to our conventional moral thought and practice. Conventionally speaking, moral assessment is a fair thing. It would be unconventional or revisionary to make moral assessment out to be an unfair thing.

---

<sup>17</sup>This is Foot's view in Philippa Foot, "Morality as a System of Hypothetical Imperatives," *The Philosophical Review* 81 (1972): 305-316.

Alas, this concludes my preliminary argument in favor of moral rationalism. I just described (hopefully accurately) two moral phenomena which I interpret as evidence in favor of moral rationalism.

#### 1.4 A Strike Against Intrinsic Moral Rationalism

So far I have raised an important metamoral question, introduced some different ways of answering it, and offered a preliminary argument in favor of moral rationalism unqualified. But as I discussed in section 2, there are different qualified versions of moral rationalism: intrinsic moral rationalism and extrinsic moral rationalism. My preliminary argument in favor of moral rationalism was neutral with respect to any disagreement between intrinsic moral rationalism and extrinsic moral rationalism. But now I will argue that there is a crucial problem with intrinsic moral rationalism, namely that it includes a problematic element of mystery, and since extrinsic moral rationalism escapes this problem, it is the preferable candidate theory. In this section, I will discuss this crucial problem with intrinsic moral rationalism by way of a famous argument against intrinsic moral rationalism that has been made by Phillipa Foot.<sup>18</sup>

To make sense of Foot's argument, I need to begin by explaining her distinction between categorical and hypothetical imperatives. Foot actually uses the terms categorical and hypothetical imperatives in two different senses.<sup>19</sup> In what follows, I am concerned particularly with these terms in just one of the senses she uses. In the sense I am concerned with, what Foot

---

<sup>18</sup> Ibid.

<sup>19</sup> In one sense, a *categorical imperative* is a moral prescription that necessarily provides agents with a reason to act, while a *hypothetical imperative* is a moral prescription that only contingently provides agents with a reason to act. In another sense, a *categorical imperative* is a moral prescription that has an all pervasive jurisdiction. That means that it necessarily applies to all agents even though it might not necessarily provide them with a reason to act. In this second sense, a *hypothetical imperative* is a moral prescription that has a limited jurisdiction. That means that it applies to agents only in certain contexts, and does not necessarily provide them with a reason to act.

calls an imperative is very close to what I call a (prima facie, justifying) reason. I interpret the distinction that Foot makes between categorical and hypothetical imperatives as parallel to the distinction I make between intrinsic and extrinsic reasons which goes as follows. An intrinsic reason is a consideration in favor of doing something or other in virtue of no further consideration, whereas an extrinsic reason is a consideration in favor of doing something or other in virtue of a further consideration. Categorical imperatives are to intrinsic reasons as hypothetical imperatives are to extrinsic reasons. Moving forward with the parallel, I will primarily use the language of intrinsic and extrinsic reasons.

Like me, Foot holds that moral convention suggests that moral rationalism is true. Only she believes that it is particularly intrinsic moral rationalism that moral convention suggests to be true, for, by her lights, conventional morality suggests that there is an intrinsic reason (a categorical imperative) to do the moral thing. But, on the contrary, she thinks that there must only be an extrinsic reason (a hypothetical imperative) to do the moral thing, and she makes her case by drawing an analogy between morality and etiquette. Roughly, the analogy is that, both with etiquette and morality, at first glance, there might seem to be an intrinsic reason to do what each requires. At second glance, however, it is most mysterious how there could be an intrinsic reason to do the polite thing, and analogously, most mysterious how there could be an intrinsic reason to do the moral thing. She concludes, then, that there must only be extrinsic reasons for doing the polite thing, and analogously, only extrinsic reasons for doing the moral thing. So, by her lights (and mine), intrinsic moral rationalism must be false.

I will say a little more about her argument against intrinsic moral rationalism, attempting to make it a little more clear here with an example. Supposing Smith is at dinner at a fancy restaurant, it might seem that Smith has an intrinsic reason to practice good etiquette and refrain

from eating his steak with his hands even if he really desires to eat his steak with his hands. There is some sense in which, at least while he is at dinner, Smith is under the jurisdiction of etiquette, and he might have a reason to practice it regardless of whether he wants to or not. And so it is with morality. Only the jurisdiction of morality is supposedly larger—even inescapable; it seems that agents always necessarily have an intrinsic reason to do the moral thing. Yet there are agents who do not care about etiquette or morality, and it is just plain mysterious according to Foot, how these agents necessarily have an intrinsic reason to do the polite thing, or the moral thing. So, at second glance, she concludes that there must be only extrinsic reasons to practice etiquette and, analogously, only extrinsic reasons to do the moral thing. Smith, after all, has no mysterious intrinsic reason to practice etiquette and refrain from eating his steak with his hands. Similarly, the agent who cares nothing about morality has no intrinsic reason to do the moral thing.

I think Foot succeeds at bringing out the element of mystery in intrinsic moral rationalism by drawing the analogy between morality and etiquette. In the case of etiquette, at second glance, it becomes clear that an agent only has a reason to do what etiquette requires in virtue of *some* further consideration, for example, that he likes etiquette and desires to do what it requires, or that he does not desire to face the consequences of breaking the rules of etiquette, or that it is in his self-interest to follow the rules of etiquette. The view that etiquette intrinsically provides agents with reasons to act is absurdly mysterious; this view grants *etiquette* some strange, magical force that allows it to impose upon agents and somehow give them reasons to act. Intrinsic moral rationalism is analogously, absurdly mysterious; intrinsic moral rationalism just holds that morality has this strange, magical force. So I agree with Foot that the conclusion we

should draw is that moral judgments have no better claim to be categorical imperatives [or, provide agents with intrinsic reasons to act] than do statements about matters of etiquette.<sup>20</sup>

Elaborating on Foot now, the mysteriousness of intrinsic moral rationalism is furthermore brought out when we think about how unsatisfactorily the intrinsic moral rationalist would answer questions like the following:

- (1) How does morality necessarily provide agents with a reason to act?
- (2) Why do agents necessarily have a reason to do the moral thing?

The intrinsic moral rationalist answer to (1) must be something like the following: morality necessarily provides agents with a reason to act in that morality is just the sort of thing that, by its very nature, necessarily provides agents with a reason to act. And the intrinsic moral rationalist answer to (2) must be something like the following: agents necessarily have a reason to do the moral thing because morality is itself necessarily reason-giving. There seems to be at least one thought missing from the intrinsic moral rationalist's account, namely the further thought that is needed to satisfactorily answer (1) and (2), that is, a thought that does more than restate the thesis of intrinsic moral rationalism. The intrinsic moral rationalist's account is mysterious; it is mysterious just *how* morality provides agents with a reason to act and *why* agents necessarily have a reason to do the moral thing.

In contrast, extrinsic moral rationalism includes the further thought that is needed to satisfactorily answer (1) and (2). To (1), the extrinsic moral rationalist would say: morality necessarily provides agents with a reason to act in that morality is necessarily connected with some other consideration (whether it be desire or self-interest) that intrinsically provides agents with a reason to act. To (2), the extrinsic moral rationalist would say: agents necessarily have a

---

<sup>20</sup>Ibid., 312.

reason to do the moral thing because doing the moral thing is necessarily connected with some other consideration (whether it be desire or self-interest) and there is an intrinsic reason to do what one desires, or what is in one's self-interest. The conclusion I reach is that because intrinsic moral rationalism includes a problematic element of mystery, and since extrinsic moral rationalism appears to escape this problem, it is the preferable candidate theory. In the following section, I will further defend this conclusion against two intrinsic moral rationalist replies.

### 1.5 Two Intrinsic Moral Rationalist Replies

The first intrinsic moral rationalist reply is that Foot's point that intrinsic moral rationalism is mysterious poses a threat to extrinsic moral rationalism as well; the worry just surfaces at the second level. The worry seems to be that it is just mysterious how any consideration can be intrinsically reason-giving; just as intrinsic moral rationalism has to account for the intrinsically reason-giving character of morality, extrinsic moral rationalism has to account for the intrinsically reason-giving character of desire, or self-interest. I call this *the worry with intrinsic reasons*.

Even more interesting is that the worry with intrinsic reasons poses a potential threat to Foot's own account of intrinsic reasons, as she holds that an agent's ends provide the agent with intrinsic reasons to act. I derive her account of reasons primarily from the following quotation: "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."<sup>21</sup> From this it follows that rational actions are those in which an agent acts in accordance with his own purposes or ends. So in my interpretation, Foot is committed to the view that there is an intrinsic reason to act in accordance

---

<sup>21</sup>Ibid., 310.

with one's own ends. Just as the intrinsic moral rationalist has to account for the intrinsically reason-giving character of morality, Foot has to account for the intrinsically reason-giving character of an agent's ends.

This reply is supposed to help the intrinsic moral rationalist because, if it turns out that any and all accounts of intrinsic reasons include a problematic element of mystery, then intrinsic moral rationalism is no worse off as a candidate theory. My point was that *only* intrinsic moral rationalism is plagued by this problematic element of mystery and that other accounts of intrinsic reasons, particularly versions of extrinsic moral rationalism, are not. So, if the intrinsic moral rationalist can show that other accounts of intrinsic reasons also include a problematic element of mystery, my point would no longer work against intrinsic moral rationalism.

I will argue that my point does work because the intrinsic reason-giving character of morality is *uniquely* mysterious; the intrinsic moral rationalist cannot and does not genuinely try to account for the intrinsic reason-giving character of morality. In contrast, it is much easier for Foot or an extrinsic moral rationalist to account for the intrinsic reason-giving character of whatever it is that they claim is intrinsically reason-giving. All the other candidate options namely ends, desires, and self-interest are considerations that are quite intuitive candidate sources for intrinsic reasons. That is, it is quite natural to think that things like an agent's ends, desires, and self-interest are the sort of thing that intrinsically provide an agent with reasons to act. These things are not foreign, alien, or external things like morality and etiquette. No strange or mysterious force is needed to explain how an agent's ends, desires, or self-interest could provide the agent with intrinsic reasons to act. Thus there is no mystery in the view that an agent's ends, desires, or self-interest necessarily provide the agent with intrinsic reasons to act. In contrast, morality, like etiquette, is foreign, alien, and external to the agent. If intrinsic moral

rationalism was true, morality would need some strange force behind it that could allow it to impose upon agents and give them reasons to act regardless of their ends, desires, or self-interest. So I conclude that intrinsic moral rationalism is *uniquely* mysterious, that is, mysterious in a way that other accounts of intrinsic reasons are not.

The second intrinsic moral rationalist reply is that the element of mystery in intrinsic moral rationalism is not a fatal flaw; the claim is that we have no choice but to embrace the mystery.<sup>22</sup> Shafer-Landau defends this claim by bringing up a view which I call *epistemic rationalism*, which holds that certain non-moral facts such as  $2+2=4$  are themselves intrinsically reason-giving. According to Shafer-Landau, just as the non-moral fact that  $2+2=4$  is itself a reason to believe that  $2+2=4$ , so the moral fact that A is morally wrong is itself a reason not to do A. Non-moral facts such as  $2+2=4$  share the same mysteriously, intrinsically reason-giving character as moral facts, and yet the mystery does not stop people from accepting epistemic rationalism. The mysteriousness of the intrinsically reason-giving character of non-moral facts does not appear to be an impediment to accepting epistemic rationalism. Similarly, the mysteriousness of the intrinsically reason-giving character of moral facts should not be an impediment to accepting moral rationalism.

But I do not think Shafer-Landau's point here really helps intrinsic moral rationalism. By my lights, his point is that intrinsic moral rationalism does not look so bad, that is, does not look so *problematically* mysterious, when we consider its parallel to the view I am calling epistemic rationalism. But Shafer-Landau's point only works if it is accepted that epistemic rationalism is not problematically mysterious. I do not accept epistemic rationalism because I think it is just as mysterious as intrinsic moral rationalism, and just as problematically mysterious. Here is why.

---

<sup>22</sup>Shafer-Landau, 205.

First, if epistemic rationalism were true, facts like  $2+2=4$  would need some strange force behind them that could allow them to impose upon agents and give them reasons to act regardless of their ends, desires, or self-interest. It is much more plausible that agents have a reason to believe facts like  $2+2=4$  in virtue of some further consideration like they desire to believe what is true, or it is in their self-interest to believe what is true. Second, the epistemic rationalist, paralleling the intrinsic moral rationalist, has no answer to the following questions that do not merely restate the thesis of epistemic rationalism: (1a) how do non-moral facts such as  $2+2=4$  necessarily provide agents with a reason to believe them? and (2a) why do agents necessarily have a reason to believe non-moral facts such as  $2+2=4$ ? So this intrinsic moral rationalist reply does not work either. We are not forced to accept the mysteriousness involved in the view that facts, whether moral or non-moral, somehow impose upon agents and provide them with reasons to act, as the intrinsic moral rationalist would have us think. Rather, we may hold the more plausible view that things like an agent's desires, or self-interest are what provide agents with intrinsic reasons to act.

## 1.6 Conclusion

In chapter 1, I raised an important metamoral question, namely whether or not there is necessarily a reason to do the moral thing, and if there is, what that reason might be. I then surveyed some different ways of answering that question. Next I made a presumptive case that the correct way of answering the question is some form of moral rationalism. Lastly, I made a case that intrinsic moral rationalism is not the correct way of answering the question. At this point then, we have a preliminary case in favor of extrinsic moral rationalism. In chapter 2, I

will argue that some versions of moral antirationalism and extrinsic moral rationalism are untenable, namely the ones that hold that the desire-satisfaction view is true.

## Chapter 2

### Against the Desire-Satisfaction View

My main project for chapter 2 is to argue that some versions of moral antirationalism and some versions of extrinsic moral rationalism are untenable. Specifically, I will argue that the desire-satisfaction view, the thesis that the agent's desires necessarily provide him with intrinsic reasons to act, is untenable. Some versions of moral antirationalism pair the desire-satisfaction view with the thesis that there is no necessary connection between morality and an agent's desires, and so there is not necessarily a reason to do the moral thing; some versions of extrinsic moral rationalism pair the desire-satisfaction view with the thesis that there is a necessary connection between morality and an agent's desires, and so there is necessarily a reason to do the moral thing. So by pointing out the flaws with the desire-satisfaction view, I will attempt to dispense with all such versions of moral antirationalism and extrinsic moral rationalism.

In what follows I will distinguish between two main versions of the desire-satisfaction view: the actual desire-satisfaction (ADS) view and the idealized desire-satisfaction (IDS) view. I will argue that both are untenable but for different reasons. In that any version of the desire-satisfaction view is bound to be plagued by either the problems of the ADS view or the IDS view, no versions of the desire-satisfaction view are tenable.

#### 2.1 The Actual Desire-Satisfaction View

The actual desire-satisfaction view (ADS) holds that an agent's actual desires necessarily provide him with an intrinsic reason to act. Foot seems to hold a version of this view although

she believes that it is particularly an agent's ends that necessarily provide him with intrinsic reasons to act.<sup>23</sup> For ease of reference, I use the term *desire* to refer to all the elements of an agent's subjective motivational set keeping in mind that the elements of an agent's subjective motivational set can include what we conventionally refer to as desires as well as a number of other things such as ends, goals, projects, wants, beliefs, or commitments.<sup>24</sup> I am not too concerned with which elements of an agent's subjective motivational set are claimed to necessarily provide agents with an intrinsic reason to act; I am, at least for now, just concerned with the more general thesis that the actual elements of an agent's subjective motivational set necessarily provide him with intrinsic reasons to act.

This thesis seems plausible at first glance. It seems natural to think that an agent's actual desires necessarily and intrinsically provide him with at least some reason to act. Desires, for example, may include desires to become a professional philosopher, to continue an exciting romance, to drink a pot of green tea, or to finish a thesis paper. Desires are, in some sense, very close to the agent, they perhaps constitute what might be called the agent's character, and in many cases, they may be what an agent lives for.<sup>25</sup> At first glance, it would seem strange to think that these things do not themselves necessarily and intrinsically provide the agent with at least some reason to act. For example, it would seem strange to think that my desire to complete my thesis does not itself necessarily provide me with at least some reason to complete my thesis. Another virtue of the desire-satisfaction view, as I attempted to point out in chapter 1, is that it does not include the element of mystery that intrinsic moral rationalism includes. And finally,

---

<sup>23</sup>Foot,310.

<sup>24</sup>Williams, "Internal and External Reasons," 101-113.

<sup>25</sup>Bernard Williams, "Persons, Character, and Morality," *Moral Luck* (New York: Cambridge University Press, 1981) 1-19.

when agents lack integrity, that is, when agents do not act in accordance with their desires, the desire-satisfaction view can allow us to blame those agents for being irrational. This might be taken as another virtue of the desire-satisfaction view, although I will have more to say about the concept of blame, as it relates to the desire-satisfaction view, below.

## 2.2 Problems for the ADS View

Despite the preliminary appeal of the ADS view, I will argue that the view is untenable because of two main problems for the view. I will provide a preliminary sketch of these problems and then consider how the ADS theorist might respond to them. As I consider the potential responses that are open to the ADS theorist, I will further explain how these problems are so devastating for the ADS view.

The first main problem for the view is that at least some agents have actual desires for things that are self-destructive, cruel, or meaningless, and it is difficult to understand how an agent could necessarily have an intrinsic reason to fulfill desires such as these. Assuming that Wilson has the desire to torture himself in all the most gruesome ways, or torture others in all the most gruesome ways, or to count all the blades of grass in his yard, the ADS view would hold that Wilson necessarily has an intrinsic reason to do these things. This is a problem for the ADS view because the ADS theorist is forced to conclude that agents like Wilson necessarily have intrinsic reasons to do the most self-destructive, cruel, or meaningless things, things that we would not conventionally think that agents have any genuine reason to do. This is an apparent flaw with the ADS view. And importantly, this amounts to a flaw for the ADS view whether or not the agent knows that his desires are for things that are self-destructive, cruel, or meaningless.

To be more clear, on one hand, the agent may actually desire to take some action A without knowing that A is self-destructive, cruel, or meaningless. If he knew that A was self-destructive, cruel, or meaningless, he may or may not desire to do it anymore. On the other hand, the agent may actually desire to do the self-destructive, cruel, or meaningless thing (knowing that it is, in fact, the self-destructive, cruel, or meaningless thing and wanting to do it in virtue of this fact). This distinction does not help the ADS theorist, who holds that the agent's actual desires necessarily provide him with intrinsic reasons to act.<sup>26</sup> Whether the agent actually desires to do A without knowing that A is self-destructive, cruel, or meaningless, or the agent actually desires to do the self-destructive, cruel, or meaningless thing (knowing that it is, in fact, the self-destructive, cruel, or meaningless thing and wanting to do it in virtue of this fact), the ADS theorist will be committed to the proposition that the agent necessarily has an intrinsic reason to do something self-destructive, cruel, or meaningless, and it is difficult to understand how an agent could necessarily have a genuine reason to do something like this.

The second main problem for the view is that, if ADS theorists are forced to conclude that agents necessarily have a reason to do the most self-destructive, cruel, or meaningless things if they desire to, then it would be unfair to blame these agents for doing such apparently blameworthy things. It would be unfair because it is unfair to blame agents for doing things that they have a reason to do. This is a problem for the ADS view because ADS theorists are forced to either unfairly blame agents for doing things that they necessarily had an intrinsic reason to do, or to (fairly) refrain from blaming agents for doing the most self-destructive, cruel, or meaningless things.

---

<sup>26</sup>The is the de re/de dicto distinction. In the de re case, the agent does not know that the action he desires to take is self-destructive, cruel, or meaningless. In the de dicto case, the agent knows that the action he desires to take is self-destructive, cruel, or meaningless, and he desires to take the action in virtue of this fact.

My conclusion is that the actual desire-satisfaction view leads to two very implausible propositions:

- (1) If agents desire to do the most self-destructive, cruel or meaningless things, then agents necessarily have an intrinsic reason to do the most self-destructive, cruel or meaningless things.
- (2) Either we ought to unfairly blame agents for doing things that they necessarily had an intrinsic reason to do, or we ought not blame agents for doing the most self-destructive, cruel, or meaningless things if they desired to do such things.

And so we have solid grounds for rejecting the ADS view.

### 2.3 ADS View Reconsidered

There are a few ways in which the ADS theorist might attempt to defend the ADS view. First, the ADS theorist might bite the bullet on (1) and then claim that it is not a problem for the ADS view. They might defend this claim by arguing that there is nothing involved in the concept of reasons or rationality that precludes agents from having reasons to do the most self-destructive, cruel, or meaningless things. The ADS theorist might claim that *reasons* and *rationality* are often, in a sense, overrated or over-estimated; that is, it is a mistake to think that reasons will always point to actions that are not self-destructive, cruel, or meaningless. Rather, there can be considerations in favor of doing the most self-destructive, cruel, or meaningless things. The idea is that agents can be rational or rational enough, that is, appropriately attentive to reasons, and yet their reasons can still indicate the most self-destructive, cruel, or meaningless

things if they desire to do such things. Essentially, this ADS reply is meant to work for the ADS theorist by challenging the intuition that there is something a matter with the view that agents can necessarily have intrinsic reasons to do the most self-destructive, cruel or meaningless things.

But I do not think this reply will work. Recall that I am talking about justifying reasons, that is, genuine considerations in favor of doing something or other. I do not see how there can be any genuine considerations in favor of doing the most self-destructive, cruel, or meaningless things *just because an agent desires to*. This is brought out when we think about how, say, Hitler—assuming now that Hitler is an ADS theorist who also, in fact, had a desire to kill the Jews—would answer the following question: what justifying reason did you have to kill the Jews? His answer would be something like the following: “what justifies my killing the Jews is that I desired to kill the Jews.” I just do not see how this can count as a genuine reason.

For those who do not share my intuition here, it might help here, once again, to bring out the distinction between *justifying* and *explanatory* reasons. In contrast with justifying reasons, explanatory reasons are considerations, provided by elements of an agent's subjective motivational set, that can serve in an explanation as to why an agent takes some action or other. Hitler would indeed have an explanatory reason for killing the Jews. That is, his having the desire to kill the Jews could serve in an explanation as to why he killed many Jews. Similarly, Wilson's having desires to do the most destructive, cruel, or meaningless things could serve in an explanation as to why he fulfilled such desires.

If explanatory reasons are the kind of reasons on the table, then the ADS view works just fine. In this case, the ADS view would just amount to the view that the elements in an agent's subjective motivational set provide considerations that can serve in explanations as to why an

agent takes some action or another. This is unproblematic. But the kind of reasons on the table are justifying reasons not explanatory reasons. This being the case, I do not see how (1) can be defended; I do not see how one could defend the claim that an agent necessarily has a justifying reason to do the most destructive, cruel, or meaningless things *just because he desires to*.

Another way that the ADS theorist might try to argue that (1) is not a problem for the ADS view is by emphasizing that the ADS view is just the view that an agent's actual desires necessarily provide him with prima facie reasons to act; it is not the view that an agent's actual desires necessarily provide him with absolute reasons to act. Recall the distinction between prima facie reasons and absolute reasons from chapter 1. A *prima facie reason* is a justifying reason that can justify an action unless there are countervailing reasons. An *absolute reason* is a justifying reason that justifies an action, all things considered. So the ADS theorist could claim that although on the ADS view we are forced to conclude that agents like Wilson necessarily have a prima facie reason to do the most self-destructive, cruel, or meaningless things if they desire to, we are not forced to conclude that agents like Wilson necessarily have an absolute reason to do the most self-destructive, cruel or meaningless things. The ADS theorist could further claim that the problem is with the proposition that agents like Wilson necessarily have an absolute reason to do the most self-destructive, cruel, or meaningless things; there is not a problem with the proposition that agents like Wilson necessarily have a prima facie reason to do the most self-destructive, cruel, or meaningless things. In that the ADS theorist is only committed to the latter claim, there is no problem for their view.

I do not think this defense will work for the ADS theorist. Granting that ADS theorists are only committed to the weaker conclusion that agents like Wilson necessarily have prima facie reasons to do the most self-destructive, cruel, or meaningless things, but not the stronger

conclusion that agents like Wilson necessarily have an absolute reason to do the most self-destructive, cruel or meaningless things, the problem still stands; the weaker conclusion is problematic enough. I do not understand how one could support the claim that one has a genuine, justifying (even though just *prima facie*) reason to do the most self-destructive, cruel, or meaningless things just because one desires to. In my view, this claim totally deflates the notion of *reasons*. A persistent ADS theorist might dig in his heels and stand by a claim like the following: I have at least a *prima facie* reason to count all the blades of grass in my yard just because I desire to. But I believe that this ADS position relies, for its argumentative force, on an equivocation on the term *reasons*. Again, I am talking about genuine, justifying reasons. I think the ADS theorist must be talking about explanatory reasons. The persistent ADS theorist, in this case, certainly has an explanatory reason to count all the blades of grass in his yard. If the agent counts all the blades of grass in his yard, his having the desire to do so would explain his action. Still, just because an agent has an explanatory reason for counting all the blades of grass in his yard, it does not follow that an agent has a genuine, justifying reason for taking such action.

We can also anticipate an ADS reply that has to do with (2) which, recall, goes as follows: either we ought to unfairly blame agents for doing things that they necessarily had an intrinsic reason to do, or we ought not blame agents for doing the most self-destructive, cruel, or meaningless things if they desired to do such things. (2) is the second problematic proposition that the ADS theorist is committed to. The ADS reply might go something like this: ADS theorists can indeed fairly blame agents like Wilson, not for their irrationality, but for having desires for things that are self-destructive, cruel, or meaningless. The ADS reply is that Wilson is adequately rational; he pays sufficient attention to reasons. He desires to do the most self-destructive, cruel, and meaningless things, and he acts in accordance with his desires; he does

what he necessarily has an intrinsic reason to do. There is nothing a matter with his rationality. Yet we can still blame him for having blameworthy desires. This reply is supposed to work for the ADS theorist because it gives him a way of blaming agents for taking conventionally blameworthy actions. However, this reply will not do, for it is difficult to understand how an agent's desires could count as a fair basis for blame if there is not necessarily a reason to cultivate non-blameworthy desires. That is, if there is no reason to have certain desires rather than others, then blaming an agent for the desires he has is a completely arbitrary endeavor. For the ADS theorist, there is no reason to have certain desires rather than others. Reasons come from the actual desires that one has. If the ADS theorist does not want to commit himself to the practice of blaming agents arbitrarily, then the other option for the ADS theorist is to endorse the practice of withholding blame even when agents commit the most self-destructive, cruel, or meaningless actions, as long as they desired to do such things. Neither of these options are plausible.

This concludes my discussion of the ADS view. I stand by my initial conclusion that the ADS view leads to two highly implausible propositions, and that this is grounds for rejecting the view.

## 2.4 Idealized Desire-Satisfaction Views

The next version of the desire-satisfaction view that I will consider is what I am calling the *idealized desire-satisfaction (IDS) view*. Actually, there are many different views that could count as IDS views. These views differ from the ADS view in that they deny, in some sense, that it is an agent's *actual* desires that necessarily provide him with an intrinsic reason to act. Rather, the IDS views all hold that the agent's desires, with some degree of qualification or idealization,

necessarily provide him with intrinsic reasons to act. The great virtue of the IDS views is that they share the preliminary plausibility of the ADS views yet are able, if they build in enough idealization, to escape the problems that I raised for the ADS view. The rest of my discussion of the desire-satisfaction views will go as follows. First, I will discuss Williams' version of the IDS view, and in particular how it is able to escape the problems that I raised for the ADS view.<sup>27</sup> Then, I will argue that this opens Williams up to another crucial problem. I will follow this up with an argument that any versions of the IDS view are bound to be vulnerable to this other crucial problem.

For ease of reference, I will hereafter refer to the troubling commitment to the implausible propositions (1) and (2) as *the ADS problem*. The crucial problem for William's view, and all versions of the IDS view, I will call *the IDS problem*. The IDS problem is that the IDS views' appeal to idealization is troublingly ad hoc.

## 2.5 Williams' IDS View

Now I will get into the details of Williams' view and how it is different from the ADS view. Williams' view includes two major qualifications to the thesis that an agent's desires necessarily provide him with an intrinsic reason to act. The first qualification is that an agent's desire *d* does not necessarily provide him with an intrinsic reason to act if the existence of *d* is dependent on a false belief. Using Williams' example now, we can imagine an agent, Jim, who actually desires to drink a gin and tonic.<sup>28</sup> Before him is a drink that looks like a gin and tonic,

---

<sup>27</sup>Williams, "Internal and External Reasons," 101-113.

but is in fact, a petrol and tonic. Jim has the false belief that the drink before him is a gin and tonic, and he *actually* desires to drink the drink before him. Yet Williams would say he does not have a reason to drink the drink before him, for his desire to drink the drink before him is dependent upon the false belief that the drink before him is a gin and tonic.

Williams' second qualification is that an agent's *d* does not necessarily provide him with an intrinsic reason to act if the agent would not still have *d* after putting it through the process of practical deliberation. Williams mentions that the process of practical deliberation includes more than just deliberation about what could work as a causal means for an agent to satisfy a desire; the process may include considerations such as what might be the most convenient, economical, or pleasant way to satisfy a desire, and the process may include a number of other considerations as well. Williams, on purpose, leaves open what could be involved in the process of practical deliberation but describes, to some extent, how it could work:

A clear example of practical reasoning is that leading to the conclusion that one has reason to  $\emptyset$  because  $\emptyset$ -ing would be the most convenient, economical, pleasant, etc. way of satisfying some element in *S* [*S* represents the agent's subjective motivational set], and this of course is controlled by other elements in *S*, if not necessarily in a very clear or determinate way. But there are much wider possibilities for deliberation, such as: thinking how the satisfaction of elements in *S* can be combined, e.g. by time-ordering; where there is some irresolvable conflict among the elements of *S*, considering which one attaches most weight to (which, importantly does not imply that there is some one commodity of which they provide varying amounts); or, again, finding constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment.<sup>29</sup>

Going back to the Jim example now, suppose Jim puts his desire to drink a gin and tonic through the process of practical deliberation. He realizes that his desire to drink a gin and tonic is dependent on his desire to relax. He desires to drink a gin and tonic just because he believes it

---

<sup>28</sup>Ibid., 102.

<sup>29</sup>Ibid., 104.

will satisfy his desire to relax. He also recalls that last time he drank a gin and tonic, he could not stop himself from drinking several more gin and tonics, and he wound up in quite a bit of trouble. Maybe drinking a gin and tonic would satisfy his desire to relax, but remembering his past troubles with gin and tonics, he loses his desire to drink a gin and tonic.<sup>30</sup> We can suppose that he gains the desire to practice meditation because he believes that practicing meditation will help him satisfy his desire to relax. In a case like this, for Williams, Jim would not have a reason to satisfy his desire to drink the gin and tonic because his desire to drink the gin and tonic would not make it through the process of practical deliberation.

Now, suppose alternatively that Jim has the desire to drink a gin and tonic, and that his desire to drink a gin and tonic is not dependent on any false beliefs. Suppose also that Jim's desire to drink a gin and tonic would survive the process of practical deliberation. In this case, on Williams' view, Jim would necessarily have an intrinsic reason to drink a gin and tonic.

## 2.6 Williams' View and the ADS Problem

Now I will discuss how William's view is able to escape the ADS problem. Recall that the ADS problem is the troubling commitment to propositions (1) and (2). Proposition (1) holds that if agents actually desire to do the most self-destructive, cruel or meaningless things, then agents necessarily have an intrinsic reason to do the most self-destructive, cruel or meaningless

---

<sup>30</sup>Notice that when Jim recalls his past trouble with gin and tonics, a relevant true belief is introduced into his rational deliberation process, namely the belief that last time he drank a gin and tonic, he could not stop himself from drinking several more gin and tonics, and he wound up in quite a bit of trouble. With the introduction of this relevant true belief, Jim loses his desire to drink a gin and tonic; his desire to drink a gin and tonic does not survive the rational deliberation process. So just as an agent's  $d$  does not necessarily provide him with an intrinsic reasons to act if  $d$  is based upon false belief, an agent's  $d$  does not necessarily provide him with an intrinsic reason to act if he would no longer desire to fulfill  $d$  given the introduction of relevant true belief.

things. First of all, Williams is not committed to (1) for Williams does not hold that all of an agent's actual desires necessarily provide the agent with an intrinsic reason to act; Williams holds that an agent's actual desire only provides the agent with an intrinsic reason to act if the agent's desire is not based upon a false belief, and would survive the practical deliberation process. However, Williams is committed to a proposition that might be just as problematic at first glance, namely (1\*): if agents desire to do the most self-destructive, cruel, or meaningless things, and their desires to do the most self-destructive, cruel, or meaningless things are not based upon false belief and would survive the practical deliberation process, then agents necessarily have an intrinsic reason to do the most self-destructive, cruel, or meaningless things. But even though Williams is committed to (1\*), this commitment is not a problem for his view for he can claim that the antecedent of (1\*) will always turn out to be false by denying the second part of the conjunct of the antecedent; he could admit that some agents actually have the most self-destructive, cruel, or meaningless desires, but he could deny that the most self-destructive, cruel, or meaningless desires could ever survive the practical deliberation process. And so he would never be forced to conclude that agents necessarily have reasons to do the most self-destructive, cruel, or meaningless things just in virtue of their desiring to do such things. This would be an easy move for Williams to make especially because his notion of the rational deliberation process is roomy enough to include whatever it might take to eliminate desires such as these:

There is an essential indeterminacy in what can be counted a rational deliberative process. Practical reasoning is a heuristic process, and an imaginative one, and there are no fixed boundaries on the continuum from rational thought to inspiration and conversion.<sup>31</sup>

Now, I should be careful because Williams might not actually hold that the most self-destructive, cruel, or meaningless desires would never survive the practical deliberation process.

---

<sup>31</sup>Ibid., 110.

But if he does not hold this, then his view would be devastated by the ADS problem. In other words, if Williams allows that the most self-destructive, cruel, or meaningless desires could survive the practical deliberation process (and also not be dependent on false beliefs), Williams would be committed to allowing that the consequent of (1\*) could obtain. Williams' way out of the ADS Problem is to deny that the consequent of (1\*) could obtain by denying that an agent's desires for things that are self-destructive, cruel, or meaningless could ever survive the practical deliberation process. So, to escape the ADS problem, Williams must be committed to the proposition that an agent's desires for self-destructive, cruel, or meaningless things could never survive the practical deliberation process.

Indeed, we can suppose that Williams would accept this proposition just in order to think about how he can escape the ADS problem. If he holds this, then he is not forced to accept the consequent of (1\*) namely that agents necessarily have intrinsic reasons to do the most self-destructive, cruel, or meaningless things, and so he is not forced to accept the other problematic conclusion that the ADS theorist is committed to, namely the proposition which holds that we ought to unfairly blame agents for doing things that they necessarily had an intrinsic reason to do, or we ought not blame agents for doing the most self-destructive, cruel, or meaningless things if they desired to do such things. Recall that the problem with (2) is the commitment to either unfair, arbitrary blaming or withholding blame even when agents commit the most blameworthy actions. Williams is not committed to (2) because he can hold that agents would never have a reason to do the most self-destructive, cruel, or meaningless things. So if agents did such things, he could *fairly* blame agents for doing things that they did not have a reason to do. This is how Williams could escape the ADS problem.

## 2.7 Williams' View and the IDS Problem

Although Williams' view is able to escape the ADS problem, it is vulnerable to another crucial problem, namely the IDS problem. The IDS problem is that the IDS views' commitment to the idealization process is troublingly ad hoc. In Williams' case, it is his appeal to the practical deliberation process that is troublingly ad hoc. In this chapter section, I will explain what makes Williams' appeal to the practical deliberation process troublingly ad hoc, and I will explain how other versions of the IDS view will be ad hoc in the same way. I will conclude that this is grounds for rejecting the IDS view.

Recall what I referred to as the great virtue of the IDS views: the IDS views share the preliminary plausibility of the ADS views yet are able to escape the problems that I raised for the ADS view. Indeed, I have discussed how Williams' IDS view is able to escape the ADS problem. Williams is able to escape the ADS problem because he can claim that agents' desires for things that are self-destructive, cruel, or meaningless would never survive the practical deliberation process. But it is Williams' appeal to the practical deliberation process that is troublingly ad hoc. To be clear, a theoretical mechanism is ad hoc if it saves the day for a theory by helping the theory to answer objections or overcome problems that are raised for it, but is itself unmotivated. Indeed, Williams' appeal to the practical deliberation process saves the day, in a sense, for Williams' view because it can allow his view to escape the ADS problem. However, Williams' appeal to the practical deliberation process is itself unmotivated.

To explain how Williams' appeal to the practical deliberation process is unmotivated, I will make an analogy. My argument in what follows is largely inspired by David Enoch.<sup>32</sup> Suppose I want to know the time. I will look at a watch. But not just any watch will do; I need a

---

<sup>32</sup>David Enoch, "Why Idealize?" *Ethics* 115 (July 2005): 762.

watch that is accurate. My father's Rolex will not do. It is only accurate when set properly and then worn regularly, and although it was set properly, no one has worn it in quite awhile. So my father's Rolex is not accurately tracking the time. I need a watch that is accurately tracking the time; fortunately, my watch is accurately tracking the time. And importantly, it seems like the facts about the time hold independently of what any watch says. What is so great about my accurate watch is that it tracks the time well, and that is why I need a look at it. But just suppose that instead of conceiving of the time-facts as holding independently of what any watch says, we conceive of time-facts just as determined by what my accurate watch says. So instead of my accurate watch *tracking* the right time, we come to believe that the right time is just determined by what my accurate watch says. But then it is difficult to see why we call my accurate watch *accurate*; there is nothing for it be doing *accurately* since we have now come to believe that the right time is just whatever my purportedly accurate watch says. It is no longer clear why I should consult my watch rather than my father's Rolex. Thus, if we conceive of time-facts as determined by what my watch says, we no longer have any special motivation for consulting my watch rather than my father's Rolex.

Now suppose we want to know what an agent has a reason to do. We could start by determining what the agent actually desires, but this would be like consulting my father's Rolex. The Rolex is not accurately tracking the time. And analogously, the agents actual desires do not seem to line up with what the agent has reasons to do. For the agent may have desires that are for things that are self-destructive, cruel, or meaningless, or he may have desires that are dependent on false beliefs, or he may have desires that he would no longer have with the introduction of relevantly true beliefs, and so on; these sorts of desires do not seem to provide agents with reasons to act. So instead, in order to find out what the agent has a reason to do,

perhaps we should look at his *idealized* desires. That is, we should look at the set of desires that are not dependent upon false belief, and that would survive the practical deliberation process, or that would be introduced during the practical deliberation process. Looking at the agent's idealized desires would be like consulting my watch, the supposedly accurate one. Just as my accurate watch tracks the time well, the agent's idealized desires would, by hypothesis, line up accurately with what the agent has reasons to do. In other words, the appeal to the practical deliberation process is supposed to help us track what it is that the agent has reasons to do just like the accurate watch is supposed to help us track the time.

But unfortunately for Williams, he holds that the agents' idealized desires determine what it is that agents necessarily have reasons to do; this view is analogous to holding that time-facts are determined by what my watch says. And now it becomes apparent that Williams' appeal to the practical deliberation process is unmotivated; for, if it is not the case that the agent's idealized desires line up with, or track, that which is reason-giving, then there is no longer any motivation for considering the agents' idealized desires. We might as well just consult the agent's actual desires. The initial motivation for considering the agent's idealized desires rather than his actual desires was that his idealized desires were thought to line up accurately with that which grounded his reasons to act. But on Williams' view, facts about what agents have reasons to do is determined by their idealized desires, and so it seems that there is no longer any special motivation, or independent grounds, for consulting agents' idealized desires rather than their actual desires. Now, perhaps Williams could come up with some other independent motivation for consulting agents' idealized desires rather than their actual desires, but the natural motivation—that idealized desires *track*, rather than determine, what agents have reasons to do—cannot be employed by Williams because he holds that facts about what agents have reasons to

do are determined by agents' idealized desires. In the following chapter section, I will consider how Williams, or other IDS theorists, could claim that there some other motivation for the IDS appeal to idealization, but for now, I am confident in concluding that Williams' appeal to idealization is troublingly ad hoc.

I have focused on how Williams' appeal to the practical deliberation process is ad hoc, or unmotivated, but it should be easy to see how any IDS appeal to idealization would be ad hoc in just the same way. Recall that the IDS views all hold that the agent's desires, with some degree of qualification or idealization, necessarily provide him with intrinsic reasons to act. It is the move from actual desires, to qualified or idealized desires, that is ad hoc. The motivation behind the move from actual desires to qualified or idealized desires is rooted in the idea that qualified or idealized desires will be more in line with what it is that necessarily provide the agent with reasons to act. But, desire-satisfaction theorists deny that there is anything further that necessarily provides agents with reasons to act; for desire-satisfaction theorists, there is nothing over and above the agents' idealized or qualified desires that is thought to provide agents with reasons to act. So there is nothing, then, for idealized or qualified desires to line up with. And so there is no reason to say that that the agents' idealized or qualified desires are preferable to the agent's actual desires. There is no longer any motivation for appealing to the agents' idealized or qualified desires. Recall that a theoretical mechanism is ad hoc if it saves the day for a theory by helping the theory to answer objections or overcome problems that are raised for it, but is itself unmotivated. Indeed, the IDS appeal to idealization saves the day, in a sense, for the IDS view because it can allow the IDS view to escape the ADS problem. However, the IDS appeal to the practical deliberation process is itself unmotivated.

## 2.8 Independent Support for Idealization

In this section, I will consider some attempts that the IDS theorist could make to argue that the IDS appeal to idealization is not ad hoc. In order to show that the IDS appeal to idealization is not ad hoc, the IDS theorist would need to show that the IDS appeal to idealization is well motivated. There are two ways I can imagine that the IDS theorist might do this.

The first attempt that the IDS theorist could make to establish that the IDS appeal to idealization is well motivated could go as follows. The IDS theorist might claim that the ADS view is well motivated, and the IDS appeal to idealization simply supplies the needed amendments to the ADS view. And so the IDS appeal to idealization is well motivated. Or maybe the IDS theorist could say that the desire-satisfaction view, in general, is well motivated, and since the IDS view is a version of the desire-satisfaction view, the motivation that the desire-satisfaction view has, in some way, carries over to the IDS view, even with its appeal to idealization. But it is appropriate to dismiss this IDS attempt quite quickly, for this attempt just points back to the IDS problem; sure, the IDS view can supply the amendments that are needed to help the desire-satisfaction view avoid the ADS problem, but what the IDS theorist must do is provide some *independent* motivation for the move to idealization. So this first attempt will not help the IDS theorist to escape the IDS problem.

The second attempt that the IDS theorist could make to establish that the IDS appeal to idealization is not ad hoc could go as follows. The IDS theorist could argue that the IDS view provides a solid explanation for the phenomenon on the table. The phenomenon on the table is that agents seem to have reasons to act, and it is just not clear what grounds agents' reasons to act. An argument that the IDS view provides the best explanation for the phenomenon on the table, if sound, would get the IDS theorist out of the IDS problem. For if the IDS view,

including its appeal to idealization, could explain very well, or better than other theories, just what it is that necessarily provides agents with reasons to act, then this would count as good motivation for the appeal to idealization. And indeed, the IDS view, at first glance, might seem like a solid explanation of what it is that necessarily provides agents with reasons to act. After all, the IDS view, with its appeal to idealization, can provide us with an account of what it is that necessarily provides agents with reasons to act, all while paying due regard to the importance of desires, but simultaneously, capturing our intuition that desires to do self-destructive, cruel, or meaningless things do not necessarily provide agents with intrinsic reasons to act; in other words, the IDS view provides the explanation that agents' desires—but not their faulty ones, only their well-considered ones—necessarily provide agents with reasons to act. So, it might seem that the IDS view provides a decent explanation of the phenomenon on the table.

But I do not think that the IDS view provides a very good explanation of the phenomenon on the table. And to show that the IDS appeal to idealization is well motivated, the IDS theorist would need to make a strong case that the IDS view provides a solid explanation of the phenomenon on the table. What, then, is wrong with the explanation that the IDS view provides? Well, the IDS view does not provide an explanation regarding why it is preferable to consult the agent's idealized desires rather than his actual desires when it comes to determining what the agent has reasons to do. To see how this is so, let us briefly return to the accurate watch case. The conventional view about the time, in general, and what goes on when I want to know the time, goes something like this. When I want to know the time, I consult my watch, and not my father's Rolex. This is because I believe that my watch tracks the facts about the time well, and my father's Rolex does not track the facts about the time well. The facts about the time hold independently of what any watch says, and my watch does a good job at telling us the facts about

the time, and so that is why I consult my accurate watch rather than the Rolex.

Let us consider an analogous explanation that can account for agents' having reasons to act and what grounds their reasons. When I want to know what reasons an agent has to do something or other, I think about his idealized desires rather than his actual desires. His actual desires do not seem to line up well with his reasons to act; for sometimes agents desire to do things that are cruel, self-destructive, meaningless, and so on. Instead, I think about the set of the agent's desires that are not rooted in false beliefs, or that would survive the practical deliberation process, or that would be introduced during the practical deliberation process. This idealized set of desires seems to match up better with the agent's reasons to act. And there are facts about the agent's reasons to act that hold independently of his actual or idealized desires. The virtue of idealized desires, analogous to the accurate watch, is that they are thought to line up well with what agents have reasons to do. In other words, by hypothesis, what an agent would desire after putting his desires through the practical deliberation process is what he would have reasons to do; for the practical deliberation process helps him to track what he has reasons to do. This is a simple explanation of what is going on with agents and their reasons to act, and it provides an independent explanation regarding why we consult the agents' idealized desires rather than his actual ones.

In contrast, the IDS theorist cannot explain why we consult the agent's idealized desires rather than his actual ones. This is because the IDS theorist holds that there is nothing over and above the agent's idealized desires that provide him with reasons to act. Because of this commitment, the choice, to consult the idealized desires rather than the actual ones when attempting to determine what an agent has reasons to do, seems arbitrary. There appears to be no grounds that the IDS theorist can use to motivate the appeal to idealization except that the appeal

to idealization saves the desire-satisfaction view from the ADS problem. But this is just what is ad hoc about the appeal to idealization.

Summing up now, I have considered the claim that the IDS view provides the best explanation of reasons for action and what grounds them. If this claim is true, it would count as independent motivation for the IDS appeal to idealization, and thus the IDS view could escape what I have called the IDS problem. My counter to this claim is just that the IDS view does not provide, or at the least, it is not clear that the IDS view provides a solid explanation of reasons for action and what grounds them. To escape my objection, the IDS theorist would need to establish that the IDS view provides a solid explanation of reasons to act and what grounds them. I do not think that the IDS theorist can accomplish this. There is indeed another candidate view, that is, another view that provides an explanation of reasons and what grounds them that seems very natural and intuitive. In the previous paragraph, I hinted towards this view. It is, very roughly, the view that facts about agents' reasons to act hold independently of their actual or idealized desires. Of course this thesis needs support of its own, and interestingly, providing support for a thesis like this is largely what I will be up to in chapter 3. But at least this thesis seems to provide a much better explanation of the phenomenon on the table for it can explain why we would consult the agents' idealized desires rather than his actual desires when trying to determine what his reasons are to act. And that is all I need at this point to uphold my objection to the IDS view. Given the IDS problem, the burden of proof is on the IDS theorist to show that the IDS view provides a solid explanation of reasons to act and what grounds them. I have, at the very least, raised a solid reason for doubting that the IDS view provides a good explanation of reasons to act and what grounds them; in chapter 3, I will make the case that there is another view that provides the best explanation of reasons to act and what grounds them. Since the IDS

theorist cannot, it seems, establish that the IDS view provides a solid explanation for the phenomenon on the table, it seems that the IDS theorist fails to provide independent motivation for the IDS appeal to idealization, and thus is unable to escape the IDS problem. Now perhaps the IDS theorist could provide some independent support for idealization some other way, but I am just not sure how this would go. And so I am comfortable concluding that the IDS view is untenable.

## 2.9 Conclusion

Recall that the desire-satisfaction view is a thesis which purports to explain just what it is that necessarily provides agents with intrinsic reasons to act. The desire-satisfaction view holds that it is an agent's desires, in some sense, that necessarily provide the agent with intrinsic reasons to act. Contrast the desire-satisfaction view with another candidate theory, namely intrinsic moral rationalism, which also purports to explain just what it is that necessarily provides agents with intrinsic reasons to act. Intrinsic moral rationalism holds that it is morality that necessarily provides agents with intrinsic reasons to act. I attempted to dispense with intrinsic moral rationalism in chapter 1, and I have attempted to dispense with the desire-satisfaction view in chapter 2. I have distinguished between two versions of the desire-satisfaction view namely the ADS view and the IDS view. I rejected the ADS view because it leads to the ADS problem, namely the troubling commitment to propositions (1) and (2). I rejected the IDS views because their commitment to idealization is troublingly ad hoc. In that the desire-satisfaction view is untenable, we have grounds for rejecting all versions of extrinsic moral rationalism and moral antirationalism that hold that the desire-satisfaction view is true.

In chapter 1, I made a preliminary case for moral rationalism unqualified, but I argued against intrinsic moral rationalism, suggesting that there are strong reasons for considering extrinsic moral rationalism. By my lights, the most plausible versions of extrinsic moral rationalism hold that either desires or self-interest necessarily provide agents with reasons to act. In chapter 2, I argued against the thesis that desires necessarily provide agents with reasons to act. So we have strong grounds for considering the thesis that self-interest necessarily provides agents with reasons to act. I will attempt to defend this thesis in chapter 3.

## Chapter 3

### Self-Interest and Reasons to Act

I will spend the rest of this paper arguing for the thesis I am calling *rational egoism*; it holds that self-interest necessarily provides agents with intrinsic reasons to act. By self-interest, I mean, roughly, that which is objectively (not necessarily morally) good, valuable, or desirable for an agent.<sup>33</sup> I will proceed by arguing that rational egoism captures the most important theoretical merits of its major competitor theories and yet avoid the problems that its major competitor theories are vulnerable to. However, even if it is not totally clear that rational egoism can capture every theoretical merit of its competitor theories and avoid every problem that its competitor theories are vulnerable to, or if it is not clear that rational egoism is a better theory all things considered, I will conclude that rational egoism is, at least, on equal footing with its competitor theories, and deserving of equal consideration as a theory about what it is that necessarily provides agents with intrinsic reasons to act. I think this is important because, at least in the contemporary literature, rational egoism appears to have been neglected. As I go along, I will consider objections and replies to rational egoism. I will then close with some considerations regarding rational egoism and moral rationalism.

---

<sup>33</sup>I do not here distinguish between *the good*, *the valuable*, and *the desirable*.

### 3.1 What is Self-Interest?

I follow Railton's account of *self-interest*.<sup>34</sup> I have chosen to keep with Railton because there are two main ideas about self-interest which I think an account of self-interest should capture, and Railton's account captures both of them. The first idea is that an agent's self-interest is something different from what that agent desires, and the second idea, very much related to the first, is that an agent's self-interest is what is, in fact, good for him. My view is that, although an agent may, of course, desire what is good for him, facts about what is in an agent's self-interest are not relativized to what an agent desires. I have chosen to keep with Railton because he captures these ideas about self-interest in a way that I find plausible. I will now get into the details of his account.

Railton explains what it means for something to be in an agent's self-interest by way of what it means for something to be in an agent's *subjective interest*. Something *x* is in an agent *A*'s subjective interest just in case *A* actually desires *x*. For example, drinking a gin and tonic is in Jim's subjective interest just in case Jim actually desires to drink a gin and tonic. Each of an agent's subjective interests, according to Railton, can be reduced to qualities of the agent, qualities of the object of the agent's subjective interest, and qualities of the surrounding environment. Railton calls this complex set of qualities *the reduction basis* of the subjective interest. Some of the qualities that form the reduction basis of Jim's subjective interest in drinking a gin and tonic might be, for example, Jim's desiring to relax, and gin's ability to produce the relaxing effect.

---

<sup>34</sup>As a point of clarification, Railton does not use the term *self-interest*; he uses the term, *non-moral good*. I, for the most part, keep with the language of self-interest. Peter Railton, "Moral Realism," *The Philosophical Review* 95 (1986): 163-207.

Next, Railton introduces the notion of an *objectified subjective interest* of an agent. He asks us to consider an agent A, who is given

unqualified cognitive and imaginative powers, and full factual and nomological information about his [A's] physical and psychological constitution, capacities, circumstances, history, and so on...whose instrumental rationality is in no way defective.<sup>35</sup>

Given these powers, we now call him A+. The *objectified subjective interest* of agent A is that which A+ would want if A+ were in the actual circumstances and condition of A. And just as each of A's subjective interests has a corresponding reduction basis, each of A's objectified subjective interests has a corresponding reduction basis,

namely, those facts about A and his circumstances that A+ would combine with his general knowledge in arriving at his views about what he would want to want were he to step into A's shoes.<sup>36</sup>

Now, importantly, x is in A's self-interest just in case x is in A's objectified subjective interest.

Before going any further, a brief comment is in order about the distance between A and A+. A common worry with Railton's account is that A+ could turn out to be so different from A that it would be strange to connect what is in A's self-interest with what A+ would want were he to step into A's shoes. If A+ were too different from A, what A+ would want would be impertinent to A. Of course, it is difficult to imagine what a virtually omniscient agent would be like, but by hypothesis, A+ is not very different from A after all. A+ is almost identical to A; A+ just has all the knowledge and full factual and nomological information about A's physical and psychological constitution, capacities, circumstances, history, and so on, and has perfect practical

---

<sup>35</sup>Ibid., 173-174.

<sup>36</sup>Ibid., 174.

deliberation skills. So, very basically, what A+ would want for A, is what A would want for himself if he knew best.

Another important clarification is in order about what fixes the truth about A's self-interest. It is the reduction basis of A's objectified subjective interest that makes it the case that A has a certain objectified subjective interest, and so it is the reduction basis of A's objectified subjective interest that makes it the case that something x is in A's self-interest. An important point is that it is not the case that what A+ would desire if he were in the actual condition and circumstances of A that determines what is in A's self-interest; it is the reduction basis of A's objectified subjective interest that makes it the case that A+ would desire something or other if he were in the actual condition and circumstances of A, and so it is the reduction basis of A's objectified subjective interest that determines what is in A's self-interest. Perhaps there is a simpler way of making this point. It is not what an ideal version of an agent would want for the actual version of himself that determines what is in the actual agent's self-interest; it is a complex set of facts about the actual agent and his environment that determines what is in the actual agent's self-interest. The ideal version of the agent, though, would have access to this complex set of facts and would be able to deliberate perfectly and so, by hypothesis, he would desire just what is in the actual agent's self-interest if he were in the circumstances and condition of the actual agent. The point is that it is not the case that x is in the actual agent's self-interest because an idealized version of the actual agent would desire x; an idealized version of the actual agent, by hypothesis, would desire x because x is in the actual agent's self-interest. Importantly, the reference to A+ or the idealized version of the actual agent is just meant as a useful conceptual tool to help us think about the sort of things that are in our self-interest.

It might help to take a look at Railton's example. He asks us to imagine Lonnie, a traveler in a foreign country who is homesick and dehydrated.<sup>37</sup> Lonnie actually desires to drink a glass of milk, and he believes that drinking a glass of milk will ease his homesickness and upset stomach, which, unbeknownst to him, is caused by his dehydration. So it is in Lonnie's subjective interest to drink a glass of milk. As a matter of fact, drinking a glass of milk will only worsen his dehydration, and accordingly, his stomach upset. Lonnie-Plus knows that Lonnie's upset stomach is caused by dehydration, Lonnie-Plus knows that Lonnie desires a glass of milk, and Lonnie-Plus knows that drinking a glass of milk will only worsen Lonnie's dehydration, and accordingly, his stomach upset. Lonnie-Plus also knows that abundant clear liquids will improve Lonnie's stomach upset and homesickness. Indeed, it is in Lonnie's objectified subjective interest to drink abundant clear liquids, and so, it is in Lonnie's self-interest to drink abundant clear liquids. Lonnie-Plus knows this, and so if Lonnie-Plus were in the actual circumstances and condition of Lonnie, Lonnie-Plus would desire, for Lonnie, abundant clear liquids rather than milk.<sup>38</sup> And importantly, it is the reduction basis of Lonnie's objectified subjective interest, that is, a complex set of facts about Lonnie, clear liquids, and Lonnie's environment that makes it true that it is in Lonnie's self-interest to drink abundant clear liquids. These facts may include: the fact that Lonnie is dehydrated, the fact that clear liquids aid in the healing of dehydration, the fact that Lonnie would feel relief from his symptoms if he were not dehydrated, and so on. Basically, Lonnie is so constituted that, in the sort of circumstance he is in, it is in his good to drink abundant clear liquids.

---

<sup>37</sup>Ibid., 174.

<sup>38</sup>Just suppose that, alternatively, for Lonnie, milk cures dehydration and homesickness, and that this is a part of Lonnie's essence. In other words, suppose that Lonnie is just the sort of creature for which milk cures dehydration and homesickness. If this were the case, then Lonnie-Plus would be aware of this fact, and, if he were in Lonnie's shoes, he would desire milk rather than clear liquids.

### 3.2 Rational Egoism

Since I have now presented a working account of self-interest, I will continue with my argument in favor of rational egoism by discussing what sort of case needs to be made for a thesis like rational egoism. Let us here recall that a *reason*, as I use the term, is a genuine consideration in favor of doing something or other. If agents do indeed have the sorts of reasons that I am talking about, there must be something in virtue of which agents have these reasons. The question is: in virtue of *what* do agents have reasons to act? Rational egoism holds that it is in virtue of their self-interest, or what is, in fact, in their own good, that agents have reasons to act. Another candidate theory, namely intrinsic moral rationalism, holds that it is in virtue of morality that agents have reasons to act, and a third candidate theory, namely the desire-satisfaction view, holds that it is in virtue of their desires that agents have reasons to act. So which theory seems the most plausible? In chapter 1, I argued that intrinsic moral rationalism is untenable because it includes a problematic element of mystery. In chapter 2, I argued that the desire-satisfaction view is untenable because it is either vulnerable to what I call the ADS problem or the IDS problem. Because of the problems with these theories, we have solid grounds for considering rational egoism.

In addition to the problems with its competitor theories, there are solid, pre-theoretical, common sense grounds for considering rational egoism. It is plausible to think that that which is in an agent's self-interest, or that which is, in fact, good for an agent, necessarily provides him with intrinsic reasons to act. In virtue of broccoli being good for me, it seems that there is necessarily an intrinsic reason to eat it, even if I do not desire it.

At this point it is also useful to make the point that intrinsic moral rationalism, rational egoism, and the desire-satisfaction view are not mutually exclusive. That is, one can consistently

accept all three candidate theories. One could hold that morality, self-interest, and desires all necessarily provide agents with intrinsic reasons to act. Of course, one who accepted more than one of the three theories would need an account of how to weigh considerations against each other in cases in which they were conflicting. For example, if one accepted both instrumentalism and rational egoism, one would need to provide an account of whether it would be rational or more rational to do that which one desired or that which was in one's self-interest. This is not a matter that I am too concerned with here. However, as I proceed with a case in favor of rational egoism, it is useful to note that accepting rational egoism does not ultimately rule out the desire-satisfaction view, and intrinsic moral rationalism. So readers who are strongly pulled by either the desire-satisfaction view or intrinsic moral rationalism do not need to give up their predilections in order to accept rational egoism.<sup>39</sup>

Now, provided the problems with its competitor theories and the common sense grounds in its favor, I will continue with a case in favor of rational egoism. In order to make a case in favor of rational egoism, I will try to show that rational egoism captures the most important theoretical virtues of its major competitor theories and avoids the problems that its major competitor theories are vulnerable to. I will also try to defend rational egoism against objections, and show that rational egoism is not vulnerable to devastating, special problems of its own. I will proceed as follows. In section 3.3, I will argue that rational egoism avoids the major problem that the ADS version of the desire-satisfaction view is vulnerable to. In section 3.4, I will argue that rational egoism captures the major theoretical virtue of the desire-satisfaction view. In section 3.5, I will argue that rational egoism avoids the major problem that the IDS

---

<sup>39</sup>This is not to say that each theory will yield the same answer. For example, the desire-satisfaction view might indicate that Jim has a reason to drink a gin and tonic, and rational egoism might indicate that Jim has a reason to meditate instead. One could accept both the desire-satisfaction view and rational egoism, and hold that Jim has a prima facie reason to drink a gin and tonic, and a prima facie reason to meditate. But then, some account would be needed that could explain what Jim has an absolute reason to do.

versions of the desire-satisfaction view are vulnerable to. In section 3.6, I will argue that rational egoism avoids the major problem that intrinsic moral rationalism is vulnerable to. Lastly, in section 3.7, I will argue that rational egoism might be able to capture the virtue of intrinsic moral rationalism, but even if it cannot, this would not be grounds for rejecting rational egoism. Along the way, I will defend rational egoism against objections that could be raised against it.

### 3.3 Rational Egoism and the ADS Problem

In this section, I will argue that rational egoism escapes the major problem that the ADS version of the desire-satisfaction view is vulnerable to. Recall that the ADS view holds that agents' actual desires necessarily provide agents with intrinsic reasons to act. The first part of the ADS problem is that the ADS view is committed to the troubling proposition that even an agent's desires to do the most self-destructive, cruel, and meaningless things necessarily provide him with reasons to act—that is, an agent's desires necessarily provide him with intrinsic reasons to do the most self-destructive, cruel, or meaningless things. This is a problem for the ADS view because it seems inaccurate to say that desires such as these necessarily provide agents with intrinsic reasons to act, or that agents necessarily have reasons to do things like these. For example, it seems inaccurate to say that Hitler, who desires to kill the Jews, necessarily has a reason to do so in virtue of him desiring it. It seems inaccurate to say that Jim, given all his trouble with gin and tonics, necessarily has a reason to drink a gin and tonic just in virtue of him desiring to. And it seems inaccurate to say that an agent who desires to count all the blades of grass in his yard necessarily has a reason to count all the blades of grass in his yard, just in virtue of him desiring to count all the blades of grass in his yard. The second part of the ADS problem is the ADS view's commitment to the troubling proposition that either we ought to unfairly blame

agents for doing things that they necessarily had a reason to do, or we ought not blame agents for doing the most destructive, cruel, or meaningless things if they desired to do such things.

Now, how could the ADS problem, or a similar one, possibly become a problem for rational egoism? The ADS problem, or a similar one, could become a problem for rational egoism if it turned out to be in an agent's self-interest to do the most self-destructive, cruel, or meaningless things because then the rational egoist would be committed to the troubling proposition that agents necessarily have reasons to do the most self-destructive, cruel, or meaningless things. In contrast, if it never turned out to be in an agent's self-interest to do the most self-destructive, cruel, or meaningless things, then, for the the rational egoist, it would never turn out that agents necessarily have a reason to do the most self-destructive, cruel, or meaningless things. I will proceed here by considering why we might think that it could ever turn out to be in an agent's self-interest to do the most self-destructive, cruel, or meaningless things. Then I will argue that even if there are cases in which it is in an agent's self-interest to do the most self-destructive, cruel, or meaningless things, these cases would be so out of the ordinary that the problem created for rational egoism would not be relevantly analogous to the ADS problem, and it would not be a devastating problem for rational egoism. For the ADS theorist is quite often forced to conclude that agents necessarily have reasons to do the most self-destructive, cruel, or meaningless things because agents quite often desire to do things that are self-destructive, cruel, or meaningless, and it is only in relatively rare cases that it will ever turn out to be in an agent's self-interest to do something self-destructive, cruel, or meaningless. So although the rational egoist might be vulnerable to a problem similar to the ADS problem, it will not be a devastating problem for the rational egoist.

First, why might we think that it could ever turn out to be in an agent's self-interest to do the most self-destructive, cruel, or meaningless things? We may think about this question in two ways: could there ever be a complex set of facts that would make it the case that it was in an agent A's self-interest to do the most self-destructive, cruel, or meaningless things, or would A+, if he were in A's circumstances, ever desire to do the most self-destructive, cruel, or meaningless things?

Let us think about self-destructive things first. It seems that it will almost never turn out to be in an agent's self-interest to do something self-destructive. Let us again consider Lonnie and his desire to drink milk. Drinking milk, in his situation, would be self-destructive for Lonnie because it would worsen his dehydration and discomfort. Lonnie-Plus knows all the facts about Lonnie and his environment, and so he knows that drinking milk would be self-destructive for Lonnie because it would worsen his condition. Even though Lonnie desires a glass of milk rather than abundant clear liquids, Lonnie-Plus, if he were in Lonnie's situation, would desire abundant clear liquids because he knows that abundant clear liquids will improve Lonnie's condition, and Lonnie-Plus is a perfect rational deliberator. Similarly, we could imagine Jim-Plus, an idealized version of Jim, who, given Jim's trouble with gin and tonics, would presumably not desire gin and tonics if he were in Jim's shoes. Knowing that gin and tonics lead to the worse sorts of trouble for Jim, and being able to deliberate perfectly, Jim-Plus would likely desire something other than gin and tonics.<sup>40</sup>

---

<sup>40</sup>Maybe Jim-Plus would desire gin and tonics. But knowing what we know about how drinking alcohol in excess can have a negative impact on a human's well-being, it seems natural to doubt that Jim+ would desire to drink gin and tonics if he were in Jim's shoes, especially given Jim's past trouble with gin and tonics. However, Jim could be experiencing some awful circumstances, and gin and tonics might be the only thing that makes his life bearable. If this were the case, then maybe Jim+ would desire gin and tonics if he were in Jim's shoes. But interestingly, if drinking gin and tonics were the only thing making Jim's life bearable, we might not call his drinking gin and tonics a self-destructive habit.

Next, let us consider if it could ever turn out to be in an agent's self-interest to do something meaningless. If something were in the self-interest of an agent, it seems like it would be meaningful in virtue of it serving the agent's self-interest. So it seems impossible that it could be in an agent's self-interest to do anything meaningless. Also, it might help to consider, for example, an actual agent, Larry, who desires to count all the blades of grass in his yard. Suppose also that counting all the blades of grass in one's yard is, in fact, a meaningless activity. It seems likely that Larry-Plus who has all the knowledge about Larry and his environment, and could deliberate perfectly, would come to desire something other than counting all the blades of grass in his yard. Larry-Plus would know that, although counting all the blades of grass in the yard serves some strange desire of Larry's, using the time to do philosophy, create art or music, work for positive social change, exercise, or do something else, would probably better serve Larry's self-interest. Larry-Plus would know that human beings are the sorts of creatures whose self-interest is not served by doing repetitive tasks for no purpose other than to satisfy strange compulsions. And with his perfect rational deliberation skills, Larry-Plus would likely come to desire something other than counting all the blades of grass in his yard.

Lastly, let us consider cruel things. It does seem that, in some cases, it could serve an agent's self-interest to do something cruel. Perhaps, in prison for example, an inmate might do something cruel in order to scare other inmates away from potentially harming him or bullying him in the future. Performing one act of cruelty might serve the inmate's self-interest because it would ensure that he was protected from future physical harm or bullying by other inmates. But equally, we can imagine that refraining from doing the cruel thing and taking another course of action could even better serve the inmate's self-interest. Instead of guaranteeing his own physical safety through acts of cruelty, the inmate could start some sort of self-improvement

movement within the prison that could create community, brotherhood, and respect among the inmates. Suppose this would guarantee his physical safety and freedom from bullying. We might also imagine a shipwrecked sailor who is forced by natives to choose to either do something cruel, or face death. It seems like it would be in the sailor's self-interest to do the cruel thing in order to preserve his own life. But equally, we might imagine that the sailor might never be able to live with himself after doing the cruel thing. Perhaps his guilt would plague him, and he would suffer for the rest of his life disgusted by the cruel deed that he had done. In this case, we might think it would have better served his self-interest to accept death rather than a life of misery.<sup>41</sup>

My point here is that self-interest is complicated and, in many cases, it can be difficult to determine what is in an agent's self-interest. I think it is at the very least difficult to imagine cases in which it would be in an agent's self-interest to do the most self-destructive, cruel, or meaningless things. However, there will never be certainty about what is in agents' self-interest because we can always discover new facts that might overturn propositions that we thought were true about agents' self-interest. Thus, on the account of self-interest that I use, I must admit that it could turn out to be in an agent's self-interest to do the most self-destructive, cruel, or meaningless things. But I think these cases would be rare in contrast with cases in which actual agents desire to do things that are self-destructive, cruel, or meaningless; cases in which actual agents desire things that are self-destructive, cruel, or meaningless seem relatively ordinary. We can imagine all sorts of cases like Jim's: people who desire to drink too much, eat too much, smoke too much, or in general, desire to do things that are self-destructive, or meaningless, or even cruel. So, it is acceptable for the rational egoist to say, okay, in some *relatively rare* cases,

---

<sup>41</sup>We could run these same examples using meaninglessness or self-destructiveness. For instance, we could imagine that the natives force the sailor to do a meaningless or self-destructive thing.

it might turn out to be in an agent's self-interest to do something self-destructive, cruel, or meaningless, and so in some *relatively rare* cases, agents necessarily have reasons to do the most self-destructive, cruel, or meaningless things. But these cases are so out of the ordinary that it is not a devastating problem for the rational egoist. The ADS theorist, in contrast, will quite ordinarily be forced to conclude that agents necessarily have reasons to do the most self-destructive, cruel, or meaningless things. So the rational egoist is not as troubled by the ADS problem, or an analogously severe one; the problem is not a devastating one for the rational egoist like it is for the ADS theorist.

### 3.4 Rational Egoism and the Theoretical Virtue of the Desire-Satisfaction View

In this section I will argue that rational egoism is able to capture the major theoretical virtue of the desire-satisfaction view. The major theoretical virtue of the desire-satisfaction view is that it is able to acknowledge the importance of desires, and account for the relationship between an agent's desires and his reasons to act. Recall that the desire-satisfaction view holds that an agent's desires necessarily provide him with intrinsic reasons to act. *Desires*, as I use the term, include ends, goals, projects, wants, and commitments among other things that could be counted as elements of an agent's subjective motivational set. Desires are important in the following sense: they are, in some way, close to the agent, they perhaps constitute what might be called the agent's character, and in many cases, their fulfillment might be what an agent lives for.<sup>42</sup> This is why I consider it to be a virtue of the desire-satisfaction view that it is able to acknowledge the importance of desires, and account for the relationship between an agent's desires and his reasons to act.

---

<sup>42</sup>Williams, "Persons, Character, and Morality," 1-19.

Now, I will discuss the way in which rational egoism is also able to acknowledge the importance of desires, and account for the relationship between an agent's desires and his reasons to act. Rational egoism captures the intuition that an agent's actual desires provide him with reasons to act for rational egoism holds that, indeed, some of the agent's actual desires—the desires for things that are in his self-interest—do necessarily provide him with reasons to act, just in virtue of them being for things that are in the agent's self-interest. It might be objected, though, that rational egoism does not do well at capturing the idea that actual desires provide agents with reasons to act for it turns out that rational egoism holds that only desires for objects that are in an agent's self-interest necessarily provide agents with reasons to act. In reply, I maintain that rational egoism does well at capturing the idea that actual desires provide agents with reasons to act because rational egoism can grant that some of the agents' actual desires necessarily provide agents with reasons to act, and at the same time, rational egoism is not vulnerable to the ADS problem. Indeed, it is not a virtue of the ADS view that it holds that all of an agent's actual desires—even the desires for things that are self-destructive, cruel, or meaningless—necessarily provide the agent with intrinsic reasons to act; the virtue of the desire-satisfaction view is that it is able to acknowledge the importance of desires and provide an account of the relationship between an agents' desires and his reasons to act. Rational egoism shares this virtue all while escaping the ADS problem.

Of course this objection to rational egoism could be developed if we imagine agents who desire only things that are not in their self-interest. According to rational egoism, agents like this would have no reasons to pursue the objects of their desires. Alas, I think this is as it should be. Furthermore, I think agents like these are rare and that normally there will be quite an overlap between actual desires and self-interest. Since facts about self-interest are reducible to facts

about the agent, the objects of his desires, and his environment, facts about the agent's actual desires will play a part in determining what it is that is in an agent's self-interest. In sum, rational egoism gets things right concerning the relationship between desires and reasons to act; rational egoism safeguards a place for desires and their connection with reasons for action without being vulnerable to the ADS problem.

### 3.5 Rational Egoism and the IDS Problem

Next, I will discuss how rational egoism is not vulnerable to the IDS problem. Recall that the IDS problem is that the IDS appeal to idealization is troublingly ad hoc. A theoretical mechanism is ad hoc if it, in some sense, saves the day for the theory by helping it to answer objections or overcome problems that are raised for it, yet is itself unmotivated. The IDS appeal to idealization saves the day, in some sense, for the IDS view because it can allow the IDS view to escape the ADS problem, yet the IDS appeal to idealization is itself unmotivated.

So how is rational egoism not ad hoc? Quite simply, rational egoism does not necessarily include any idealization; the reference to the idealized agent is just a device that helps us conceptualize what it is that is in an agent's self-interest. In other words, rational egoism is not ad hoc because rational egoism includes no additional theoretical mechanism that is itself unmotivated. To further illustrate how rational egoism is not ad hoc, we may compare it to the ADS view and contrast it with the IDS view. The ADS view just holds that the agent's actual desires *unqualified* necessarily provide him with intrinsic reasons to act; there is no additional theoretical mechanism. Similarly, rational egoism holds that the agent's self-interest *unqualified* necessarily provides him with intrinsic reasons to act; also, there is no additional theoretical mechanism. Contrastingly, the IDS view holds that the agent's desires *idealized* necessarily

provide him with intrinsic reasons to act. It is the appeal to the *idealized* part, or the idealization process, that is ad hoc.

It might be objected that the idealization, or the reference to the idealized agent A+, included in Railton's account of self-interest is ad hoc. Since I follow Railton's account of self-interest, my version of rational egoism would be vulnerable to this objection. How might this objection go? Recall that it is the appeal to idealization that makes the IDS view ad hoc.

Railton's account includes something like this, namely the reference to the idealized agent A+.

So one might object that Railton's reference to the idealized agent A+ is ad hoc. But this

objection is very easy for the rational egoist to handle because what A+ would want if A+ were in the actual circumstances of A does not fix the truth about what is in A's self-interest. Appeal

to A+ is supposed to be helpful because, by hypothesis, A+ would desire what is in A's self-

interest, and thinking about what A+ would want is meant to help us think about what is in A's

self-interest. But facts about an actual agent's self-interest are fixed by facts about the actual

agent, the objects of his interest, and his environment. So ultimately, the rational egoist can let

go of any reference to A+ if need be. In contrast, the IDS theorist cannot let go of the

idealization process in their account because then their view would just collapse into the ADS

view and it would be vulnerable to the ADS problem. Since reference to A+ or the idealized

agent is not essential to rational egoism, rational egoism is not ad hoc.

It might be objected that although Railton's account of self-interest is not ad hoc, it is,

nonetheless, ontologically problematic. Railton holds that the truth about an agent's self-interest

is fixed by the reduction basis of the objectified subjective interest of the agent, and the reduction

basis of the objectified subjective interest of the agent is a complex set of facts about the agent,

the agent's desires, and the surrounding environment. The objection is that the reduction basis is

incredibly complex, or just too difficult to make sense of, and since the reduction basis is what purportedly fixes the facts about the agent's self-interest, we ought to reject Railton's account of self-interest. The complexity of the reduction basis that Railton is committed to is just too much to accept.

There are two ways that the rational egoist could reply to this ontological objection. Keeping with Railton's account of self-interest, the rational egoist could argue that there is really nothing the matter with the incredible complexity of the reduction basis. After all, the universe is complex and we are complex creatures; it is no surprise that facts about agents' self-interest are fixed by complex sets of facts about the agents, their desires, and the environment. This reply works for the rational egoist because it challenges the intuition that there is a problem with the complex ontology that Railton is committed to.

A second reply open to the rational egoist is to grant that there are problems with Railton's ontology, and to claim that this is a problem with Railton's account of self-interest rather than a problem with rational egoism. The rational egoist could then claim that all that is needed is just a better account of self-interest, and that it would be open to anyone to supply this. Indeed, it surely seems that agents have a self-interest, or that there are things that are good for agents, and that this *good* for agents is different from and not relativized to their desires. It might just be a matter of finding the best account of this phenomenon.

### 3.6 Rational Egoism and the Problem with Intrinsic Moral Rationalism

The main problem with intrinsic moral rationalism is that it contains a troubling element of mystery. I will now argue that rational egoism avoids the mystery problem that intrinsic moral rationalism is vulnerable to, or at least that rational egoism is no worse off than intrinsic moral

rationalism with regards to the mystery problem. It will help to recall from chapter 1, that roughly, the mystery problem is that intrinsic moral rationalism seems to grant morality some mysterious force that allows it to impose upon agents and somehow give them reasons to act. If intrinsic moral rationalism is true, then morality seems to be like a mystical umbrella hanging over the heads of everyone and giving them reasons to act; this is a mysterious view. It is mysterious how morality necessarily provides agents with intrinsic reasons to act and mysterious why agents necessarily have an intrinsic reason to do the moral thing.

Why might one think that rational egoism is also mysterious? One might think that rational egoism, like intrinsic moral rationalism, is mysterious because one might think that the thesis of rational egoism grants *self-interest* a mysterious power that allows it to impose upon agents and give them reasons to act, or that self-interest, for the rational egoist, is like a mystical umbrella hanging over the heads of everyone and giving them reasons to act. One might argue that it is mysterious how self-interest necessarily provides agents with intrinsic reasons to act and mysterious why agents necessarily have an intrinsic reason to do that which is in their self-interest.

And I will grant that surely rational egoism is mysterious in some sense. In chapter 1, I raised what I called *the worry with intrinsic reasons*. The worry is that it is just mysterious how any consideration can be intrinsically reason-giving; just as intrinsic moral rationalism has to account for the intrinsically reason-giving character of morality, rational egoism has to account for the intrinsically reason-giving character of self-interest. And the desire-satisfaction view, too, has to account for the intrinsically reason-giving character of desires.

But intrinsic moral rationalism is mysterious in a second sense. And here is why. The considerations that rational egoism and the desire satisfaction view hold to be intrinsically

reason-giving, namely self-interest and desires, are considerations that are quite intuitive candidate sources for intrinsic reasons. That is, it is quite natural to think that things like an agent's self-interest and desires are the sort of things that necessarily provide agents with intrinsic reasons to act. These things are not foreign, alien, or external to the agent like morality is. No mysterious force is needed to explain how an agent's desires or self-interest could provide the agent with intrinsic reasons to act. Thus rational egoism does not include the further element of mystery that intrinsic moral rationalism does. In contrast, morality, like etiquette, is foreign, alien, and external to the agent. If intrinsic moral rationalism was true, morality would need some mysterious force behind it that could allow it to impose upon agents and give them reasons to act regardless of their desires or self-interest. Thus, I maintain that intrinsic moral rationalism is *uniquely* mysterious, that is, mysterious in a way that our other accounts of intrinsic reasons are not.

So I am confident in concluding that rational egoism avoids the mystery problem that intrinsic moral rationalism is vulnerable to. But suppose the intrinsic moral rationalist were to dig in her heels and dwell on the point that, just as the intrinsic moral rationalist cannot explain how morality necessarily provides agents with intrinsic reasons to act, or why agents necessarily have an intrinsic reason to do the moral thing, neither can the rational egoist explain how self-interest necessarily provides agents with intrinsic reasons to act, or why agents necessarily have an intrinsic reason to do that which is in their self-interest. For this analogous lack of explanatory power, the intrinsic moral rationalist might conclude that rational egoism is just as mysterious as intrinsic moral rationalism.<sup>43</sup> Even if she were right, it would still turn out that rational egoism is no worse than intrinsic moral rationalism with regards to the mystery problem, and that is all I really needed to demonstrate here. As stated in the introduction to this chapter,

---

<sup>43</sup>Shafer-Landau makes this sort of argument. Shafer-Landau, 209-211.

although I make some arguments for the conclusion that rational egoism is able to capture the merits of its competitor theories all while avoiding their flaws, my primary goal is smaller, and that is just to show that rational egoism is equally deserving of consideration as a theory about what it is that necessarily provides agents with intrinsic reasons to act.

### 3.7 Rational Egoism, Moral Egoism, and Extrinsic Moral Rationalism

In this chapter section, I will provide some grounds for believing that rational egoism is able to capture the main virtue of intrinsic moral rationalism, and then I will argue that, even if rational egoism fails to completely capture the main virtue of intrinsic moral rationalism, it would not be a devastating problem for rational egoism. To be clear on the main virtue of intrinsic moral rationalism, it will help to recall from chapter 1 my presumptive argument in favor of moral rationalism (unqualified) in which I argued that moral rationalism fits well with our conventional ways of thinking and talking about morality. Recall that moral rationalism is the thesis that there is necessarily a reason to do the moral thing. Moral rationalism is neutral with regard to the debate between intrinsic moral rationalism and extrinsic moral rationalism. That is, moral rationalism leaves open whether it is morality itself that intrinsically provides agents with a reason to do the moral thing, or it is in virtue of a necessary connection between morality and some other intrinsically reason-giving sort of consideration that morality necessarily provides agents with a reason to do the moral thing. As a version of moral rationalism, intrinsic moral rationalism fits well with our conventional ways of thinking and talking about morality, and I take this to be the main virtue of intrinsic moral rationalism.

Now, note that extrinsic moral rationalism, if it were true, as another version of moral rationalism, would also fit well with our conventional ways of thinking and talking about

morality. Just to be clear, as a version of moral rationalism, extrinsic moral rationalism holds that morality necessarily provides agents with a reason to do the moral thing, and this is the thesis about moral reasons that our conventional ways of thinking and talking about morality seem to point us to. Thus, extrinsic moral rationalism, like intrinsic moral rationalism, has the theoretical virtue of fitting well with our conventional ways of thinking and talking about morality. So, if rational egoism can be paired with extrinsic moral rationalism, then rational egoism can capture the main virtue of intrinsic moral rationalism. But for rational egoism to be paired with extrinsic moral rationalism, the following thesis about normative morality will have to be true: morality just requires that agents do that which is in their individual self-interest.

I will hereafter refer to this as the thesis of *moral egoism*. For if moral egoism is true and rational egoism is true, we will be left with the following formulation of extrinsic moral rationalism: agents necessarily have a reason to do the moral thing because agents necessarily have a reason to do what is in their individual self-interest, and morality just requires agents to do that which is in their individual self-interest. Since moral egoism plus rational egoism gets us to a version of extrinsic moral rationalism, if moral egoism is true, then rational egoism will fit well with moral convention and thus share the main virtue of intrinsic moral rationalism. In this chapter section, I will have more to say about moral egoism and why we might believe that it is true.

But I must also consider the other possibility. If moral egoism is not true, then rational egoism is not quite capable of capturing the virtue of intrinsic moral rationalism. If moral egoism is not true, then rational egoism will be committed to a version of moral antirationalism, the thesis that there is not necessarily a reason for agents to do the moral thing. For if moral egoism is not true, it is possible that there will be cases in which morality requires agents to do

things that are not in their individual self-interest, and rational egoism will have it that there might not be a reason for agents to do these moral things. However, depending upon what morality requires of agents, it might turn out that the version of moral antirationalism that rational egoism is committed to, if moral egoism is false, is not a very troubling version of moral antirationalism. To be more specific, depending upon what morality requires, rational egoism could lead us to the thesis that agents *usually* have a reason to do the moral thing. If rational egoism leads us to this proposition, it comes close to what moral convention tells us, and thus comes close to capturing the virtue of intrinsic moral rationalism. For if morality usually indicates that agents ought to do that which is in their individual self-interest, and rational egoism is true, then most of the time, there will be a reason to do the moral thing. After I provide some grounds for believing that moral egoism is true, I will argue that even if moral egoism is not true, there are grounds for believing that normative morality usually indicates that agents ought to do that which is in their individual self-interest, and so it is likely that rational egoism comes close—perhaps close enough—to capturing the virtue of intrinsic moral rationalism.

Perhaps the worst case scenario for rational egoism is that it turns out that, quite often, normative morality requires agents to do that which is not in their individual self-interest. If this is the case, then rational egoism would lead us to the proposition that there is rarely a reason to do the moral thing, and this proposition does not match moral convention. If rational egoism leads us to this proposition, this might be considered a strike against rational egoism. I will argue, though, that this would not be a devastating strike against rational egoism, and that it would still be a theory about intrinsic reasons that is worthy of our consideration.

Now, I will say more about moral egoism and why we might think it is true. I must note that although I endorse moral egoism, I do not have space to present an all things considered

argument for it here. I will nonetheless attempt to provide some considerations in its favor.<sup>44</sup> To start, presumably, normative morality, even if it is primarily about how we should behave with regards to others, should at least have something to say about how we should behave with regards to ourselves, especially given that the way we behave with regards to ourselves is likely to have quite an impact on others. For example, a moral agent will be able to be more altruistic if he acts in a way towards himself that allows him to be a more altruistic person. We may think again of Jim, the agent who desires to drink a gin and tonic even though whenever he drinks one gin and tonic, he ends up drinking several more and getting into the worst sorts of trouble. Suppose that Jim is extraordinarily kind and has the most altruistic intentions. If he gives in to his desire for a gin and tonic on a regular basis, he will just wind up in the worse sorts of trouble, and will not be able to fulfill his altruistic goals. He will perhaps wind up in jails, detox centers, or places that he never intended to go. Rather than contributing to the welfare of others, he will likely be in a position of having to accept their aid. On the other hand, if Jim were to be a good moral egoist, and take the course of action that was in his self-interest, namely resisting his desire to drink that first gin and tonic, it is much more likely that he would be able to be kind, and fulfill his altruistic intentions. I take it as evidence in favor of moral egoism that it gives prominence to that part of morality that has to do with how agents behave with regards to themselves, especially because how agents behave with regards to themselves is quite connected to the impact they have on others. Just to clarify, I am claiming that, conventionally, many moral philosophers and others seem to think that morality is largely about how agents should act with

---

<sup>44</sup>Although I do not have space to develop or defend the following connection, I do think it is useful here to gesture at the similarity between what I call moral egoism and certain versions of virtue ethics, arguably held by Plato and Aristotle, which holds that morality just requires agents to do that which will allow them to flourish, or live the best possible lives. Aristotle, *The Nichomachean Ethics*, ed. Richard McKeon, *The Basic Works of Aristotle*, 2001 Modern Library Paperback Edition (New York: Random House, 1941) 935-1112; Plato, *Apology* and *Crito*, ed. John M. Cooper, *Plato: Complete Works* (Indianapolis: Hackett Publishing Company, 1997) 17-48.

regards to other agents. Interestingly though, how agents act with regards to others has quite a bit to do with how agents act with regards to themselves. And since moral egoism gives prominence to that part of morality that has to do with how agents behave with regards to themselves, instead of neglecting this part of morality as other normative moral theories might, we have a consideration, however small, in favor of moral egoism.

Another point in favor of moral egoism is that it might be important for a normative moral theory to account for the character and separateness of agents, and moral egoism, as a normative moral theory, does a fine job of this, especially in contrast with its competitor theories.<sup>45</sup> The fact is that agents are separate from one another and agents each have their own characters, and it seems that a normative moral theory should give prominence to these, what seem like moral, concerns. Suppose, for example, there is an agent, Tina, who loves to play music, and is a virtuoso musician. Her playing brings much meaning, pleasure, and benefit to the world. Her music is part of her character, and something that distinguishes her from others. It could be said that playing music is what she lives for. Indeed, it at least seems that it is in her self-interest to continue playing music, and moral egoism would require that she keep it up, as long as it is in her self-interest to keep it up. And I believe that this is a virtue of moral egoism. Suppose though, that as a matter of fact, Tina could do much more good overall by volunteering her time at soup kitchens and clinics all over the world. Some other normative moral theories that give prominence to impartiality, rather than the individual characters and separateness of agents, would require Tina to give up her music and volunteer her time at soup kitchens and clinics all over the world. I take this to be a flaw of these other normative moral theories, and correspondingly, a virtue of moral egoism that it would not require Tina to give up her music.

---

<sup>45</sup>In Williams, "Persons, Character, and Morality," 1-19, Williams makes the case that consequentialism and Kantian moral theory neglect the character and separateness of persons, and that this is a crucial flaw of these theories.

Perhaps some readers may be moved by my comments here; indeed they may agree that morality is largely about how agents should act with regards to themselves, and that the character and separateness of agents are important moral concerns. These readers may be drawn to moral egoism. On the other hand, some readers may still lean more strongly towards versions of normative morality that give prominence to impartiality. Such readers might believe that the moral thing is for Tina to volunteer at soup kitchens and clinics around the world. Although I am critical of normative moral theories that give perhaps prominence to impartiality and would require that Tina give up her music, I lack space to fully assess these sorts of views here. So now, I will argue that even if moral egoism is not true—and that perhaps some normative moral theory that gives prominence to impartiality is true—there are, nonetheless, grounds for believing that morality *usually* indicates that agents ought to do that which is in their individual self-interest, and so it is likely that rational egoism still comes close to capturing the virtue of intrinsic moral rationalism. I will proceed by discussing a few popular normative moral theories, and suggesting that each of these theories will usually, even if not always, require agents to do that which is in their individual self-interest. I must admit that there are many versions of each moral theory that I discuss and I do not have time to go into the details about each version of each moral theory. What I hope to get across is that there is generally much consensus among normative moral theorists about what the moral thing to do is, although not much consensus about what makes the moral thing to do the moral thing to do. I hope to provide some grounds for believing that most normative moral theories will share what I take to be the common sense notion that morality rarely requires agents to do that which is not in their individual self-interest.

First, let us consider consequentialism, the moral theory which holds that the rightness or wrongness of actions depends upon the value of their consequences.<sup>46</sup> So roughly, agents are required to do that which will bring about the consequences with the greatest overall value. There are grounds for thinking that agents, like Jim for example, who nurture their individual self-interest, that is, who always or usually act in accordance with their individual self-interest, will be in a better position to act in a way that will bring about the consequences with the greatest value.<sup>47</sup> If Jim does the thing that is in his self-interest, here supposing that it is in his self-interest to stay sober, he will become the sort of person who is able to fulfill his altruistic goals and bring much overall value to the world. On the other hand, it also seems to be true that, in some cases, consequentialist moral theory would require agents to do something that might not be in accordance with their self-interest in order to bring about the consequences with the greatest value. For example, consequentialism could be taken to require Tina to give up her music in order to work at soup kitchens and clinics around the world. Still, the point is that, most of the time, it seems that consequentialist moral theory will allow or even require agents to act in accordance with their individual self-interest because it is likely that agents who act in accordance with their individual self-interest are the kind of agents that are able to bring about the consequences with the greatest value. So, if consequentialist moral theory is true, there are grounds for believing that rational egoism will still match up well with our moral conventions although perhaps not as well as intrinsic moral rationalism; rational egoism will have it that, in most cases, there are reasons to do the moral thing, although in some cases in which morality

---

<sup>46</sup>Different versions of consequentialism give prominence to things like rules or character traits rather than actions. I focus on the version of consequentialism that gives prominence to actions.

<sup>47</sup>Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Ethical Theory 2: Theories About How We Should Live*. Oxford: Oxford University Press, 1998, 222-255.

requires of agents that they do things that are not in their self-interest, there will be a reason to refrain from doing the moral thing.

Next, let us consider Kantian moral theory. I will focus primarily on Kant's two formulations of the Categorical Imperative: the Formula of Universal Law and the Formula of the End in Itself. The Formula of Universal Law goes as follows: "Act only according to that maxim by which you can at the same time will that it should become a universal law."<sup>48</sup> I follow Korsgaard in my interpretation of the Formula of Universal Law.<sup>49</sup> Korsgaard holds that the Formula of Universal Law requires agents to act only on maxims that they can *without contradiction* will to be universal laws. In other words, by Korsgaard's interpretation, the moral requirement is that agents act only according to maxims that would not become self-defeating if universalized. Another way of saying this is that any action that would become ineffectual for the achievement of the agent's purpose, if everyone tried to use the same action for the same purpose, is immoral. An example of an immoral thing is a false promise.<sup>50</sup> Suppose a man in financial difficulties borrows a sum of money with the intention of never paying it back. By borrowing a sum of money with the intention of never paying it back, the man would be acting on the following maxim or one very similar to it: anyone in need can promise what he pleases with the intention of not fulfilling the promise in order to get what he needs. However, if this maxim were made into universal law, no one would believe the promises of others, and so the agent's action and his purpose behind it would become ineffectual. Nobody would believe his

---

<sup>48</sup>Immanuel Kant, *Foundations of the Metaphysics of Morals*, trans. Lewis White Beck, Library of Liberal Arts: 1956 quoted in Christine M. Korsgaard, "Kant's Formula of Universal Law," *Pacific Philosophical Quarterly* 66, no. 1-2 (1985): 24.

<sup>49</sup>Christine M. Korsgaard, "Kant's Formula of Universal Law," *Pacific Philosophical Quarterly* 66, no. 1-2 (1985): 24-47.

<sup>50</sup>*Ibid.*, 20.

promise, nobody would give him a loan, and so he would not succeed at borrowing a some of money with the purpose of never paying it back in order to escape his financial difficulties.

Next, let us take a look at Kant's Formula of the End in Itself, which requires that agents act only on maxims that "treat humanity whether in your own person or in the person of any other never simply as means, but always at the same time as an end."<sup>51</sup> I follow O'Neill in my interpretation of the Formula of the End in Itself.<sup>52</sup> O'Neill holds that the Formula of the End in Itself requires that agents act only on maxims that do not involve using agents as mere means, and that additionally, treat agents as ends in themselves. For O'Neill, not using agents as means requires providing others the opportunity to consent or dissent to how they are being treated. And treating agents as ends in themselves requires neither disregarding, taking over, or lending no support to their ends. So, the Formula of the End in Itself can be understood in terms of two requirements. The first requirement is to refrain from acting on maxims that treat agents as mere means, or in other words, to refrain from acting on maxims that do not provide agents the opportunity to consent to or dissent to how they are being treated. The second requirement is to refrain from acting on maxims that do not treat others only as ends in themselves, or in other words, to refrain from acting on maxims that disregard, take over, or lend no support to the ends of others.

So now, given a rough sketch of the requirements of Kantian morality, we can ask the following question: will Kantian morality require, or at least allow, agents to act in ways that are in their individual self-interest, most of the time? More specifically, will it turn out to be in an

---

<sup>51</sup>Immanuel Kant, *Groundwork*, trans. H. J. Paton, as *The Moral Law* (London: Hutchinson), 1953, Prussian Academy Pagination quoted in Onora O'Neill, "Between Consenting Adults," *Philosophy and Public Affairs*, Volume 14, No. 3, (Summer, 1985): 262.

<sup>52</sup>Onora O'Neill, "Between Consenting Adults," *Philosophy and Public Affairs*, Volume 14, No. 3, (Summer, 1985): 252-277.

agent's self-interest to act only on maxims that can, without becoming self-defeating, be universalized? And will it turn out to be in an agent's self-interest to act only on maxims that provide others the chance to consent or dissent, and that neither disregard, take over, or lend no support to the ends of others? Indeed, the following seems to be true: it would be in an agent's self-interest to be the sort of person who acts only on maxims that can, without becoming self-defeating, be universalized. For the sort of maxims that cannot be universalized, as Korsgaard points out, usually involve unfairness, deception, or cheating.<sup>53</sup> Generally speaking, it seems that, most of the time, it is not in an agent's self-interest to be unfair, a deceiver, or a cheater. On the other hand, in some cases, it might be in an agent's self-interest to be unfair, a deceiver, or a cheater. At this point, it helps to return to the account of self-interest that we are working with. An agent's self-interest is that which an idealized version of the agent would desire for the actual agent were the idealized agent to step into the situation and circumstances of the actual agent. The idealized agent has full knowledge of the actual agent, his environment, and the objects of the actual agent's desires, and perfect rational deliberation skills. Would the idealized agent desire that the actual version of himself deceive or cheat? Without begging the question about the idealized agent's honesty, I nonetheless think that the idealized agent would usually not desire that the actual agent deceive or cheat. And here is why: with full knowledge and perfect rational deliberation skills, the idealized agent would likely know of other, more honest ways to achieve his ends, and thus he would not desire to place the actual agent in a position of having to face the negative practical consequences, or the future frustration of his ends that could come along with deceiving and cheating others. I do grant that we could probably imagine special cases in which the idealized agent would desire for the actual agent to deceive or cheat. My point is just that it seems that the idealized agent with full information and rational deliberation skills would usually

---

<sup>53</sup>Korsgaard, "Kant's Formula of Universal Law," 44.

not desire for the actual agent to cheat, not because the idealized agent is honest, but out of prudence, or the lack of willingness to face the potential negative practical consequences, or the possible future frustration of ends, involved in cheating and deceiving others. Thus, based upon our considerations about the idealized agent, there are grounds for believing that it would usually turn out to be in an agent's self-interest to do what is required by the Formula of Universal Law.

There are similar grounds for believing that it would usually turn out to be in an agent's self-interest to do what is required by the Formula of the End in Itself. Roughly, the Formula of the End in Itself forbids agents from acting on maxims that either use agents or do not treat them as ends in themselves. Out of prudence, we can imagine the ideal agent not desiring that the actual agent act on maxims that either use agents or do not treat them as ends in themselves. With full knowledge and perfect rational deliberation skills, it is likely that the idealized agent would know of ways that the actual agent could achieve his ends without ever having to act on maxims that either use agents or do not treat them as ends in themselves. I do grant that we could probably imagine special cases in which the idealized agent would desire that the actual agent act on a maxim that either uses agents or does not treat them as ends in themselves. My point is just that it seems that the idealized agent with full information and rational deliberation skills would usually not desire that the actual agent act on maxims that use agents or do not treat them as ends in themselves, not because he is respectful or kind, but out of prudence, or the lack of willingness to face the potential negative practical consequences, or potential frustration of his ends, involved in using others and not treating them as ends in themselves. Thus, based upon our considerations about the idealized agent, there are grounds for believing that it would usually turn out to be in an agent's self-interest to do what is required by the Formula of the End in Itself. So, if Kantian moral theory is true, there are grounds for believing that rational egoism will still

match up well with our moral conventions although perhaps not as well as intrinsic moral rationalism; rational egoism will have it that, in most cases, there are reasons to do the moral thing, although in some cases in which morality requires of agents they they do things that are not in their self-interest, there will be a reason to refrain from doing the moral thing.

Let us now consider one last normative moral theory, virtue ethics. Virtue ethics holds that morality requires agents to act virtuously, cultivate certain moral virtues, or to live in a way that will allow them to flourish. Interestingly, I believe that moral egoism would qualify as a version of virtue ethics. Following this line of thought, we may think of moral egoism as a version of virtue ethics that provides *self-interest* as an account of what it means for an agent to flourish. But I will continue this line of thought no further here. In what follows, I am going to suppose that moral egoism is not a version of virtue ethics, and that *self-interest* and *flourishing* are two different concepts. I do this so that we can consider what other versions of virtue ethics may require of agents. Virtue ethicists may disagree about which traits or characteristics are to count as moral virtues, and what it means for an agent to flourish. They may also disagree about whether it is the virtues or the concept of flourishing that deserves prominence. I do not have space to get into these disagreements, or to discuss in any detail the virtues, or the concept of flourishing. My goal here is just to provide some grounds for believing that any normative moral theory that holds that morality requires agents to act virtuously, cultivate certain moral virtues, or to live in a way that will allow them to flourish will usually require, or at least allow, agents to do that which is in their individual self-interest. Since virtue ethics gives prominence to moral virtue or flourishing—which I am supposing is something different than self-interest—virtue ethics could require agents to do that which is not in their self-interest, specifically when cultivating some moral virtue or doing that which will allow them to flourish does not turn out to

be in their individual self-interest. Nonetheless, there are grounds for believing that agents who nurture their individual self-interest, that is, who always or usually act in accordance with their individual self-interest, will be virtuous people doing virtuous things, and flourishing. For example, we may imagine Jim cultivating the virtue of temperance; it seems very likely that this is what would help him to flourish, and it seems that this is what it would be in his self-interest to do. Generally, it seems that it is in the self-interest of agents to do virtuous things, to refrain from doing vicious things, and to do that which will help them flourish. So it seems that virtue ethics and moral egoism will generally require the same sorts of things of agents. However, we can imagine that, in some cases, it might not be in an agent's self-interest to do the morally virtuous thing or to do what is required to flourish. On some accounts, it might be the case that the morally virtuous thing or what is required to flourish might be for agents like Tina to give up their personal projects like music, and instead, volunteer at soup kitchens and clinics around the world. So, if virtue ethics is true, there are grounds for believing that rational egoism will still match up well with our moral conventions although perhaps not as well as intrinsic moral rationalism; rational egoism will have it that, in most cases, there is a reason to do the moral thing, although in the cases in which morality requires of agents they they do things that are not in their self-interest, there will be a reason to refrain from doing the moral thing.

I will close this chapter section by considering the worst case scenario for rational egoism: the possibility that, quite often, morality requires agents to do that which is not in their individual self-interest. Supposing that morality indeed requires, most of the time, that agents do that which is not in their individual self-interest, then the rational egoist will be committed to the proposition that there is rarely a reason to do the moral thing, and this is quite contrary to moral convention. But just what does this mean for rational egoism? I think it can be taken as a strike

against rational egoism; many times, when a theory does not fit well with our conventions, we take that as a strike against the theory. However, convention can be mistaken. It is possible that, although it seems that there is necessarily a reason to do the moral thing, the fact of the matter might be that there is rarely a reason to do the moral thing.<sup>54</sup> If rational egoism does not fit well with moral convention, this might turn out to be more of a strike against moral convention than it is against rational egoism. This especially seems true when we think about how well rational egoism fares when compared to its competitor theories intrinsic moral rationalism and the desire-satisfaction view. At best, rational egoism gets around the problems that intrinsic moral rationalism and the desire-satisfaction views are vulnerable to. At worst, rational egoism is troubled by the problems of its competitor theories, but is no worse off.<sup>55</sup> Surely, that rational egoism might not fit with moral convention is not grounds for rejecting it outright. I will grant, though, that maybe rational egoism will not match our moral conventions, and this *could* be taken to count as a strike against rational egoism.

In this chapter section, I have presented three possibilities: either (1) moral egoism is true which means that morality always requires agents to do that which is in their self-interest, (2) morality usually requires or at least allows agents to do that which is in their individual self-interest, or (3) morality rarely allows agents to do that which is in their individual self-interest. If (1) is true, then rational egoism can capture the virtue of intrinsic moral rationalism. If (2) is true, then rational egoism will lead to the proposition there is usually a reason to do the moral thing, and this is close to our moral convention. If (3) is true, then rational egoism will lead to

---

<sup>54</sup>Susan Wolf, "Moral Saints," *The Journal of Philosophy* 79, no. 8, (August, 1982): 419-439. Williams, "Internal and External Reasons," 101-113.

<sup>55</sup>Maybe rational egoism would be worse off than intrinsic moral rationalism at least in one sense, if rational egoism did not turn out to fit well with moral convention, but rational egoism is surely no worse off than intrinsic moral rationalism with regards to the mystery problem.

the proposition that there is rarely a reason to do the moral thing, and this does not fit with moral convention. This might count as a strike against rational egoism, but not a devastating one. In this chapter section, I have presented some grounds for believing that either (1) or (2) is true. And I have argued that even if (3) is true, it would not be a devastating problem for rational egoism. I believe (1) is true because, although I was not able to provide an all things considered defense of it here, I strongly suspect that moral egoism is true.

### 3.8 Conclusion

In Chapter 3, I made a case in favor of rational egoism by presenting some common sense grounds for believing that it is true, and then by arguing that it fares better, or no worse, than its major competitor theories with regards to both their theoretical merits, and the problems that they are vulnerable to. The conclusion I reach with regards to rational egoism is that, at best, we have solid grounds for believing that it is true; at worst, we have grounds for keeping it on the table with its competitor theories for further consideration.

But where does this leave us in terms of the metamoral question that I raised at the beginning of chapter 1? Is there necessarily a reason to do the moral thing, and if so, what is that reason? I have made a case that it is a particular version of extrinsic moral rationalism that seems to provide the best answer to this fundamental question. Indeed, the version of extrinsic moral rationalism that I have endorsed goes as follows: there is necessarily a reason to do the moral thing because there is necessarily a reason to do what is in one's self-interest, and there is a necessary connection between morality and self-interest. But what sort of case have I made in support of this view?

First, I presented a preliminary argument in favor of moral rationalism unqualified, but then I argued that we should dismiss one particular version of moral rationalism, namely intrinsic moral rationalism. These arguments together amount to a preliminary case in favor of extrinsic moral rationalism. Then, in chapter 2, I argued against the desire-satisfaction view. In that the most plausible versions of extrinsic moral rationalism seem to be committed to either the desire-satisfaction view or rational egoism, my arguments from chapter 1 with my argument against the desire-satisfaction view in chapter 2 would provide strong grounds for considering rational egoism, which I defended in chapter 3. I am confident that, in chapter 3, I, at the least, showed that rational egoism provides a worthy account of what it is that necessarily provides agents with reasons to act. I also provided some considerations in favor of moral egoism, the normative moral theory which confers a necessary connection between morality and self-interest. All in all, although I was not able to provide an all things considered, airtight argument for my version of extrinsic moral rationalism, I think I have done enough to get it on the table, and I think that just getting it on the table is a significant endeavor. In chapter 3, I also argued that even if moral egoism is not true, morality usually indicates that agents should do that which is in their self-interest. If moral egoism is not true, then my version of extrinsic of extrinsic moral rationalism would turn out to be false, and a version of moral antirationalism would be true. But I have argued that this would be a version of moral antirationalism that gets us close to moral rationalism; indeed, it would be a version of moral antirationalism which has it that there is usually a reason to do the moral thing.

Closing now, I strongly suspect that if rational egoism is true, and I have presented grounds for believing that it is, then there is always, or almost always, a reason to do the moral thing. On one normative moral theory, namely moral egoism, morality just requires agents to do

that which is in their self-interest. If moral egoism is true, and there are indeed grounds for believing that it is, and rational egoism is true, then there will always be a reason to do the moral thing. But even if moral egoism is not true, I have argued that there are solid grounds for believing that morality usually indicates that agents should do that which is in their self-interest. In this case, if rational egoism is true, there will usually be a reason to do the moral thing.