

WANTING THE BAD AND DOING BAD THINGS: AN ESSAY IN MORAL  
PSYCHOLOGY

By

PETER BRIAN BARRY

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2005

Copyright 2005

by

Peter Brian Barry

## ACKNOWLEDGMENTS

I am grateful to a number of people for their contributions to this dissertation. My two greatest debts are owed to David Copp and Anton Tupa. As my supervisory committee chair, David faced the formidable task of reading and commenting on dozens of tortured drafts. His timely support and encouragement and wise counsel are greatly appreciated and I have learned much from him. Anton too helped guide earlier drafts into this present form. I only hope that he has enjoyed and benefited from working with me half as much as I have.

Marina Oshana, Crystal Thorpe, and Dolores Albarracin all capably served as members of my supervisory committee. Marina and Crystal both provided crucial comments and suggestions for improvement. Dolores tolerated a layman's forays and her very presence on my committee forced me to look for empirical confirmation of my theses. I am also grateful to members of the philosophy departments at the University of Florida and Bowling Green State University, especially John Biro, Kirk Ludwig, Chuang Lui, Greg Ray, David Sobel, Jon Tresan, and Gene Witmer.

I also record my debt to the late Michael Robins, who served as my original dissertation advisor and sadly passed away while my dissertation was in its infancy. Mike is responsible for initiating and cultivating my interests in the philosophy of action and moral psychology. I fear that this dissertation barely resembles the project Mike briefly oversaw and that Mike may have less sympathy with my present views. I wish that I had a chance to argue with such a formidable philosopher again. He is missed.

## TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS .....	iii
ABSTRACT.....	vi
CHAPTER	
1 THE IMP OF THE PERVERSE.....	1
Introduction.....	1
Wanting the Bad and Desirability.....	10
What Intentionally Pursuing the Bad Is . . . and What it Is Not .....	16
Intentionally Pursuing the Bad and Akratic Action.....	29
Conclusion .....	34
2 DESIRE AND DESIRING THE BAD.....	35
Introduction.....	35
Humean Accounts of Desire.....	37
Nagel and Motivated Desires.....	43
Hedonic Accounts of Desire.....	47
Thomism about Desire.....	50
The Master Argument.....	64
Desiring the Bad and Intelligibility .....	71
Conclusion .....	77
3 INTENTION AND INTENDING THE BAD.....	78
Introduction.....	78
The Function of Intention and the Nature of Intention.....	79
Desire-Belief Reductions of Intention .....	84
Intentions as Evaluative Judgments.....	90
Intentions and Decisions to Act.....	96
Whither Rationality?.....	101
Conclusion .....	105

4	ACTING FOR REASONS AND REASONS FOR INTENTIONALLY PURSUING THE BAD .....	109
	Introduction.....	109
	The Problematic Syllogism.....	111
	A Taxonomy of Reasons .....	115
	‘Reason’ and Equivocation.....	122
	The Problematic Syllogism Redux .....	128
	Conclusion .....	138
5	UNDERSTANDING, CONTROL, AND INTENTIONALLY PURSUING THE BAD.....	143
	Introduction.....	143
	Goals and Having Goals .....	145
	Goal-Directed Behavior and Intentional Action.....	152
	Full-Blooded Action, Control, and Conscious Direction .....	155
	Practical Reasoning .....	163
	Passive Cases of Treating as a Reason and Pursuing the Bad.....	166
	Active Treating as a Reason and Pursuing the Bad.....	170
	Principled Decisions to Pursue the Bad.....	175
	Resisting Raz’s Challenge .....	179
	Conclusion .....	182
	LIST OF REFERENCES .....	183
	BIOGRAPHICAL SKETCH .....	193

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

WANTING THE BAD AND DOING BAD THINGS: AN ESSAY IN MORAL  
PSYCHOLOGY

By

Peter Brian Barry

August 2005

Chair: David Copp  
Major Department: Philosophy

In this dissertation, I argue that it is possible to desire what is believed to be bad and not at all good, to intentionally pursue what is believed to be bad and not at all good, and that the believed badness of an action can be an agent's reason for acting. I refer to these theses as the *desirebad*, *intentionbad*, and *reasonbad* theses, respectively.

I consider various accounts of desire and I argue that all but one is consistent with the *desirebad* thesis. The account that is inconsistent, what I call the Thomist account, implies that believing that the object of desire is good is necessary for desiring it. However, the Thomist account depends on dubious assumptions about the nature of desire and what it is to act intelligibly. I also consider various accounts of intention and I argue that the *intentionbad* thesis is consistent with all of them, including accounts that identify intentions with a kind of evaluation and the outcome of a kind of decision.

I then consider objections to my three theses. For example, it is plausible to suppose that intentional action is performed for reasons, but it seems that the intentional

pursuit of the bad cannot be action performed for reasons. This objection fails because it fails to distinguish different kinds of reasons; in particular, it fails to distinguish *normative* reasons and an *agent's* reason for acting. Here is another objection. Some philosophers claim that agents who intentionally pursue the bad cannot be in control of or understand themselves and what they do. I consider various ways that agents can possess control and understanding; for example, by having goals and performing goal-directed behavior, by deliberately forming and executing intentions, and by treating considerations as reasons. I argue that at least some agents who intentionally pursue the bad possess control and understanding, just as agents who act intentionally do.

In responding to these objections, I develop a positive account of intentional action and agency that is inclusive enough to explain non-standard agents and their actions.

## CHAPTER 1 THE IMP OF THE PERVERSE

We stand upon the brink of a precipice. We peer into the abyss—we grow sick and dizzy. . . . there grows into palpability, a shape, far more terrible than any genie, or any demon of a tale, and yet it is but a thought, although a fearful one, and one which chills the very marrow of our bones with the fierceness of the delight of its horror. It is merely the idea of what would be our sensations during the sweeping precipitancy of a fall from such a height. . . for the very reason that involves that one most ghastly and loathsome of all the most ghastly and loathsome of images of death and suffering which have ever presented themselves to our imagination—for *this very cause* do we now the most vividly want it.

Edgar Allen Poe, “The Imp of the Perverse”<sup>1</sup>

### Introduction

In some ways, Poe’s precipice dweller is no different from the standard sort of agent discussed by moral psychologists and philosophers of action: there is some action that he wants to perform and he wants to perform the action because something about it appeals to him. According to what I shall refer to as *standard accounts* of moral psychology and action, when an agent wants to perform some action, she believes that her action would be good to perform and wants to perform the action because she believes her action would be good; further, according to standard accounts, given that an agent wants something she believes to be good, the agent can proceed to act intentionally, supposing the opportunity comes up to get what she wants. But Poe’s precipice dweller is not the kind of agent that fits comfortably into standard accounts. Poe’s precipice dweller wants to fall to the rocks below, but he does not want to fall because of any

---

<sup>1</sup> Edgar Allen Poe, “The Imp of the Perverse” from *The Complete Tales and Poems of Edgar Allen Poe*, p. 282.

believed good he associates with falling. Poe's precipice dweller associates fearful and ghastly and loathsome images of death and suffering with falling from the precipice, and, in Poe's words, "for that very cause" wants to fall.

Proponents of standard accounts will insist that it is misleading to claim that the believed badness of falling to the rocks below is "the very cause", or counts as the reason, that leads the precipice dweller to want to fall. The precipice dweller surely finds images of falling to be fearful and ghastly and loathsome, claims the proponent of a standard account, but what leads him to want to fall is, for example, the anticipated rush of excitement associated with falling. Perhaps he enjoys the prospect of falling in the same way that many persons enjoy scary movies or roller coasters. I shall discuss this response further in what follows, but note that Poe's account of the precipice dweller includes *nothing* that independently suggests that the precipice dweller believes there is something good about falling to the rocks below. Poe's precipice dweller is motivated perversely, insofar as he is motivated to perform an action that deviates from what he believes is good to do. Standard accounts are forced to explain away or deny the seeming perversity of the precipice dweller. The interesting question is not whether the precipice dweller, as Poe presents him and I describe him, can be re-described in a way that explains away his perversity. The interesting question is whether the precipice dweller, as Poe presents him and I describe him, is possible.

In what follows, I shall argue for and defend three distinct but related theses. The first thesis has already been alluded to: agents can desire what they believe to be bad without also believing the thing they desire is good. I shall refer to instances where an agent desires something she believes to be bad and to be not at all good as instances of

wanting the bad or, alternatively, as instances of *desiring the bad*. I shall refer to the thesis that agents can desire the bad as the *desirebad* thesis. For future reference, my claims that the *desirebad* thesis is true and that wanting and desiring the bad are possible are simply shorthand for the thesis that it is possible for agents to desire what they believe to be bad and to be not at all good.

It will be useful to have a name for proponents of standard accounts who reject some or all of my three theses: I shall refer to proponents of standard accounts who reject some or all of my three theses as ‘Panglossians.’ The name ‘Panglossian’ is derived from Voltaire’s satirical representation of Leibniz in *Candide*. Familiarly, Leibniz claims that this world is in fact the best of all possible worlds God could have created in attempt to reconcile the omnipotence, omniscience and goodness of God with the appearance that there is unnecessary suffering. Voltaire mocks Leibniz’s theodicy by repeatedly subjecting Pangloss to undeserved suffering while having Pangloss consistently repeat a variant of Leibniz’s thesis that “all is for the best in this the best of all possible worlds,” even while things could seemingly not be much worse:

Candide came up, saw his benefactor reappear for a moment and then be engulfed forever. He tried to throw himself after him into the sea; he was prevented by the philosopher Pangloss, who proved to him that the Lisbon roads had been expressly created for the Anabaptist to be drowned in them. While he was proving this *a priori*, the vessel sank, and everyone perished except Pangloss, Candide, and the brutal sailor who had drowned the virtuous Anabaptist. . .<sup>2</sup>

My Panglossian shares at least two features with Dr. Pangloss. First, my Panglossian suffers from a certain amount of undue optimism that “all is for the best,” that agents always desire what they believe to be good, for example. Second, my Panglossian typically asserts *a priori* that desiring the bad is impossible, despite of evidence to the

---

<sup>2</sup> Voltaire, *Candide*, p. 20.

contrary, just as Dr. Pangloss continues to assert that “all’s for the best” when there is ample evidence to the contrary.

I shall refer to the arguments offered by proponents of standard accounts that challenge the truth of my three theses as “Panglossian challenges.” Panglossian challenges come in at least two varieties: weak and strong. Weak Panglossian challenges claim that explanations of action that appeal to the believed goodness of the action are always more plausible than explanations of action that appeal to the believed badness of the action. If it can be shown that it is more plausible to suppose that an agent would desire something because she believes the object of her desire is good, rather than to suppose that she desired what she believes is bad and not at all good, then there would be some reason to doubt the truth of the *desirebad* thesis, for example. My general response to weak Panglossian challenges is that the moral psychology of the Panglossian is not as rich and explanatorily potent as the moral psychology of the proponent of the thesis that the intentional pursuit is possible.<sup>3</sup> Strong Panglossian challenges purport to establish that my three theses are false by defending some conceptual thesis concerning desire or action that would imply that some or all of my three theses could not be true. I shall consider various strong and weak Panglossian challenges in subsequent chapters.

Historically, many philosophers have been Panglossians insofar as they have rejected some or all of my three theses. The *desirebad* thesis, for example, is not widely

---

<sup>3</sup> Similarly, I think there is no knockdown argument that explanations of allegedly altruistic action in terms of self-interest are mistaken; nonetheless, I find it plausible that a moral psychology that denies the thesis of psychological egoism, for example, and permits genuine altruistic actions is a much more compelling moral psychology.

defended.<sup>4</sup> Plato explicitly rejects the possibility of desiring the bad.<sup>5</sup> Aquinas claims “people desire things because they think them good” and if Aquinas means propose a necessary condition for desiring, then he is committed to denying that desiring the bad is possible.<sup>6</sup> Other philosophers come very close to rejecting the *desirebad* thesis. Sidgwick understands desires as an “impulse” directed “towards the realization of some positive future result.”<sup>7</sup> Locke understands desire as “the uneasiness a man finds in himself upon the absence of anything whose present enjoyment carries the idea of delight with it.”<sup>8</sup> Spinoza would deny that Poe’s precipice dweller could rightly be said to desire to fall to the rocks below given Spinoza defines desire as an appetite that is “the very essence of man in so far as his essence is determined to such actions as contribute to his preservation.”<sup>9</sup> Since falling to the rocks below does not contribute to an agent’s preservation, then the precipice dweller could not desire to fall on Spinoza’s account. And as I note in Chapter 4, a number of contemporary philosophers have explicitly or implicitly rejected *desirebad* thesis.

However, there is anecdotal evidence that wanting the bad is possible. It is perhaps common to be tempted to lean too far over the edge of a tall building or to wonder what it would be like to quickly steer one’s car into the median on the expressway. Women

---

<sup>4</sup> For notable exceptions, see Michael Stocker, “Desiring the Bad; An Essay in Moral Psychology”, *Journal of Philosophy* 76 (1979), pp. 738-53 and David Velleman, “The Guise of the Good”, reprinted in his *The Possibility of Practical Reason*, pp. 99-122.

<sup>5</sup> Plato, *Meno* 77c-78b, in John M. Cooper, ed., *Plato: The Complete Works*.

<sup>6</sup> Aquinas: *Selected Philosophical Writings*, Timothy McDermott, ed., p. 340.

<sup>7</sup> Henry Sidgwick, *The Methods of Ethics*, p. 46.

<sup>8</sup> John Locke, *An Essay Concerning Human Understanding*, A. D. Wood, ed., p. 160.

<sup>9</sup> Spinoza, *Ethics* (Hackett: 1992), Seymour Feldman, ed., p. 141.

suffering from postpartum depression often enough have disturbing and aggressive desires to harm their newborn children.<sup>10</sup> There are odd cases of agents who want to eat bizarre items such as dirt or carpentry nails or cigarette butts, and so forth; these are instances of the eating disorder, pica, an abnormal craving for the ingestion of non-food items.<sup>11</sup> There is some reason to think the origins of pica lie in some sort of deficiency in a person's diet. We might *make* instances of pica comprehensible to an agent by explaining the etiology of her bizarre desire, but when she is motivated to eat dirt or cigarette butts she surely does not believe there is anything good about eating dirt and cigarette butts and she very likely believes that eating dirt and cigarette butts is bad. Even the most virtuous and pious of us have perhaps desired the bad. Former President of the United States Jimmy Carter, a deeply religious man and former Sunday school teacher, publicly admitted to having committed "adultery in his heart" insofar as he sometimes looked upon women with lust. It may very well be that many agents think that there is something good about lustful leerings insofar as lustful leerings are believed to be pleasurable. Given Carter's piety, it is plausible to suppose that Carter believes that the pleasure of looking upon women with lust is bad and not at all good.<sup>12</sup> Protestant reformist Martin Luther was tormented by urges to curse Christ while praying."<sup>13</sup> Given

---

<sup>10</sup> One study conducted out of the Women's Mood Disorders Clinic in Cleveland found that 54% of women suffering from postpartum depression admitted to having such thoughts, as opposed to 21% of those with non-postpartum onset of depression; another study at the Western Psychiatric Institute and Clinic in Pittsburgh found that 41% of surveyed women admitted to having such thoughts. See Lee Baer, *The Imp of the Mind*, pp. 20-3 for discussion.

<sup>11</sup> The word 'pica' is derived from the Latin word for a magpie, a bird that will eat just about anything.

<sup>12</sup> Gary Watson discusses a case of a man who thinks his sexual inclinations are the work of a devil and believes there is nothing even *prima facie* good about his sexual desires; see Watson, "Free Agency," reprinted in *Free Will*, ed. Watson, p. 101.

<sup>13</sup> Erik H. Erickson, *Young Man Luther: A Study in Psychoanalysis and History*.

Luther's faith, it is hard to believe that Luther thought there is anything good about cursing Christ.

The second thesis I argue for is related to, but distinct from, the *desirebad* thesis: agents can act intentionally when they pursue what they believe to be bad and not at all good.<sup>14</sup> I shall refer to this thesis as the *intentionbad* thesis. Suppose that the precipice dweller desires the bad insofar as he desires to fall to the rocks below and believes that falling would be bad and not at all good. Suppose he also engages in a piece of instrumental reasoning and concludes that jumping from the cliff is an effective means for falling to the rocks below. Although the precipice dweller believes that jumping from the cliff realizes something he believes to be bad and not at all good, the precipice dweller can nonetheless act intentionally by jumping from the cliff. Or so I shall argue. I shall refer to instances in which an agent acts intentionally while believing the object of her pursuit is bad and not at all good as instances of *intentionally pursuing the bad*.

Panglossians are committed to denying the *intentionbad* thesis just as they are committed to denying the *desirebad* thesis. Any number of philosophers have rejected, or nearly rejected, the *intentionbad* thesis. Plato famously argues that no one ever willingly or knowingly chooses a lesser good over a greater good<sup>15</sup> and less famously that no one would ever deliberately embrace any of the supreme evils and that no one would voluntarily allow his soul to be corrupted.<sup>16</sup> Aristotle claims that all action aims at some

---

<sup>14</sup> Watson seems to allow that an agent can act for a desire for what she believes is bad and not at all good, but Watson denies that those actions are free; see Watson, *ibid.*, pp. 96-110, and Watson's discussion of the kleptomaniac, p. 110 in particular. I do not know if Watson endorses the *intentionbad* thesis, since the question of whether an act of pursuing the bad is intentional is different from the question of whether an act of pursuing the bad is free.

<sup>15</sup> Plato, *Protagoras* 351d-358d, in Cooper, *ibid.*

<sup>16</sup> Plato, *Laws* 731c, in Cooper, *ibid.*

good; if Aristotle's dictum implies that to act intentionally an agent must believe that what she aims at is good, then Aristotle is committed to denying the *intentionbad* thesis.<sup>17</sup> Kant attributes the thesis that "we avoid what we conceive to be bad" to Scholastic philosophers; supposing that Kant's characterization of the Scholastics is correct, then it seems that Scholastic philosophers would deny we could pursue what we conceive to be bad.<sup>18</sup> Aquinas claims that the will of human beings is ordered to seek bliss, that bliss is the goal of human life and that no one wills unhappiness or evil except accidentally by mistaking what is good for what is bad.<sup>19</sup> Leibniz claims that God has decreed that the will shall always seek the apparent good.<sup>20</sup> And many contemporary philosophers are committed to denying the *intentionbad* thesis.<sup>21</sup>

The third thesis I argue for I shall refer to as the *reasonbad* thesis. The *reasonbad* thesis implies that the believed badness of an action can be an agent's reason for acting. Poe comes remarkably close to asserting the *reasonbad* thesis when he asserts the following:

[There is] an innate and primitive principle of human action, a paradoxical something which we may call *Perverseness*, for want of a more characteristic term. . . Through its promptings we act, *for the reason that we should not*. . . this

---

<sup>17</sup> Aristotle, *The Nicomachean Ethics* 1094a2.

<sup>18</sup> Kant, *Critique of Practical Reason*, 5:59, Mary Gregor, trans.

<sup>19</sup> See Aquinas, *ibid.*, p. 338-40.

<sup>20</sup> Leibniz, *Discourse on Metaphysics*, p.49.

<sup>21</sup> See Elizabeth Anscombe, *Intention*, pp. 73-4; Dennis Stampe, "The Authority of Desire", *The Philosophical Review* 96, pp. 355-81; James Griffin, *Value Judgment*, pp. 32-3; Ronald De Sousa, "The Good and the True", *Mind* 83, pp. 534-51; David Pears, *Motivated Irrationality*, p. 198; Joseph Raz, "Agency, Reason, and the Good" from his *Engaging Reason: On the Theory of Value and Action* and "On the Moral Point of View" from *Reason, Ethics, and Society: Essays in Honor of Kurt Baier*, J. B. Schneewind, ed., pp. 70-2; Sergio Tenenbaum, "Accidie, Evaluation, and Motivation" in *Weakness of the Will and Practical Irrationality*, Stroud and Tappolet, eds., pp. 147-71.

overwhelming tendency *to do wrong for the wrong's sake*, [will not] admit of analysis, or resolution into ulterior elements.<sup>22</sup>

Augustine also appears to be sympathetic with the *reasonbad* thesis in his memorable account of stealing pears from the tree of a neighbor's orchard.<sup>23</sup> Augustine notes that we typically suppose that the *reasonbad* thesis is false when we seek to explain an action:

When we investigate a crime and ask for what reason was it committed, it is generally thought that there must be some obvious motive; either the hope of gaining or the fear of losing. . . Not even Cataline himself, then, loved the crimes he committed but the things he hoped to gain by them; which is something else entirely. (2.5.11)

However, in his own case, Augustine is unable to find any obvious motive indicative of gaining or losing. It was not poverty that led Augustine to steal, for what he stole he had in abundance. Augustine did not steal to enjoy the pears, for they were “not especially appealing to the eye or the tongue” and, at any rate, he later threw them away. Augustine did not steal to enjoy the company of his fellow hoodlums, for their company was nothing to him. Instead, Augustine claims what explains his theft of the pears was that he regarded the theft to be bad, that his act of stealing enticed him *because* it was an act of stealing, and claims that “there was no reason for my evil save evil itself.” (2.4.9) Augustine's testimony suggests that he did not act despite his belief that his action was bad and not at all good; rather, Augustine acted *just because* and *on the grounds that* he believed his action was bad and not at all good.<sup>24</sup>

---

<sup>22</sup> Poe, *ibid.*, p. 281, emphasis added.

<sup>23</sup> See Augustine's *The Confessions*, Book 2. For further discussion, see Gareth Matthews, *Augustine*, pp. 115-24. Thanks to Matthews for providing me with an advance draft of this chapter.

<sup>24</sup> Ultimately, Augustine seems puzzled by his own account: “Who can unravel this twisted bundle of knots and tangles? It is repugnant; I do not wish to see it or think of it.” (2.10.18)

The project undertaken in this chapter is to make sense of, and defend in a preliminary way, the *desirebad* and *intentionbad* and *reasonbad* theses. In Chapter 2, I argue that our best accounts of desire do permit the possibility of desiring the bad and I consider arguments against the *desirebad* thesis. In Chapter 3, I offer and argue for an account of intention that helps to explain how agents intentionally pursue the bad. In Chapter 4, I consider and reject various arguments against the *intentionbad* and *reasonbad* theses. In Chapter 5, I explain how agents who intentionally pursue the bad can control and understand themselves and what they do. Initially, I want to respond to a potential objection that threatens the plausibility of the theses I defend.

### **Wanting the Bad and Desirability**

One reason that some proponents of standard accounts have rejected any or all of the three theses I defend is that they believe that not just anything can genuinely be wanted. Elizabeth Anscombe implicitly argues against the *desirebad* thesis insofar as she argues that that not just anything can be wanted and that only what is conceived of in some way as desirable can genuinely be desired. Several proponents of standard accounts have appealed to Anscombe's arguments in rejecting any and all three of my theses.<sup>25</sup> Since Anscombe has been so influential, her arguments demand special attention.

Anscombe rejects what she takes to be a "familiar doctrine" that anything can be wanted.<sup>26</sup> Here is a pregnant passage of Anscombe's:

---

<sup>25</sup> See, for example, Joseph Raz, "Agency, Reason, and the Good", from his *Engaging Reason*, p. 22, fn. 1 and Sergio Tenenbaum, *ibid.*, p. 150, fn. 9.

<sup>26</sup> *Intention*, p. 67.

It will be instructive to anyone who thinks [that anything can be wanted] to approach someone and say: “I want a saucer of mud” or “I want a twig of mountain ash.” He is likely to be asked what for; to which let him reply that he does not want if *for* anything, he just wants it. It is likely that the other will then perceive that a philosophical example is all that is in question. . . but supposing that he did not realize this, and yet did not dismiss our man as a dull babbling loon, would he not try to find out in what aspect the object desired is desirable? . . . if the reply is: “Philosophers have taught that anything can be an object of desire; so there can be no need for me to characterize these objections as somehow desirable; it merely happens that I want it,” then this is fair nonsense.<sup>27</sup>

To be sure, Anscombe would concede that we can have odd or idiosyncratic wants, such as a desire for a saucer of mud. But Anscombe does suggest that wanting a saucer of mud is a more complicated matter than is perhaps supposed. For now, note that Anscombe accepts what I shall call *the desirability thesis*: any agent who genuinely wants something must be able to sincerely discern what aspect of the object of her want is desirable, and an agent who cannot sincerely discern what aspect of the object of her want is desirable does not really want the thing.

Moreover, Anscombe claims that if an agent is able to sincerely discern in what aspect the putative object of her want is desirable, then that agent is able to provide a *desirability characterization* of the object of her want. All of this suggests that Anscombe accepts a heuristic device for determining if an agent genuinely wants something: an interrogator can ask a series of “What for?” and “And what is the good of that?” questions until an intelligible answer has been reached.<sup>28</sup> Call this the *desirability test*. An agent who provides an intelligible answer to her interrogator’s questions brings her interrogator’s questioning to a close insofar as she provides a desirability

---

<sup>27</sup> Anscombe, *ibid.*, pp. 70-1.

<sup>28</sup> Anscombe, *ibid.*, p. 72 and 75. Anscombe also claims that an agent who insisted that he wanted something yet was unable to explain in what aspect the thing he purportedly wants is desirable is saying something unintelligible; *ibid.*, pp. 26-7.

characterization of what she purportedly wants. Thus, the desirability thesis can be restated in at least two ways. First, an agent who wants something must be able to provide a desirability characterization of what is wanted. Second, an agent who wants something must be able to pass the desirability test.

Some proponents of standard accounts have argued that Anscombe's desirability thesis implies that the *desirebad*, *intentionbad*, and *reasonbad* theses are all false. After all, not just any answer in reply to the desirability test will suffice to show that an agent genuinely wants something and it is far from clear how an agent could intelligibly want and intentionally pursue what she believes to be bad and not at all good. Therefore, claims the proponent of standard accounts, when an agent putatively wants or intentionally pursues the bad she will be unable to provide a desirability characterization of what she wants and what she pursues, and her inability to produce a desirability characterization is evidence that she does not really want or intentionally pursue what she believes to be bad and not at all good. I shall have more to say about Anscombe and her adherents in later chapters, but note that at least two responses are available.

First, a case can be made that Anscombe cannot rule out the possibility of intentionally pursuing the bad. Interestingly, Anscombe does *not* claim that passing the desirability test is necessary to intentionally pursue something. Anscombe finds it plausible that we sometimes perform actions "for no reason" and "for no particular reason" when we act without purpose or absent calculation, for example; when we act for no reason, Anscombe claims, we cannot sincerely provide any desirability characterization of what we intentionally pursue.<sup>29</sup> Now, the proponent of the

---

<sup>29</sup> Anscombe, *ibid.*, p. 73.

*intentionbad* thesis can plausibly claim that instances of the intentional pursuit of the bad are actions performed “for no reason” or “for no particular reason.” Matters will be clarified in Chapter 4, but for the moment I suggest that we act “for no reason” when we intentionally pursue the bad in the same sense as we act “for no reason” when we act on the basis of whims and passing fancies. So even if no sincere desirability characterization is possible in instances of the intentional pursuit of the bad, it does not follow that the intentional pursuit of the bad is not possible just because actions performed “for no reason” are possible.<sup>30</sup>

Second, accepting Anscombe’s desirability thesis does not commit one to denying that wanting the bad is possible. For all that has been said, an agent could claim that the believed badness of something is what she finds desirable about the thing and thus that the desirability characteristic of the thing that she wants is *that it is bad*.<sup>31</sup> Admittedly, after introducing the desirability thesis, Anscombe claims that “all that is required for our concept of ‘wanting’ is that a man should see what he wants under the aspect of some good” and this quote suggests that Anscombe believes that it is necessary to wanting something that the agent believes the thing she wants is good in at least some aspect.<sup>32</sup> Thus, besides endorsing the desirability thesis, Anscombe also appears to accept a stronger thesis that I shall refer to as the *goodness-as-desirability thesis*: for any agent

---

<sup>30</sup> It is not clear, unfortunately, that this line of argument preserves the truth of the *reasonbad* thesis. Perhaps this is acceptable: two out of three is not that bad and the truth of my first two theses would still imply that standard accounts are mistaken.

<sup>31</sup> See, for example, Stocker, “Raz on the Intelligibility of Bad Acts”, from *Reasons and Value: Themes from the Moral Philosophy of Joseph Raz*, Wallace, et al., eds., pp. 303-32. Anscombe does accept that the statement “The good of it is that it is bad” can be intelligible; *ibid.*, p. 75. However, Anscombe does not accept that something’s believed badness *itself* can be what an agent finds desirable about the thing she wants. See my discussion of Anscombe’s account of Satan below.

<sup>32</sup> Anscombe, *ibid.*, p. 75.

who genuinely desires something, that agent must be able to sincerely discern in what aspect the putative object of her want is desirable and it must be the case that what an agent finds desirable about the putative object of her want is something she believes to be good in some aspect. Now, the philosopher who accepts the goodness-as-desirability thesis cannot also consistently claim that wanting the bad is possible, but Anscombe's desirability thesis can be accepted even if Anscombe's desirability-as-goodness thesis is not; the goodness-as-desirability thesis is stronger than the desirability thesis and accepting the latter does not demand acceptance of the former. And the desirability-as-goodness thesis cannot simply be assumed without begging the question since its rejection is entailed by the *desirebad* thesis. Anscombe's adherents do of course have arguments for something like the desirability-as-goodness thesis and I discuss them further in my Chapter 4. For now, I maintain that the proponent of my three theses can argue that the believed badness of what is wanted can be what an agent finds desirable about what she pursues.

But the proponent of standard accounts will surely not give up the game so easily and will surely demand an explanation of the proposed thesis that the believed badness of something *could* be found to be desirable. To forestall a confusion, the proponent of standard accounts who suggests that it is unclear how something believed to be bad *could* be found desirable might be making either a modal claim or a normative claim. For example, I am plausibly making a normative claim when I utter "How *could* you have done that?" suggesting that what you did ought not to have been done.<sup>33</sup> But the philosopher who claims that wanting and intentionally pursuing the bad are possible is

---

<sup>33</sup> My use of the word 'claim' in this sentence is used to avoid claiming that my utterance is a proposition so as not to beg questions in favor of moral realists.

making a modal claim, not a normative one. And it is not implausible that there are explanations of the *desirebad* and *intentionbad* theses.

Michael Stocker provides at least the beginnings of an explanation of how it is possible that something believed to be bad could be regarded as desirable. Stocker imagines an embittered politician, who no longer cares for his constituents and is no longer moved to help them but rather to harm them. It is not that the politician no longer sees opportunities to do good for his constituents, nor that he fails to recognize such opportunities to help *as* good; rather, he now sees what would be good with respect to his constituents as something that should be prevented or destroyed. Stocker's politician still recognizes that various bills would benefit his constituents, but now, instead of wanting to push such bills through legislative channels, he wants to quash them. A proponent of Anscombe's desirability thesis might appeal to many different considerations that the politician might take to be good about harming his constituents. Perhaps the politician would painfully regard himself as duty's slave if he helped; thus, perhaps he fails to help because he thinks avoiding painful feelings is good.<sup>34</sup> Perhaps the politician views his constituents as ungrateful and, in a retributivist moment, he judges their suffering to be good because it is deserved. Perhaps the politician believes that their present suffering will allow him to save them later and reap the rewards of being their savior. Or perhaps he wants to be appreciated and this is a way to let his constituents know how much they need him. Or perhaps he has been reading Machiavelli and believes having a reputation of being fearsome is required to stay in power. Stocker denies that any of these considerations *must* be what his politician wants and that it could be the believed *badness*

---

<sup>34</sup> Stocker, *ibid.*, p. 742.

of harming his constituents that the politician finds desirable. If the politician is sufficiently depressed or enraged or bitter, his judgments concerning what is desirable could alter significantly, and in the right sort of mood, he might come to regard the badness of harming his constituents as what is desirable about performing actions that cause his constituents to be harmed.<sup>35</sup>

If the arguments in this section are correct, then the proponent of the *desirebad*, *intentionbad*, and *reasonbad* theses need not reject Anscombe's desirability thesis. Still, it may still be somewhat unclear just what the intentional pursuit of the bad is, as well as what instances of the intentional pursuit of the bad have in common with other types of intentional actions. In the next section, I explain what the intentional pursuit of the bad is and what it is not.

### **What Intentionally Pursuing the Bad Is . . . and What it Is Not**

I have suggested that the intentional pursuit of the bad is possible. However, it may remain unclear just what the intentional pursuit of the bad is and what it is not. In this section, I offer necessary and sufficient conditions for the intentional pursuit of the bad. The necessary and sufficient conditions for intentionally pursuing the bad are parasitic on the necessary and sufficient conditions for intentional action generally. I will not supply a full-blown analysis of intentional action, but I suggest that whatever the best full-blown analysis turns out to be, the proponent of the possibility of intentionally pursuing the bad can supply a full-blown analysis of the intentional pursuit of the bad by adding two further necessary conditions. I discuss the two necessary conditions below. I also

---

<sup>35</sup> Schopenhauer initially accepts that the suffering of others can be desirable: "the suffering of others is not a means for the attainment of the ends of its own will, but an end in itself." But Schopenhauer then retreats and claims an agent who desires the suffering of others "seeks . . . to mitigate his own suffering by the sight of the suffering of others, which at the same time he recognizes as an expression of his power." See *The Philosophy of Schopenhauer*, Irwin Edman, ed., p. 293.

explicitly distinguish the intentional pursuit of the bad from other kinds of behavior that it might be confused with.

I have already stipulated that when agents intentionally pursue the bad, they believe that the object of their pursuit is bad and not at all good. Thus, instances of the intentional pursuit of the bad are instances of intentional action but are distinguished from paradigm instances of intentional action insofar as an agent who intentionally pursues the bad believes that the object of her pursuit is bad. That an agent believes the object of her pursuit is bad is the first additional necessary condition for intentionally pursuing the bad.

To see that the first necessary condition is required for an agent to intentionally pursue the bad, consider the following two agents: Donald and David. Both Donald and David want a gin and tonic, believe that gin and tonics are good to drink, and believe the drink in front of them is a gin and tonic. Suppose that on that basis, Donald and David both proceed to intentionally drink what is in the glass before them. However, the drink before Donald is made of gin and tonic while the drink before David is made of petrol and tonic. David, unlike Donald, is motivated to pursue something that is, in fact, bad. While David performs an action that is, in fact, bad to perform, David's action is not an instance of the intentional pursuit of the bad. Cases in which we want what is in fact bad while believing what we want is good are neither uncommon nor do they present any real challenge to standard accounts of intentional action. Since Donald does not intentionally pursue the bad and given there is no relevant difference between the psychologies of David and Donald, then it follows that David does not intentionally pursue the bad either.

Consider also the following case involving two agents: Harry and Frank. Both Harry and Frank want, not a gin and tonic, but a petrol and tonic, believe that drinks made of petrol and tonic are bad and not at all good, and believe that the drink before them is a petrol and tonic. Suppose that on that basis, Harry and Frank both proceed to intentionally drink what is in the glass before them. However, the drink before Harry is, in fact, made of petrol and tonic while the drink before Frank is, in fact, made of gin and tonic. Harry is motivated to pursue something that is, in fact, bad while Frank is motivated to pursue something that is, in fact, not bad. Given my definition of the intentional pursuit of the bad, Harry does intentionally pursue the bad. Further, I submit that if it is the case that Harry intentionally pursues the bad, then so does Frank. Even though Frank intentionally pursues something that is, in fact, not bad, Frank is still a counter-example to standard accounts since Frank intentionally pursues what he believes is bad and believes that the object of his pursuit is not at all good. Since Harry does intentionally pursue the bad and given that there is no relevant difference between the psychologies of Harry and Frank, then it seems to follow that Frank intentionally pursues the bad as well.

The examples of Donald and David and of Harry and Frank suggest that it is neither necessary nor sufficient to intentionally pursue the bad that the object of one's pursuit actually is bad or bad to pursue. What is essential to intentionally pursuing the bad is that the agent *believes* that what she intentionally pursues is bad. It is the fact that Donald nor David believe the object of their pursuit is good that makes it the case that standard accounts can explain their actions, even if what David pursues is, in fact, bad. And it is the fact that Harry and Frank both believe the object of their pursuit is bad and

not at all good that makes it the case that standard accounts will have difficulty explaining their actions, even if what Frank pursues is good.

Supposing that an agent acts intentionally, the fact that she believes that what she pursues is bad is necessary, but not sufficient, for intentionally pursuing the bad. It is also necessary that, either, she believes that what she pursues is not at all good or that she fails to believe that what she pursues is at all good. Proponents of standard accounts can allow that agents can intentionally pursue what they believe to be bad provided that those agents *also* believe that what they pursue is good in some respect. If agents who intentionally pursue the bad really do believe that what they pursue is good in some respect then standard accounts can deny that intentionally pursuing the bad ever really occurs; agents who apparently intentionally pursue the bad have conflicting beliefs but their actions are no counter-example to standard accounts. It is perhaps not uncommon to believe, after all, that what one pursues has something to be said for it and something to be said against it, and it is surely rare that we believe that what we pursue is, for example, only good and not at all bad.

The proponent of standard accounts needs to argue that any and all instances of the intentional pursuit of the bad are instances in which an agent has multiple evaluative beliefs about the object of her pursuit, namely, that the object of her pursuit is bad *and* that the object of her pursuit is good in some respect. Call cases of the intentional pursuit of the bad in which an agent also believes that the object of her pursuit is good in some respect *impure* instances of the intentional pursuit of the bad. Cases of the intentional pursuit of the bad in which an agent believes that the object of her pursuit is bad and fails

to believe that the object of her pursuit is good in some respect can therefore be referred to as *pure* instances of the intentional pursuit of the bad.

Note that the proponent of standard accounts cannot necessarily accommodate even impure instances of the intentional pursuit of the bad. Instances of the impure intentional pursuit of the bad might, at least sometimes, be instances of causally over-determined intentional action. Generally, some event *E* is causally over-determined if it is the case that there are multiple causes of *E*, each of which is causally sufficient by itself to bring *E* about. The event of Jones' being killed is causally over-determined if he is simultaneously shot by two different assassins, since, presumably, Jones would have been killed if either one but not both of the assassins had shot. In impure instances of the intentional pursuit of the bad, it might be the case that an agent's desire for the bad and her desire for what she believed to be good causally over-determine her action, since, presumably, either would have sufficed to lead her to act. But proponents of standard accounts cannot allow that an agent's desire for what she believes to be bad and not at all good might be sufficient to explain her action. Therefore, impure instances of the intentional pursuit of the bad are perhaps as problematic for the proponent of standard accounts as pure instances of the intentional pursuit of the bad.

Generally, the examples of the intentional pursuit of the bad that I discuss are pure instances of the intentional pursuit of the bad and even if the proponent of standard accounts can accommodate the possibility of impure instances of the intentional pursuit of the bad, pure instances of the intentional pursuit of the bad are starker counter-examples to standard accounts. Note that it appears that there are some pure instances of the intentional pursuit of the bad. Consider an agent that I refer to as 'Carla' who

engages in self-mutilating, or “cutting”, behavior.<sup>36</sup> Carla is being treated for both anorexia and depression. During the course of her treatment, Carla reveals she cuts herself on her arms and breasts and that her targeted areas of cutting are selected because mutilation of these areas creates the greatest amount of pain with the least amount of visible damage. Carla’s self-mutilating behavior is diagnosed as the product of feelings of low-self esteem and self-hatred that are themselves the product of her parents’ unhappy marriage. But what is interesting about Carla is that this is not how things seem to her.

Some comments about self-mutilating behavior are in order to forestall confusion. First, there is no consensus that self-mutilation is a distinct mental disorder rather than a consequence of other mental disorders, such as obsessive-compulsive disorder, anxiety or depression, eating disorders, psychosis, and so forth.<sup>37</sup> While it is true that self-mutilating behavior is often concurrent with other mental illnesses, it is not true that every agent who engages in self-mutilating behavior also suffers from depression or anorexia or obsessive-compulsive disorder or borderline personality disorder, and so forth. Second, there are several competing models that attempt to explain self-mutilating behavior in terms of its function.<sup>38</sup> Cutting behavior might function as an expression of anger or anxiety that cannot otherwise be expressed; it may be the result of external pressures and signal that the patient is residing in an unhealthy environment; it may function as a “suicide replacement”; it may serve as a means for a patient to distinguish

---

<sup>36</sup> Carla’s case is based upon a case study in Steven Levenkron, *Cutting*, pp. 90-1, p. 110.

<sup>37</sup> Self-mutilating behavior is not recognized as a distinct mental disorder in the most recent DSM, although trichotillomania, the compulsive pulling out of one’s hair, is and receives its own entry.

<sup>38</sup> For discussion, see Suyemoto, “The Functions of Self-Mutilation”, *Clinical Psychology Review* 18 (1998), pp. 532-3.

herself from others<sup>39</sup>; some case studies suggest self-mutilation might be inflicted to deter sexual assault<sup>40</sup> or to reinforce the ability to suffer great pains.<sup>41</sup> No one functional model has gained consensus as *the* functional model that explains self-mutilation. As a result, there is no reason to suppose that there is any one particular desire that is *the* cause of self-mutilating behavior.

Return to Carla. Carla believes she engages in self-mutilating behavior in order to harm herself and she explicitly denies that she wants to hurt anyone else; Carla insists that it “would defeat the whole purpose” if anyone else, such as her parents, were hurt by her self-mutilating behavior.<sup>42</sup> Further, Carla denies that she “enjoys” the pain she inflicts upon herself; self-mutilators are not, by definition, sadists.<sup>43</sup> Generally, Carla does not herself believe her acts of self-mutilation are instrumental to realizing something else she believes to be good, such as the feeling of release or gaining control; some cutters report that their self-mutilating behavior is not a means to an end, unlike those who undergo the pain of a tattoo for the sake of self-expression. Levenkron claims that for the self-mutilator, “the act of creating pain, when pain is experienced. . . is *in itself* the

---

<sup>39</sup> Suyemoto, *ibid.*, pp. 537-48.

<sup>40</sup> Robert Cavanaugh, “Self-Mutilation as a Manifestation of Sexual Abuse in Adolescent Girls”, *Journal of Pediatric and Adolescent Gynecology* (2002), vol. 15, no. 97, pp. 97-100.

<sup>41</sup> Levenkron, *ibid.*, p. 120-3.

<sup>42</sup> Levenkron, *ibid.*, p. 91.

<sup>43</sup> Levenkron takes great pains to distinguish self-mutilators from sadists; *ibid.*, pp. 23-6. It is interesting to note that many cutters do not report feeling pain or painful feelings while cutting themselves; Suyemoto reports many self-mutilators do not experience pain during the act; *ibid.*, pp. 534. Note that the layman’s understanding of the sadist as someone who enjoys pain or inflicting pain is mistaken; it is not a necessary condition of being a sadist that one takes pleasure in the infliction of pain. The DSM-III lists eight possible symptoms of sadistic personality disorder, only one of which demands that an agent takes pleasure or enjoys the infliction of pain.

goal.”<sup>44</sup> Carla’s own stated purpose is that she engages in such behavior *simply to harm herself* and this is a result that she believes is a bad and not at all good.

Carla appears to be a genuine counter-example to standard accounts, insofar as she appears to act intentionally but she does not believe that the object of her pursuit, namely, the infliction of bodily harm, is at all good. It is tempting to suppose that Carla is therefore a paradigm example of an agent who intentionally pursues the bad and that self-mutilating behavior is a paradigm example of the intentional pursuit of the bad. However, supposing that Carla and self-mutilating behavior are paradigm examples can suggest some potentially misleading generalizations. Similarly, although I have suggested that agents like Augustine intentionally pursue the bad, the idea that Augustine and his actions are paradigm examples can suggest some potentially misleading generalizations. In the remainder of this section, I want to discuss a series of potential mistaken assumptions about the intentional pursuit of the bad and continue to clarify what the intentional pursuit of the bad is and what it is not.

First, it might be supposed that all agents who intentionally pursue the bad are estranged from their actions, fail to identify with the desires that produce their actions, and feel shame or regret at performing their actions. Carla’s own actions trouble her and she herself appears to have feelings of regret and shame and so forth; Carla intentionally pursues the bad *despite* her belief that engaging in self-mutilating behavior is bad and not at all good. Yet we can imagine another version of Carla, Carla\* who is not estranged from her cutting behavior, wholeheartedly identifies with her desire to harm herself, and feels no shame or regret and so forth. Carla\* does not harm herself *despite* her belief that

---

<sup>44</sup> Levenkron, *ibid.*, p. 41.

engaging in self-mutilating behavior is bad and not at all good, but instead harms herself *just because* her self-mutilating behavior is bad and not at all good. But whatever their differences, both Carla and Carla\* count as agents who intentionally pursue the bad.<sup>45</sup>

Second, it might be supposed that intentionally pursuing the bad is intentionally performing an action one believes to be *morally* bad. Augustine, for example, performs an action that is, by his own lights, morally bad. However, even if an agent performs an action she believes to be morally bad and not at all morally good<sup>46</sup>, she does not necessarily intentionally pursue the bad. Actions that are believed to be morally bad can be believed to be good in some other respect. Morally bad actions can be fun or humorous or might be just the thing to gain profitably. A Nietzschean might believe that intentionally acting in morally bad ways is expression of her liberation from a slavish morality and that liberating herself from a slave morality is a good thing. One can perfectly well believe that an action is good in virtue of being fun or humorously naughty or profitable or liberating while also believing that action is morally bad.

The remarks in the previous paragraph may suggest that instances of the intentional pursuit of the morally bad are never also instances of the intentional pursuit of the bad; this too is a mistaken generalization but one that has philosophical precedent. Consider Milton's *Paradise Lost* and Satan's infamous declaration: "Evil be thou my Good."<sup>47</sup> It appears that Satan is reversing his moral compass; as an angel he is attracted to the good and only the good, and he becomes a fallen angel when he resolves to no longer be so

---

<sup>45</sup> Augustine also appears to count as intentionally pursuing the bad and so forth, but Augustine more resembles Carla\* than Carla.

<sup>46</sup> Nothing rests on my use of the term 'bad' rather than some other moral term of disapprobation; we could just as easily speak of the belief that one is pursuing something wrong or wicked or vicious, and so forth.

<sup>47</sup> *Paradise Lost*, Book IV, line 110.

attracted. Still, Satan appears to have goals: recall that Satan claims that it is “Better to rule in Hell than to serve in Heaven,” suggesting he still believes that freedom and autonomy are good, even if he fails to believe freedom and autonomy are morally good.<sup>48</sup> Anscombe offers an interpretation of Satan where he pursues what he believes to be good, but not morally good:

‘Evil be thou my good’ is often thought to be senseless in some way. Now all that concerns us here is that “What’s the good of it?” is something that can be asked until a desirability characterization has been reached and made intelligible. If then the answer to this question at some stage is ‘The good of it is that it’s bad,’ this need not be unintelligible; one can go on to say ‘And what is the good of its being bad?’ to which the answer might be the condemnation of good as impotent, slavish, and inglorious. . . . all that is required for our concept of ‘wanting’ is that a man should see what he wants under the aspect of some good.<sup>49</sup>

On Anscombe’s interpretation of Satan, Satan wants to perform an action he believes is morally bad only because he believes that what he pursues is good in some non-moral respect. Thus, Anscombe’s Satan fails to intentionally pursue the bad. It is tempting to claim that agents who pursue what they believe to be morally bad must believe nonetheless that what they pursue is good in some non-moral respect. Anscombe’s Satan reverses his moral compass only as a means for getting what he regards as good. Here is Velleman’s complaint about Anscombe’s Satan:

What sort of Satan is this? He is trying to get things right, and so he rejects the good only because he has found respects in which it is unworthy of approval. He rejects the good, that is, only because it is slavish and inglorious, and hence only because shunning the good is a means to liberty and glory. But then he isn’t really shunning the good, after all, since the goods of liberty and glory remain his ultimate goals. Anscombe’s Satan can want evil only by judging it to *be* good, and

---

<sup>48</sup> *Paradise Lost*, Book I, line 263.

<sup>49</sup> Anscombe, *Intention*, p. 75.

so he remains, at heart, a lover of the good and the desirable—a rather sappy Satan.<sup>50</sup>

If Velleman is correct, then Anscombe's Satan simply accepts an idiosyncratic extension of 'good' and wills the means required to obtain what he believes to be good. If this is how we are to understand Satan, then we must admit Satan is no different from other agents who will the means necessary to realize their ends.<sup>51</sup> But the idea that Satan is really pursuing what he believes is non-morally good fails to preserve Satan's perversity. Note that Satan is a classic example of an agent in the grip of self-deception. Satan goes to rather extraordinary lengths to keep himself and the other fallen angels deceived with dubious accounts of their "successes" and predictions of their "victories." Satan claims that "the adverse power" of the rebelling angels shook the very throne of the Almighty, knowing full well it did not.<sup>52</sup> In response to Beelzebub's challenge that their deeds only proved "the high supremacy" of "heaven's perpetual king,"<sup>53</sup> Satan responds that their labors to pervert God's good ends will "oft-times. . . succeed, so as perhaps shall grieve him, if I fail not, and disturb his inmost counsels from their destined aim."<sup>54</sup> But Satan knows full well their labors will fail. And recall that on some accounts of the story, Satan knows full well that he will lose in the end and that his efforts are bound to fail and that he will not be able to obtain what he is now intentionally pursuing. What is perverse

---

<sup>50</sup> Velleman, "The Guise of the Good", p. 19. See also S. I. Benn's discussion of Satan in "Wickedness", *Ethics*, 95:4 (Jul., 1985), reprinted in *Ethics and Personality: Essays in Moral Psychology*, John Deigh, ed., p. 201. Benn would complain, I suggest, that Anscombe's interpretation of Satan ruins the thought that Satan is malignantly wicked, and Satan is malignantly wicked if anyone is.

<sup>51</sup> Velleman, *ibid.*, p. 18.

<sup>52</sup> See *Paradise Lost*, Book I, lines 104-105; Book VI, lines 710-12 and line 834 shows Satan's claim to be false: it is Christ's chariot and not Satan's armies which shake heaven to its foundations.

<sup>53</sup> *Paradise Lost*, Book I, lines 131-2.

<sup>54</sup> *Paradise Lost*, Book I, lines 166-8.

about Satan is not that he has an idiosyncratic account of “good” but rather that he continues to will an end that he knows will ruin him if he realizes it. It may be the case that while Satan believes the autonomy that he pursues is *pro tanto* good, since he knows that getting autonomy will ruin him, he believes that his act of pursuing autonomy in this case is bad. Thus, I suggest, it is not plausible to suppose that Satan intentionally pursues the morally bad but fails to intentionally pursue the bad simpliciter.

It is tempting to say, given the above discussion, that the agents that I am interested in do not intentionally pursue the morally bad, but intentionally pursue the bad *simpliciter*. But this is also misleading and suggests another misunderstanding. For an agent who intentionally pursues the bad may not want the bad *simpliciter*. Carla, for example, intentionally pursues something she believes is bad but Carla is not interested in pursuing just *anything* she believes to be bad. For example, Carla believes that harming her family members is bad and not at all good, but she denies that she is interested in harming members of her family. Agents who intentionally pursue the bad, I suggest, do not intentionally pursue just anything that they believe to be bad and not at all good, but only what they believe to be bad in some respect. Further, note that appealing to an agent’s emotions or moods and so forth in explaining the intentional pursuit of the bad, as Stocker suggests, helps to explain why an agent may find something she believes to be bad desirable. Carla is frustrated with herself and lacks self-esteem and thus her anger is directed at herself, rather than her parents. So it is perhaps not surprising that while Carla believes both that harming herself is bad and not at all good and that harming her parents is bad and not at all good, she only wants to harm herself and not her parents, just because Carla’s anger and frustration are directed at herself and not them.

Finally, one may suppose that an agent who intentionally pursues the bad must be irrational, or at the very least, must be in the grip of some powerful emotion or mood that makes them less than clear-headed. After all, Carla appears to suffer from a mental disorder. Further, it is often supposed that agents who perform actions they believe to be immoral or otherwise bad, like Augustine and Satan must be irrational. I shall have more to say about rationality later. Stocker is helpful here as well. Stocker claims that motivation always involves background emotions and moods and so forth.<sup>55</sup> If Stocker is correct, then it cannot be that Carla and Stocker's politician are irrational or confused *merely* because they are in the grip of some emotion or mood. If being motivated always depends upon moods or emotions, then the fact that Carla and Stocker's politician are moved as a result of their emotions or moods does not distinguish them from rational and clear-headed agents.

In this section, I have been attempting to clarify what the intentional pursuit of the bad is and what it is not by attempting to explain away potential misunderstandings. One further clarification remains: the intentional pursuit of the bad is not identical to akratic action. Not all instances of akratic action are instances of the intentional pursuit of the bad and it might be the case that not all instances of the intentional pursuit of the bad are instances of akratic action. However, akratic action has long been of philosophical interest, and thus I dedicate the following section to distinguishing the intentional pursuit of the bad from akratic action.

---

<sup>55</sup> Stocker, *ibid.*, pp. 750-3.

### Intentionally Pursuing the Bad and Akratic Action

One familiar picture of the *akrates* is the picture of an agent who is overcome by base urges in opposition to her belief or knowledge that she should not act on those urges; for example, we might describe an akratic agent who smokes as being “overwhelmed” by an urge for a cigarette, all the while knowing that she should not smoke. This familiar picture cannot be taken too seriously for at least two reasons. First, not every case of akratic action is a case where an agent is overwhelmed by some urge, much less by some base urge. I might believe that I ought to stay in bed because my bed is so warm and comfortable, but because I have an urge to take care of myself, I akratically get up to brush my teeth.<sup>56</sup> Cases of “inverse akrasia” are also perhaps relevant here. In cases of inverse akrasia, an agent fails to act as she believes she ought to, but winds up actually doing the right thing.<sup>57</sup> For example, Huckleberry Finn believes that Jim should be turned over to bounty hunters, yet when the opportunity presents itself, Huck’s conscience prevents him from doing what he believes is his moral duty and Huck akratically saves Jim out of friendship. Second, the picture of the akratic agent as “overcome” by base urge threatens our sense that akratic action is voluntary, since intentional actions are voluntary and akratic actions are intentional actions. The behavior of the *akrates* becomes difficult to construe as intentional action if we suppose the “akratic” agent is overcome by urges.

Rejecting the familiar picture, we might instead describe the *akrates* as an agent who intentionally acts while judging that some alternative action would be better or

---

<sup>56</sup> The example and the point belong to Davidson, “How is Weakness of the Will Possible?”, p. 30.

<sup>57</sup> Nomy Arpaly discusses such cases in *Unprincipled Virtue*, p. 9-10.

superior. On Davidson's persuasive analysis of akratic action, an agent acts akratically just in case she acts contrary to her all-things considered judgment about what is to be done; more formally, an agent acts akratically in performing some action  $x$  just in case she performs  $x$  intentionally, she believes there is at least some other action  $y$  open to her, and she judges all-things considered that  $y$  would be better to do than  $x$ .<sup>58</sup> Davidson's analysis captures the intuition that akratic actions are contrary to the agent's belief that she acts in a way that she ought not to. Therefore, it might be thought that an agent who intentionally pursues the bad must act akratically. After all, an agent who intentionally pursues the bad believes her action is bad and not at all good and thus it is plausible to suppose that the agent believes that there is some other action available to her that is better than intentionally pursuing the bad. Nonetheless, there are important differences between akratic action and intentionally pursuing the bad.

The tooth-brushing example suffices to show that an agent can act akratically without making a *moral* all-things-considered judgment that she ought to act differently. The tooth-brushing example *also* shows that not every case of akratic action is a case of intentionally pursuing the bad: I might act against my all-things considered judgment when I get out of bed and brush my teeth, but I still think brushing my teeth is good in some respect. Similarly, when I smoke akratically I believe that I ought not to smoke, but I still see some good in smoking. Generally, even if I fail to act in accord with my all-things considered judgment, I might still perform an action I believe is good in some

---

<sup>58</sup> The analysis is taken from Davidson, *ibid.*, p. 22. Other philosophers have proposed analyses sympathetic with Davidson's construal of akratic action as action equivalent to acting against one's all things considered judgment: see Gerasimos Santas, "Plato's *Protagoras* and Explanations of Weakness", *Philosophical Review* 65 (1966), pp. 3-33; A. Phillips Griffiths, "Acting with Reason", *Philosophical Quarterly* 8 (1958), pp. 289-99; Robert Audi, "Weakness of Will and Practical Judgment", *Nous* 13 (1979), pp. 173-96; and Al Mele, *Irrationality*, chap. 2.

respect, however minimal, and thus fail to intentionally pursue the bad while acting akratically.

Similarly, not every instance of intentionally pursuing the bad is an instance of akratic action; agents might intentionally pursue the bad while failing to believe there is some available alternative course of action that is better or superior. There are at least two sorts of instances of the intentional pursuit of the bad that are not also cases of akratic action; in both sorts of instances, the agent does not act in opposition to an all-things considered judgment that some alternative action would be better or superior.

First, note that we sometimes act just given the recognition there is a reason to act and without considering alternatives. To explain how an agent could akratically act against her all-things considered judgment, Davidson supposes the akratic actor makes two different kinds of practical judgments prior to acting: a conditional all-things considered judgment that an action would be desirable given all the evidence believed to be relevant, and a distinct unconditional judgment opposed to that all-things considered judgment.<sup>59</sup> For example, on Davidson's analysis, an agent who smokes akratically would consider reasons for and against smoking and form a conditional all-things considered judgment, presumably that she ought not to smoke because, given all the evidence believed to be relevant, smoking is not desirable. However, the akratic agent also judges unconditionally that she ought to smoke. If she acts on her unconditional judgment that she ought to smoke, she acts in opposition to her conditional all-things considered judgment that she ought not to smoke, and thus acts akratically. Note, however, that if an agent does not consider reasons for and against some course of action,

---

<sup>59</sup> For the details, see Davidson, *ibid.*, pp. 34-40. Davidson later identifies the distinct unconditional judgment with an intention to act.

then she could not form a conditional all-things considered judgment and an agent who does not form a conditional all-things considered judgment could not act akratically on the Davidsonian analysis of akratic action. Suppose an agent believes that smoking is bad while failing to believe that smoking is at all good. Suppose that in some situation, however, the agent forms the unconditional judgment that she ought to smoke without making *any* judgment that she ought *not* to smoke. Or suppose that she just straightaway decides to smoke. Given the agent does not judge all-things considered that she should not smoke, she could not smoke akratically, even if she intentionally pursues the bad.

There is another sort of case that suggests that not every case of the intentional pursuit of the bad is a case of akratic action. An agent might consider various reasons for performing various actions and form the all-things considered judgment that she ought to perform an action that she believes is bad and not at all good to perform. It is not that she is faced with a moral dilemma and decides to perform an action that is the “least-bad” of the actions she believes are available to her.<sup>60</sup> Instead, the agent I have in mind weighs the reasons for and against various courses of action, judges that the believed badness of some course of action is the strongest reason she has, and proceeds to act accordingly. Perhaps a sufficiently depressed or despairing agent would conclude that she should perform some action believed to be an inferior alternative on the grounds that it is inferior. Perhaps Augustine deliberates about whether to steal pears or apples, judges that it would be worse to steal the pears, and, on that basis, he intentionally steals the pears. Perhaps Carla believes that cutting herself is the worst thing that she could do, but

---

<sup>60</sup> This would not necessarily be a case of intentionally pursuing the bad because an agent who performs an action that is the lesser of all evils might believe that an action is good just because it is the lesser of two evils.

for that reason judges she has most reason to cut herself and acts accordingly. To be sure, this is a perverse way to judge what one's "best" or "superior" course of action is. However, the Davidsonian account of akratic action does not come with any built-in axiological theory, nor any account of what an agent's strongest or weightiest reasons are, or any account of how to go about adding up reasons. If an agent can get herself to judge all-things considered that she ought to perform an action she believes is bad and not at all good, then if she can proceed to form and execute the further unconditional judgment to perform that action, she can intentionally pursue the bad without also acting akratically.

Even if one allows for the possibility of akratic action, one might deny the possibility of intentionally pursuing the bad, for akratic actions are often performed for *something* the agent regards as good. If we stipulate that for all akratic actions, the agent still believes there is something good about the action she performs, then we can perhaps comprehend akratic action in a way that we cannot comprehend intentionally pursuing the bad. I shall do more to make the intentional pursuit of the bad comprehensible in subsequent chapters. But if one can find room in an account of moral psychology for intentional though akratic action, we might also find room for the intentional pursuit of the bad. I would suggest that the most plausible explanations of akratic action suppose that there is a "gap" between believing there are reasons for performing an action and intentionally performing that action, a gap that is to be filled in by an agent's decision or choice or intention to act. Reasons for action do not all by themselves determine what an agent will do. But if quite generally there is a gap between believing there are reasons for performing an action and acting intentionally, then it might be the case that an agent can

resist the force of reasons as she understands them or act in ways altogether differently from how such reasons prescribe. Accounts of akratic action will have to explain how it is that an agent can resist the force of reason that recommend acting in a way contrary to the way the akratic agent actually acts. And if an agent can intend and act intentionally in ways contrary the force of reasons for action, it seems equally possible that an agent could intend in opposition to what she believes would be good to do.

### **Conclusion**

In this introductory chapter, I have attempted to motivate and defend the *desire bad* thesis that agents can desire what they believe is bad and not at all good, the *intentionbad* thesis that agents can intentionally pursue what they believe to be bad and not at all good, and the *reasonbad* thesis that agents can act intentionally for the reason that their action is bad and not at all good. The project of this first chapter is simply to clarify the three theses; defending them is another matter. In the next chapter, I defend the *desirebad* thesis and in later chapters, I defend the *intentionbad* and *reasonbad* theses

## CHAPTER 2 DESIRE AND DESIRING THE BAD

... whenever the desire for something ignites in our hearts, we are moved to pursue it and seek it and, seeking and pursuing it, we are led to a thousand unruly ends.

Miguel de Cervantes Saavedra, *La Galatea*<sup>1</sup>

### Introduction

In this chapter, I defend the *desirebad* thesis: agents can desire what they believe to be bad and not at all good. I discuss competing accounts of desires and desiring and I argue that there are plausible accounts of desire that permit the possibility of desiring the bad. While there are some accounts of desire that do not appear to permit the possibility of desiring the bad, we have independent reasons for rejecting those accounts.

For the purposes of the ensuing discussion, I accept that desire always underlies intentional action, insofar as I accept that it is necessary for an agent to act intentionally that the agent has some desire that explains, or potentially explains, her acting.<sup>2</sup> As I note below, many philosophers of very different philosophical sympathies accept that desire always underlies intentional action. The thesis that desiring always underlies action may be false; indeed, some prominent philosophers reject it.<sup>3</sup> But if desiring does always underlie action, then it is plausible to suppose that an agent can intentionally pursue what

---

<sup>1</sup> Miguel de Cervantes Saavedra, *La Galatea*, Book IV.

<sup>2</sup> As noted below, the thesis that desire always underlies action might be confused with the Humean theory of motivation; I do not know that I endorse the Humean theory. My argument for the thesis that desire underlies all intentional action depends upon the truism that, for any intentional action, we would not have so acted unless we wanted to so act, and I stipulate that ‘want’ and ‘desire’ denote the same kind of psychological state.

<sup>3</sup> See, for example, Jonathan Dancy, *Practical Reality*.

she believes to be bad and not at all good only if she desires what she believes is bad and not at all good. So if desiring does always underlie intentional action, the truth of the *intentionbad and reasonbad* theses depends upon the truth of the *desirebad* thesis.

Relatedly, if the *desirebad* thesis is false, there is some reason to doubt the other two.

Two plausible assumptions will therefore guide the discussion that follows. First, I suppose that desires play an essential role in the production and explanation of intentional action. Second, I suppose that the explanation of intentional action is a kind of explanation Donald Davidson refers to as *rationalizing* explanation or *rationalization*.<sup>4</sup> In ordinary usage, ‘rationalization’ denotes a purported justification of an action, but a justification that seems bogus. In what follows, I follow Davidson’s usage of ‘rationalization’ to denote an explanation of intentional action that explains by revealing what it is that an agent saw, or thought she saw, about the action she performed that appealed to her and led her to act.<sup>5</sup> That there is something about an action that appeals to an agent rationalizes the performance of that action, and, when the explanation is a correct one, suffices to explain that action. The paradigmatic form of a rationalizing explanation is to cite the acting agent’s desire to realize some end and belief that performing this action will realize that end.<sup>6</sup> These two assumptions stand together fairly well. If desires always underlie intentional action, then any explanation of intentional action must explain by citing desires. If rationalization is the means for explaining

---

<sup>4</sup> Davidson, “Actions, Reasons, and Causes”, reprinted in his *Essays on Actions and Events*, pp. 3-19. Davidson also famously argues that rationalizing explanations are causal explanations, but this aspect of Davidson’s thesis need not be discussed here.

<sup>5</sup> Davidson, *ibid.*, p. 3.

<sup>6</sup> Davidson, *ibid.*, p. 5; while Davidson speaks of pro-attitudes rather than desires, Davidson also suggests desires are a sub-class of pro-attitudes. Therefore, if pro-attitudes can play role in rationalizing explanations, then so can desires.

intentional action, then since desire underlie intentional action, rationalizing explanations must cite an agent's desires.

### **Humean Accounts of Desire**

One popular account of desire is what I shall refer to as *the Humean Account of Desire or Humeanism*. Hume famously argues that “reason alone can never be a motive to any action of the will.”<sup>7</sup> Roughly, for Hume, ‘reason’ refers to the faculty of theoretical reason that produces belief. Contemporary Humeans often read Hume to mean that reason cannot supply the motive of any action of the will, and therefore beliefs, the product of theoretical reasoning, cannot be motives that can lead an agent to act. So what supplies and constitutes motives to act must be something besides belief: namely, desires. Contemporary Humeans have been impressed with the distinction Hume draws between desire and belief and have followed Hume in distinguishing desires from beliefs, in part, by appealing to the motivational force that desires possess and beliefs purportedly lack.<sup>8</sup> I cannot explore the full implications of the Humean account here, but I do want to examine whether desiring the bad is consistent with accounts of desire proposed by prominent contemporary Humeans. I isolate various theses that seem to be constitutive of the Humean account and then argue that the *desirebad* thesis is consistent with Humeanism about desire.

Here is one theme that is popular among Humeans: ‘desire’ functions as a class term, including in its extension all manner of propositional attitudes that can produce and explain behavior and intentional action. Since psychological states like fears and hopes

---

<sup>7</sup> David Hume, *Treatise of Human Nature*, Bk. II, Pt. III, Sect. III.

<sup>8</sup> For representative examples of contemporary Humeans, see Michael Smith, “The Humean Theory of Motivation”, *Mind* 96, pp. 36-61 and Chapter Four of his *The Moral Problem* and David Lewis, “Desire as Belief”, reprinted in his *Papers in Ethics and Social Philosophy*.

and wishes and intentions and so forth can produce and explain behavior and intentional action, then, the Humean claims, states like fears and hopes and wishes and intentions *are* desires.<sup>9</sup> Here then is a first attempt at capturing a Humean thesis about desire: a psychological state counts as a desire just in case that state can produce and explain behavior and intentional action.

Yet this first attempt at a Humean thesis cannot be constitutive of Humeanism. It is widely conceded by Humeans and non-Humeans alike that beliefs *must* play a role in the production and explanation of intentional action; even if one desires something, if one lacks any beliefs about the object of one's desire, such as the necessary or sufficient means for acquiring the object of one's desire, one is not going to do much of anything. But if we concede that beliefs do have a role to play in the production and explanation of action, then given the first attempt at a Humean thesis, beliefs are desires. Note also that some philosophers who identify themselves as *non-Humeans* argue that some beliefs, such as a belief that an action is prudent or morally required, are sufficient to motivate an agent to act absent any distinct psychological state that intuitively counts as a desire.<sup>10</sup> But these non-Humeans do not claim that motivationally efficacious beliefs are desires. The first attempt at a Humean thesis cannot do justice to a disagreement between Humeans and some non-Humeans and thus needs to be revised.

Here is a revision: 'desire' functions as a catch-all term, including in its extension all manner of *non-cognitive* propositional attitudes that are constitutively motivating.

---

<sup>9</sup> See Michael Smith, *ibid.*, p. 55 for discussion of this point.

<sup>10</sup> See, for example, Thomas Nagel, *The Possibility of Altruism*, and John McDowell, "Are Moral Imperatives Hypothetical Imperatives?", reprinted in his *Mind, Value, and Reality*.

Since beliefs are cognitive states rather than non-cognitive states, beliefs are not candidates for falling under the extension of ‘desire’ in the first place.

It is fine to suppose that desires, but not cognitive states like belief, are essential to motivation but some explanations must be offered why desires, but not beliefs, are essential to motivation. The conviction that desires but not beliefs are essential to motivation is often expressed in the claim that desire and belief have different “directions of fit.”<sup>11</sup> Desires, and non-cognitive conative states generally, have a “world-to-mind” direction of fit, insofar as they are satisfied when the world “fits” the mind. Alternatively, beliefs, and cognitive states generally, have a “mind-to-world” direction of fit, insofar as beliefs are true when the mind “fits” the world. The claim that beliefs and desires have different directions of fit helps to explain the Humean thesis that desires can produce and explain intentional action in a way that beliefs cannot. Consider Michael Smith’s analysis of what the difference in the direction of fit of desires and beliefs amounts to:

... the difference between beliefs and desires in terms of direction of fit comes down to a difference between the counterfactual dependence of a belief and a desire that *p*, on a perception that *not p*; roughly, a belief that *p* is a state that tends to go out of existence in the presence of a perception that *not p*, whereas a desire that *p* tends to endure, disposing a subject in that state to bring it about that *p*. Thus, we may say, attributions of beliefs and desires require that different *kinds* of counterfactuals are true of the subjects to whom they are attributed. We may say that this is what a difference in their directions of fit *is*.<sup>12</sup>

---

<sup>11</sup> I am uncertain of the origins of the locution “direction of fit”; I had thought the locution first appeared in Anscombe, but it seems not to appear at all, although Anscombe’s discussion of the shopkeeper pursued by a detective is often cited in the discussion of the directions of fit of belief and desire. Austin uses the locution to distinguish different sorts of speech acts; See Austin’s “How to Talk—Some Simple Ways”, *Proceedings of the Aristotelian Society*, 53, pp. 227-46. For further discussion, see Lloyd Humberstone, “Direction of Fit”, *Mind* 101 (Jan. 1992), pp. 59-83.

<sup>12</sup> Smith, “The Humean Theory of Motivation”, p. 54. For critical discussion of Smith’s proposal, see Lloyd Humberstone, *ibid.*, pp. 63-5; G. F. Schueler, “Pro-attitudes and Direction of Fit”, *Mind* 100, pp. 277-81; David Sobel and David Copp, “Against Direction of Fit Accounts of Belief and Desire”, *Analysis*

In a similar vein, Robert Stalnaker has proposed that “to desire that *P* is to be disposed to act in ways that would tend to *bring it about* that *P* in a world in which one’s beliefs whatever they are, were true.”<sup>13</sup> Both Smith and Stalnaker suppose that desires are desires *that P*, where *P* is understood as some proposition. We do not strictly desire beer or dissertations, but rather we desire *that we have a beer* or *that our dissertation be finished in a timely manner*. At the risk of obscuring a thesis that demands a technical formulation, we may say that to have a desire is to be motivated to bring it about that some proposition be made true. Thus, it is no surprise that there is an intimate connection between desire and action according to Smith and Stalnaker’s dispositional accounts of the direction of fit of desires and beliefs: to desire that *P* just is to be disposed to make *P* true and the paradigm means for making some proposition *P* true is by performing an action. I make it the case that I am enjoying a beer or finishing my dissertation by intentionally getting a beer or by intentionally revising my dissertation. But to believe that *P* is not necessarily to be disposed to bring it about that *P*. If one truly believes that *P*, then *P* is already the case and there is nothing to bring about. There are cases in which I believe something is the case that I am not disposed to bring about: I might believe that I am in great pain without being disposed to bring it about that I am in great pain. There are also cases in which I desire what is the case: I might desire to eat pasta salad while eating pasta salad. Once we are clear about what kind of dispositional state desires and beliefs are, it is no surprise that desires but not beliefs, are intimately connected to motivation and action.

---

61, pp. 44-53; Al Mele, “Motivation: Essentially Motivation-Constituting Attitudes”, *The Philosophical Review* 104:3, pp. 394-5.

<sup>13</sup> Stalnaker, *Inquiry*, p. 15, emphasis added.

It is not my intent to judge the adequacy of Humean accounts of desire or direction of fit accounts of desire and belief or to judge dispositional accounts of desire and belief. My interest is in the implications of Humean accounts of desire for the possibility of desiring the bad: the Humean account of desire does not preclude desiring the bad and actually helps to explain the possibility of desiring the bad.

The Humean might follow Hume in claiming that desires and beliefs are “distinct existences”; if desires and beliefs are distinct existences, then absent some further argument, it is plausible that there is no necessary connection between an agent having some belief and having some corresponding desire. In particular, absent some further argument, the Humean is not forced to claim that we can only desire what we believe good nor that the belief that something is bad crowds out a desire for that thing. I have supposed that Humeans include states like whims in the extension of ‘desire,’ but as I understand whims, it is possible to have a whim to perform an action without believing that what one pursues is good. If desires and beliefs are distinctive states in our psychological economies and if there are no necessary connections between them, then it should not be surprising to find that desires and evaluative beliefs can come apart such that we can desire what we believe is bad and not at all good.

One difficulty with the thesis that desires and beliefs are distinct existences and that there are no necessary connections between desiring and believing is that agents do seem with great regularity to desire what they believe is good or valuable or otherwise worth pursuing. But if beliefs do not themselves motivate agents in a way that desires do and if there are never any necessary connections between beliefs and desires, it is difficult to explain why agents regularly come to desire what they believe to be good. In response,

the Humean might claim that agents have a general desire to pursue what they believe is good. The Humean does not need suppose that the desire to pursue what one believes is good is irresistibly strong, but only strong enough to ensure the alleged regularity obtains between believing something is good and being motivated to pursue that thing. However, this picture seems to allow the possibility of agents having desires to perform actions they believe to be bad and to desire to pursue what they believe to be bad and not at all good. So Humeans can explain the regular connection between believing something is good and being motivated to pursue it without ruling out the possibility of desiring the bad.

Further, understanding desires as dispositions helps to explain how agents could desire something while believing that the object of the desire is bad and not at all good. Consider an agent like Carla who desires to harm herself in spite of believing that harming herself is bad and not at all good. Given Smith's analysis of desires as dispositions, insofar as Carla desires to harm herself, then given her perception that she is not presently harming herself, her to desire to harm herself disposes her to bring it about that she harms herself. If desires generally are dispositions in the way Smith and Stalnaker suggest, then desires for the bad might persist and constitute an agent's motivation despite an agent's belief that the object of desire is bad and not at all good.

Suppose that Humeanism about desire is constituted by the thesis that desires are non-cognitive propositional attitudes that can produce and explain behavior and intentional action and that dispose agents to bring about the content of the that-clause that is the object of the desire. The Humean account of desire does not imply that the *desirebad* thesis is false nor does it preclude desiring the bad. Unfortunately for the proponent of the *desirebad* thesis, the Humean account is not the only plausible account

of desire available. I consider some alternatives below. So while the proponent of the possibility of desiring the bad should be relieved she has an ally in the Humean, she cannot quite rest contented.

### **Nagel and Motivated Desires**

The Humean account of desire may be the dominant account of desire but it has not been endorsed by some of our best philosophers. Some contemporary philosophers with Kantian leanings have argued that some psychological states that are more like beliefs than Humean desires can supply motivation. In a brief but influential argument, Thomas Nagel attempts to resist accounting for all motivation in terms of Humean desires.<sup>14</sup> Since Nagel's account of desire has been rather influential, it will be useful to determine whether desiring the bad is consistent with Nagel's account of desiring.

Nagel rejects Humeanism about desire insofar as he rejects the assumptions that "all motivation has desire at its source" and that "belief by itself cannot produce action."<sup>15</sup> Surprisingly, however, Nagel does not reject the thesis that desire underlies action. Instead, Nagel admits the truth of a "trivial" claim that it is a "logically necessary condition" that a desire be present given that an agent is motivated.<sup>16</sup> Nagel claims that it "simply follows from the fact that. . . [some] considerations motivate me" that I have a desire: if the prospect of my future happiness motivates me, then it simply follows, according to Nagel, that I desire my future happiness.<sup>17</sup> Yet this admission does not

---

<sup>14</sup> Nagel, *ibid.*, pp. 29-32.

<sup>15</sup> Nagel, *ibid.*, p. 27.

<sup>16</sup> Nagel, *ibid.*, pp. 29-30. See also Wallace, "How to Argue About Practical Reason", p. 360.

<sup>17</sup> Nagel, *ibid.*, p. 29.

commit Nagel to Humeanism about desire for Nagel distinguishes between different kinds of desires, not all of which are amenable to Humeanism.

Nagel distinguishes desires that are *motivated* from desires that are *unmotivated*. Initially, Nagel claims that motivated desires, like beliefs, are the product of decision or deliberation and are to be distinguished from unmotivated desires that “simply come to us” and “assail us” in the way that appetites and some emotions do.<sup>18</sup> Presumably, Nagel means that motivated and unmotivated desires should be jointly exhaustive of desires, such that every desire is either motivated or unmotivated. Unfortunately, not every desire is either the product of decision or deliberation or simply “assails us.” Long-standing desires for one’s health or for the flourishing of one’s children need not result from decision or deliberation, but neither do they have the sort of phenomenology such that they “assail” the desiring agent. Fortunately, Nagel’s intent is clear. Nagel surely intended to distinguish motivated and unmotivated desires and given Nagel’s own suggestions, we can distinguish them as follows: motivated desires are conative states acquired as a result of decision or deliberation, while unmotivated desires are conative states not acquired as a result of decision or deliberation.<sup>19</sup>

While Nagel accepts that desire underlies every action, he does not accept that *unmotivated* desire underlies every action. Given the distinction between motivated and unmotivated desires, Nagel claims we cannot conclude that all motivation is ultimately grounded in *unmotivated* desires since it is an open question whether there must always be an unmotivated desire that initiates decision or deliberation that produces motivated

---

<sup>18</sup> Nagel, *ibid.*, p. 29.

<sup>19</sup> I borrow here from Schueler, *Desire*, pp. 20-1. See also Wallace, *ibid.*, p. 363.

desires.<sup>20</sup> Presumably, the Humean needs the claim that all action must ultimately be the result of some unmotivated desire since motivated desires appear to be more like cognitive states like beliefs rather than non-cognitive states like desires as the Humean understands them. But if some cognitive state can produce and explain intentional action, then Humeanism about desire must be rejected.

In the same way that it is not my present purpose to evaluate the validity of the Humean account of desire, it is not my present purpose to evaluate Nagel's critique of the Humean account nor to assess whether motivated desires all by themselves are capable of motivating agents to act. Rather, I want to assess whether Nagel's thesis that action can be produced by motivated desires implies that desiring the bad is not possible and that the *desirebad* thesis is false. As far as I can tell, the *desirebad* thesis is consistent with Nagel's thesis.

Note that I have not disputed that the origins of some desires lie in decision or deliberation. Admittedly, my account of Carla, for example, leaves it open whether her desire to harm herself is motivated rather than unmotivated, but I can see no reason to suppose that the proponent of the *desirebad* thesis should deny that motivated desires for the bad are possible. Suppose that some genuine desires are the product of decision or deliberation. This leaves open what sort of deliberation or decision produces those desires and what the basis of the agent's deliberation or decision is. For all that has been said, an agent might decide to perform some action and thus acquire a desire to act because she believes that the action is bad and not at all good. If the *reasonbad* thesis is correct, then agents can decide to act for the reason that their action is bad and not at all

---

<sup>20</sup> Nagel, *ibid.*, p. 30.

good to perform; if an agent acquires a desire to act as a result of her decision to act for the reason that her action is bad and not at all good, then she has acquired a motivated desire for the bad. Suppose, for example, that an agent engages in practical deliberation about what she is to do but becomes frustrated as a result of her inability to come to any practical conclusion. As a result of her frustration, she decides to perform an action that she believes utterly lacks any good consequences and will only result in bad consequences and she proceeds to punch a hole in the wall. Since my frustrated agent came to be motivated to punch the wall as a result of her decision or deliberation, her desire to punch the wall was a motivated desire for what she believes to be bad and not at all good. Thus, it appears that desires for the bad can be motivated.

It may be objected that, by Nagel's own lights, not just any deliberation or decision can produce motivated desires but only "rational" decision or deliberation. After claiming that a desire to shop for groceries upon discovering that the refrigerator is empty is a motivated desire, Nagel claims "*Rational*. . . explanation is just as much in order for that desire as for the action itself."<sup>21</sup> I admit it is unclear to me what Nagel means by rational deliberation; were that he said more.<sup>22</sup> Perhaps any decision or deliberation that produces a desire for the bad is not rational. If not, then even if an agent comes to be motivated by her non-rational decision or deliberation, it does not follow that her desire is motivated in Nagel's sense. Suppose that no rational deliberation could produce a motivated desire for what one believes to be bad and not at all good so that there could be

---

<sup>21</sup> Nagel, *ibid.*, p. 29, emphasis added.

<sup>22</sup> Wallace focuses upon Nagel's claim that motivated, but not unmotivated, desires are susceptible to rational explanation. Wallace argues that insofar as motivated, but not unmotivated desires, are susceptible to rational explanation, motivated desires can be rationalized by other propositional attitudes that an agent has; see Wallace, *ibid.*, p. 364.

no motivated desires for the bad. It would still be the case that desiring the bad is possible: even if there cannot be motivated desires for the bad, there still might be unmotivated desires for the bad. Even if it can be shown that motivated desires for the bad are not possible, the proponent of the *desirebad* thesis can retreat to the claim that all desires for the bad are unmotivated.

I conclude, then, that even if Nagel's motivated desires function as he thinks it does not follow that desiring the bad is not possible. Humeans and non-Humeans about desire can both claim that desiring the bad is possible and that the *desirebad* thesis is true.

### **Hedonic Accounts of Desire**

Desires, according to the Humean, are non-cognitive states that dispose an agent to bring something about. The kinds of desires Nagel is interested in have a particular kind of etiology that non-cognitive states disposing an agent to act might lack. Humeans like Smith and non-Humeans like Nagel disagree about what kinds of origins desires might have. It might be complained, however, that both accounts of desire miss something that is crucial to understanding the nature of desire. What is essential for something to be a desire, it might be claimed, is not that it disposes an agent to act nor that it has any particular kind of etiology. Instead, it might be claimed, it is essential to desiring that an agent who desires *P* is disposed to feel pleasure if it seems to her that *P* and perhaps that she is disposed to feel pain if it seems that not-*P*; this is a *hedonic* account of desire.<sup>23</sup>

John Stuart Mill appears to accept something like a hedonic account of desire in claiming “desiring a thing and finding it pleasurable, aversion to it and thinking of it as painful, are

---

<sup>23</sup> Timothy Schroeder discusses hedonic accounts of desire and calls them by this name in his *Three Faces of Desire*, pp. 27-35. I borrow his formulation, *ibid.*, p. 27.

phenomena entirely inseparable or, rather, two parts of the same phenomenon.”<sup>24</sup> More recently, Galen Strawson appears to accept something like a hedonic account of desire in claiming that “the primary linkage of the notion of desire to a notion other than itself is not to the notion of action or behavior but rather to the notion of being pleased.”<sup>25</sup> Neither Mill nor Strawson is guilty of advocating a kind of psychological hedonism according to which our only motive for doing anything is to derive pleasure or avoid pain. Mill and Strawson are only committed to the claim that it is a necessary condition of desiring something that one would feel pleasurable sensations, for example, upon satisfying the desire: desiring might only be necessarily correlated with pleasurable sensations according to hedonic accounts of desire. On hedonic accounts of desire, to desire something *is*, perhaps *inter alia*, to feel pleasure upon acquiring the object of desire or perhaps to be relieved of painful feelings. According to hedonic accounts of desire, at least one face of desire is its hedonic face.

It is far from clear that hedonic accounts of desire are correct, especially if ‘desire’ is understood as having the broad sort of extension that, for example, Humeans suppose that it has. No doubt if I desire to finish my dissertation on time I will at least enjoy the sensation of relief that sweeps over me when I submit my final draft, but I might also have a yen to do a cartwheel down the hall and feel neither pleasure nor pleasurable relief upon actually doing a cartwheel down the hall. I can also imagine creatures incapable of feeling either pleasure or pain, call them super-stoics, who nonetheless perform intentional actions; even super-stoics might want to do things on occasion. But suppose

---

<sup>24</sup> John Stuart Mill, *Utilitarianism*, p. 49.

<sup>25</sup> Galen Strawson, *Mental Reality*, p. 280.

that hedonic accounts of desire are correct and that it is part of the very nature of desire that a desiring agent feels pleasure or the absence of painful feelings upon satisfying a desire. Would the validity of hedonic accounts of desire imply that the *desirebad* thesis is false?

It might appear so. Most of us believe that there is something good about feeling pleasurable sensations and almost all of us believe that there is something good about being relieved of painful sensations. Therefore, it might appear that any agent who desires anything at all will believe that there is something good about obtaining the object of her desire: obtaining the object of her desire will produce pleasurable sensations or will relieve her of painful feelings and that is something good. However, we are rather easily talked out of the thesis that pleasure is always believed to be good. To adapt an example from Gary Watson, imagine a stoic monk who believes that his sexual desires are the work of the devil, that the very fact that he has sexual desires shows that he is corrupt to the bone, and that any pleasure he derives from satisfying his sexual desires is altogether bad.<sup>26</sup> It is far from clear that the monk must believe that there is something good about the pleasure that he derives from satisfying his sexual desires.<sup>27</sup> To borrow a locution from Watson, the monk does not believe that satisfying his sexual desires is represented by a positive entry, however small, on his “desirability matrix.”<sup>28</sup> Since the monk’s sexual desires are desires for what he believes is bad and not at all good and since he believes that the pleasure he would derive from satisfying his sexual desires is bad and

---

<sup>26</sup> Gary Watson, “Free Agency”, reprinted in *Free Will*, Watson, ed., p. 101.

<sup>27</sup> He may think there is something good about getting rid of his sexual desires, but that is surely not the same thing as satisfying his sexual desires and it is surely some other desire that leads him to rid himself of his sexual desires if he does.

<sup>28</sup> Watson, *ibid.*, p. 101.

not at all good, it seems to follow that the monk desires the bad. I conclude that even if hedonic accounts of desire are correct and it is necessary to desire something that one feels pleasurable sensations or the relief of painful feelings upon satisfying the desire, the *desirebad* thesis might still be correct.

I have surveyed Humean accounts of desire, Nagel's non-Humean account of motivated desires, and hedonic accounts of desire and I have concluded that none of them must be at odds with the *desirebad* thesis. There remains one account of desire that does pose a challenge to the *desirebad* thesis, however. I consider this account below.

### **Thomism about Desire**

In what follows, I shall refer to the following claim by Thomas Aquinas as Aquinas' thesis: "people desire things because they think them good."<sup>29</sup> In tribute to Aquinas, I shall refer to the large number of philosophers who endorse something like Aquinas' thesis as *Thomists* and I suppose that endorsement of something like Aquinas' thesis is *Thomism* about desire. The foundations of Thomism precede Aquinas himself: Aristotle claims that all desire aims at some good<sup>30</sup>, and Kant attributes a version of Aquinas' thesis to scholastic philosophers.<sup>31</sup> It is what more recent Thomists have had to say about desire that concerns me here.<sup>32</sup> In what follows, I investigate what a

---

<sup>29</sup> From *Aquinas: Selected Philosophical Writings*, Timothy McDermott, ed., p. 340.

<sup>30</sup> Aristotle, *The Nicomachean Ethics*, 1094a2.

<sup>31</sup> At 5:59 of the *Critique of Practical Reason*, Kant refers to the "old formula of the schools." In Mary Gregor's translation, the formula is rendered: "We desire nothing except under the form of the good; nothing is avoided except under the form of the bad."

<sup>32</sup> Elizabeth Anscombe, *Intention*; Warren Quinn, "Putting Rationality in its Place" reprinted in *Value, Welfare, and Morality*, Frey and Morris, eds., pp. 26-50; Thomas Scanlon, *The Importance of What We Owe Each Other*, chap. 1; Sergio Tenenbaum, "Accidie, Evaluation, and Motivation" from *Weakness of Will and Practical Irrationality*, Stroud and Tappolet, eds., pp. 147-71; Joseph Raz, "On the Moral Point of View" from *Reason, Ethics, and Society*, J. B. Schneewind, ed. pp. 70-2 and "Agency, Reason, and the

commitment to Thomism entails. It seems that if Thomism is correct, then the *desirebad* thesis is false: if desiring implies thinking that the object of desire is good, then desiring what one believes is bad and not at all good is impossible.

Note that Aquinas stops short of claiming we desire things because we believe them good. Contemporary Thomists have tended to follow Aquinas on this point. Tenenbaum suggests desiring is conceiving something to be good from some evaluative perspective.<sup>33</sup> Raz claims that what is desired must in some way be seen as good.<sup>34</sup> Stampe claims that what is characteristic of desire is that the object of desire seems to an agent as if the thing would be good.<sup>35</sup> But one can think that P, conceive that P is the case, one can see that P, and it can seem to one that P, without believing that . So Thomists, including Aquinas, are not committed to the thesis that desiring necessarily implies believing the object of desire is good, even if the Thomist is committed to claiming that desiring demands being in a belief-like state in which the object of desire is represented as being good. I shall use the term 'thinking' to refer to this cognitive state.

Note also that Aquinas' thesis is consistent with all of the following:

- **T1:** Necessarily, we desire  $\phi$  if we think  $\phi$  is good
- **T2:** Necessarily, we desire  $\phi$  only if we think  $\phi$  is good
- **T3:** Necessarily, we desire  $\phi$  if and only if we think  $\phi$  is good
- **T4:** A desire for  $\phi$  is identical to thinking  $\phi$  is good.<sup>36</sup>

---

Good" from his *Engaging Reason*, pp. 22-45; Dennis Stampe, "The Authority of Desire", *The Philosophical Review* 96, pp. 335-81.

<sup>33</sup> Tenenbaum, *ibid.*, p. 148.

<sup>34</sup> Raz, *ibid.*, p. 72.

<sup>35</sup> Stampe, *ibid.*, p. 356. See also Quinn, *ibid.*, p. 29.

<sup>36</sup> Compare Tenenbaum, *ibid.*, p. 148.

Different contemporary Thomists accept some or all of T1-T4. T4 is the strongest of all the above formulations of Aquinas' thesis, for only T4 implies that a desire is a cognitive state.<sup>37</sup> The thesis that desires are cognitive states is not without precedent<sup>38</sup> but it is a controversial thesis that is explicitly denied by Humeans, for example.<sup>39</sup> At any rate, T4 is only true if the weaker formulations of Aquinas' thesis are true. Similarly, T3 is stronger than either T1 and is only true if both T1 and T2 are. T1 amounts to a sufficiency condition for desiring and T2 amounts to a necessary condition. T1 cannot be constitutive of Thomism since T1 does not preclude the possibility of desiring the bad: that thinking something is good is sufficient for desiring does not imply that only thinking that something is good is sufficient for desiring. Further, T1 is plausibly regarded as false; thinking that something is good is not sufficient for desiring that thing. It is commonly supposed that an agent can fail to be sufficiently or appropriately motivated to pursue something the agent believes to be good, if the agent is weak of will or suffers from accidie and so forth.<sup>40</sup> The worry is not just that an agent suffering from extreme accidie might be too weakly motivated to actually pursue what she thinks is good, but rather that she might fail to be motivated to pursue it altogether. Desires are conative states. An agent who lacked any motivation to pursue  $\Phi$  would not be in a conative state with obtaining  $\Phi$  as its object. But an agent who lacked a conative state

---

<sup>37</sup> Tenenbaum, for example, claims that desiring should be *identified* with conceiving of something as good from an evaluative perspective. See Tenenbaum, *ibid.*, p. 157 and 158. See also Stampe, *ibid.*, pp. 358-9 where Stampe identifies desires with a kind of perception.

<sup>38</sup> See, for example, Jonathan Dancy, *Practical Reality* and Huw Price, "Defending Desire-as-Belief", *Mind* 98, pp.119-27.

<sup>39</sup> For arguments that desires must be distinct from beliefs and that cognitive states cannot also be conative states, see Smith, *ibid.*, p. 54.

<sup>40</sup> See Stocker, *ibid.*, pp. 741-6.

directed at  $\Phi$  would therefore lack a desire to pursue  $\Phi$ . If all this is correct, then I submit that if extreme cases of accidie are possible, then T1 is false just because an agent can fail to desire what she thinks is good. If T1 is false, then T3 and T4 are false as well.

This leaves T2 as the most plausible and least controversial version of Thomism. Admittedly, Aquinas' thesis is not identical to T2: Aquinas claims that people desire things because they think them good, and perhaps the 'because' of Aquinas' explanation is the 'because' of causality. T2 only states that thinking the object of desire is good is necessary for desiring and does not imply that there is a causal relation between thinking good and desiring. But T2 does imply that thinking good and desiring are necessarily correlated and if two events are not necessarily correlated then they do not necessarily stand in a causal relation. So, the truth of Aquinas' thesis depends upon the truth of T2. Thus, in what follows I shall be concerned with T2 and suppose that T2 at least partly constitutes Thomism about desire.

Unfortunately, some philosophers who appear to be Thomists about desire do not always seem to endorse T2. Quinn suggests that, of necessity, we evaluate the object of desire as good, and this suggests that Quinn accepts that thinking the object of desire is good is necessary for desiring.<sup>41</sup> However, Quinn also claims only that an agent may desire something that he evaluates as pleasant or interesting or advantageous or decent, for example.<sup>42</sup> Again, Raz claims that what is desired must in some way be seen as good, and this suggests that Raz accepts that thinking that the object of desire is good is

---

<sup>41</sup> Quinn, *ibid.*, p. 40.

<sup>42</sup> Quinn, *ibid.*, p. 40. Quinn claims that "to call an experience pleasant or unpleasant is already to bring it under an evaluative concept"; *ibid.*, p. 37. I wish Quinn had said more about this, for I am uncertain what it is to bring an experience under a concept, but for my part, I have no trouble imagining an agent who lacks the concept of goodness who nonetheless truly and sincerely calls an experience 'pleasurable.'

necessary for desiring. But Raz, apparently, does not claim that an agent who desires something must think that the thing is good. Instead, Raz only claims that an agent desires something only if she thinks that, roughly, the thing instantiates some property where that property is a good-making property.<sup>43</sup> Quinn and Raz thus appear to accept, not T2, but something like the following variant of T2:

- **T2\***: Necessarily, we desire  $\Phi$  only if we think  $\Phi$  is G, where G is some good-making property.<sup>44</sup>

I suggest that T2\* is not a viable Thomist alternative to T2. There are two difficulties awaiting the Thomist who endorses T2\*. First, T2\* appears to be consistent with the *desirebad* thesis, as I explain below. A philosopher who accepts the truth of the *desirebad* thesis seems to be forced to reject the thesis that thinking the object of desire is good is necessary for desiring, and thus a philosopher who accepts T2\* appears to be forced to reject T2 and thus to reject Thomism. Second, T2\* has the counter-intuitive implication that what is actually bad cannot be desired.

Consider my suggestion that T2\* is consistent with the *desirebad* thesis. T2\* only demands that an agent thinks, roughly, that this action instantiates some property G where G is a good-making property, but not that an agent thinks that this action is good. Suppose that being pleasurable is a good-making property. My ascetic monk might think that some act of fornication, D, is pleasurable, but also think that performing pleasurable actions is bad and not at all good. Since my monk does think that his act of fornication

---

<sup>43</sup> My evidence for this is contained in Raz's following response to Stocker: "Stocker is mistaken to think that I attribute a kind of high order reflectiveness to people: for example that they not only think of their actions as pleasurable or thrilling or beneficial . . . but also think of them as good in virtue of possessing those properties. All I ever claimed is that people act for considerations which we classify as a belief in the possession of a good making property." See Stocker's "Raz on the Intelligibility of Bad Acts", pp. 305-6, for discussion.

<sup>44</sup> I borrow this formulation of T2\* from a suggestion by Stocker, *ibid.*, p. 306.

instantiates a property that is in fact a good-making property, then according to T2\*, my monk can desire to perform the act of fornication. So it appears that, consistently with T2\*, my monk can desire what she believes is bad and not at all good. Hence, T2\* is consistent with the *desirebad* thesis. If the Thomist wishes to deny the *desirebad* thesis, her most obvious move is to reject T2\* and to revert to T2.

At any rate, Thomists have independent reason to reject T2\*: T2\* implies that agents cannot desire something that is actually bad, even if it is believed to be good. It is surely plausible that I can desire something that, in fact, instantiates no good-making properties and thus is not at all good insofar as I mistakenly believe there is something good about it. I might mistakenly believe that the drink in front of me is a gin and tonic when it is really a petrol and tonic. I might come to desire to drink insofar as I believe there is something good about gin and tonics, but in fact there is nothing at all good about the petrol and tonic that is actually before me since it will not quench my thirst and will surely make me sick. However, T2\* rules out the possibility of desiring something that is actually bad and not at all good, since T2\* states that it is necessary to desire something that it instantiates some good-making property. So the Thomist has independent reason not to endorse T2\*.

I have suggested that Thomists seem committed to claiming that the *desirebad* thesis is false. However, some Thomists who endorse T2 argue that we *can* desire the bad.<sup>45</sup> To explain the possibility of desiring the bad, some Thomists have noted we can have multiple evaluative thoughts about an object of desire. Drawing an analogy

---

<sup>45</sup> Tenenbaum, *ibid.*, p. 158. Scanlon suggests that it is possible to have a desire to purchase a new computer, for example, while one one's considered judgment is that there is no reason to purchase a computer; Scanlon, *ibid.*, p. 43. This suggests that Scanlon might accept that we could desire the bad.

between desire and belief is perhaps illustrative. It is possible to “think” that there is an oasis in the middle of a desert highway—that is, for it to seem as if there is an oasis in the middle of the highway—while also believing that there is not and could not be an oasis there.<sup>46</sup> Similarly, some Thomists claim it is possible to “think” that having a cigarette is good, for example, while also believing that having a cigarette would be bad and not at all good. An agent with conflicting evaluative beliefs is in a problematic mindset, but just so long as I have the thought that having a cigarette is good, I can desire what I *also* believe is bad and not at all good consistently with T2.

Perhaps Thomists can accommodate some instances of desiring the bad in the way suggested in the previous paragraph, such that the Thomist can perhaps allow for instances of weakly desiring the bad: agents who weakly desire the bad believe that the object of their desire is bad and not at all good, but they also think or conceive of the object of their desire as good in some aspect.<sup>47</sup> Thomists cannot, however, accommodate the possibility of strongly desiring the bad: agents who strongly desire the bad believe that the object of their desire is bad and not at all good and do not also conceive or think of the object of their desire as at all good. The paradigmatic instances of desiring the bad that I have discussed are instances of strongly desiring the bad. Carla, for example, believes that harming herself is bad and does not think of harming herself as at all good. Instances of strongly desiring the bad would be counter-examples to T2 since, by

---

<sup>46</sup> The word ‘think,’ as I employ it, is a term of art intended to denote a cognitive state with representational content that is not necessarily a belief. The sense of ‘think,’ as I employ it, is intended to include not only beliefs in its extension, but also cognitive states like seemings and conjectures and hypotheses and supposings and so forth. The word ‘thought,’ as I employ it, is a noun that denotes the mental state had by an agent who is thinking in the above sense.

<sup>47</sup> Actually, I suspect that some Thomists cannot even allow that weak desires for the bad are possible but the explanation depends upon discussing what Thomists claimed about the explanation and production of intentional action. I pursue this matter below.

definition, an agent who strongly desires the bad fails to think the object of her desire is good. The Thomist must therefore claim either that strong desires for the bad are not possible or that strong desires for the bad are not really desires properly understood. In what follows, I shall suppose that the *desirebad* thesis is a thesis about the possibility of strongly desiring the bad.

Thomists have claimed that any number of psychological states are not really desires properly understood even if they are “desire-like.” We speak not only of having desires, but also of having urges and impulses and cravings and so forth. Like desires, states like urges and impulses and cravings can be part of what produces and explains intentional action. However, for reasons to be discussed below, some Thomists have claimed that urges and impulses and cravings and the like are not desires, even if they are “desire-like.”<sup>48</sup> I shall refer to the class of desire-like states Thomists have distinguished from desires proper as faux desires. It is intuitively possible to have an urge for what one believes to be bad and not at all good. But if urges are faux desires and not desires proper, then that we can have an urge for the bad does not show that T2 is false, since T2 is a thesis about desires proper. Further, even if an urge is part of the explanation of why an agent acts, a Thomist might deny that urges rationalize our actions. I can surely satisfy my urge to have a cigarette by smoking intentionally and thus my urge to smoke is part of what explains my smoking intentionally. But a Thomist might point out that my urge to smoke is painful when not satisfied and claim that it is only insofar as I desire to rid myself of the pain that I smoke intentionally.<sup>49</sup> Thus, the Thomist might claim that what

---

<sup>48</sup> Raz claims that “it is wrong to take them [urges] as the basis for an analysis of wants and desires,” *ibid.*, p. 70.

<sup>49</sup> Raz, *ibid.*, p. 71.

rationalizes my smoking is my desire to avoid painful feelings, not my urge to smoke. So the Thomist can allow that we can have urges for what we believe to be bad and not at all good, and he can allow that urges can produce and explain intentional action, while denying that urges are desires proper.

It should not be surprising, then, if Thomists attempt to assimilate strong desires for the bad with faux desires rather than desires proper. Raz comes close to claiming that strong desires for the bad are faux desires and not desires proper when he claims that “some people. . . believ[e] that one can want anything, and not only what appears to one to be good or of value. This equates a desire for something with an urge for it which attacks one.”<sup>50</sup> If the Thomist can distinguish desires proper from faux desires and if strong desires for the bad are to be assimilated to faux desires rather than desires proper, then strong desires for the bad are not counter-examples to T2.

However, the Thomist must provide some basis for distinguishing faux desires from desires proper, and the Thomist must provide some reason for claiming that desires for the bad are faux desires rather than desires proper. Below, I examine how some Thomists have attempted to distinguish desires proper from faux desires. I suggest Thomists have not shown strong desires for the bad are faux desires rather than desires proper.

Scanlon suggests a means for distinguishing desires proper from faux desires. Scanlon is primarily interested in desire in “the directed-attention sense” such that to have a desire is to have one’s attention directed “insistently toward considerations that

---

<sup>50</sup> Raz, *ibid.*, p. 70

present themselves as counting in favor of” that thing.”<sup>51</sup> If I have a desire in the directed-attention sense for a glass of water, then it must be the case that my attention is insistently drawn to considerations that count in favor of drinking, such as the promised relief of a dry throat or the anticipated pleasure of drinking.<sup>52</sup> By contrast, if I suffer from pica and I find myself wanting to eat cigarette butts, there are no considerations that insistently present themselves to me as counting in favor of eating cigarette butts. But perhaps we are inclined to say that I do not have a desire to eat cigarette butts, but rather that I am in the grip of an urge or craving. If the genuine desire to drink a glass of water is paradigmatic of desires proper and the urge to eat cigarette butts is paradigmatic of faux desires, then perhaps there is a basis for distinguishing desires proper from faux desires: desires proper are marked by having my attention drawn insistently to considerations that count in favor of the object of my desire, while my attention is not so drawn when I have a faux desire.

Raz claims that the *absence* of a phenomenological feel is essential to desiring proper: desires proper, according to Raz, lack any “felt quality.”<sup>53</sup> Absent believing there is anything good about doing so, I may be “drawn” to count blades of grass, “attacked” by an urge to paint potatoes green, or “propelled by a force beyond my control” to stick my finger in a bit of goo. In all these cases, Raz would deny that I have a desire proper. Desires proper, according to Raz, emerge only as the result of a belief or belief-like state that there is something good about the object of one’s desire and “disappear” if the belief

---

<sup>51</sup> Scanlon, *ibid.*, p. 39. See also Quinn, *ibid.*, p. 36. Scanlon also claims that when I have a desire in the directed-attention sense, it must “seem” to me that there are reasons in favor of pursuing that thing which I desire; *ibid.*, p. 65.

<sup>52</sup> Scanlon, *ibid.*, p. 65.

<sup>53</sup> Raz, *ibid.*, p. 71.

is abandoned or lost.<sup>54</sup> Cognitive states need not be accompanied by a phenomenological feel, I suppose, so perhaps it is no surprise that desires proper are not accompanied by any phenomenological feel. By contrast, it is characteristic of urges and impulses and cravings and passions that it feels as though they “attack.” suggesting that urges and impulses and the like are not the product of an evaluative belief or belief-like state. But then, *ex hypothesi*, urges and impulses and the like are not desires proper.

It is controversial whether the phenomenology of desiring can reveal anything about the nature of desire. Some philosophers have rejected altogether the thesis that desires are or are necessarily correlated with any phenomenological feel.<sup>55</sup> Further, appealing to the phenomenology of some experience does not generally reveal the nature of what is experienced: when I act intentionally, I may feel as if it is up to me that I act as a result of my free will, but it does not follow that I have free will, much less that my intentional action depends on the exercise of my free will.

There is also a methodological problem with both Scanlon’s and Raz’s attempt to distinguish faux desires from desires proper. Both Scanlon and Raz note particular features of paradigm cases of both desires proper and faux desires and then suppose that those features will be present in all cases of desires proper and faux desires. But the instances that Scanlon and Raz generalize from may be anomalous or *sui generis* instances that are not representative of desires or urges or passions generally. At any rate, it is far from clear why desires for one’s children to prosper, to take bloody revenge,

---

<sup>54</sup> Raz, *ibid.*, p. 71. Actually, I have interpreted Raz here. Raz claims both that desiring something depends upon finding one’s desire to be “backed by reasons” and then that “Only what is seen in some way as good can be [desired]”; *ibid.*, pp. 71-2. It is puzzling how these claims are related and why they occur so closely unless Raz supposes that there is a relationship between believing that what is desired is good and finding the desire to be backed by reasons.

<sup>55</sup> Michael Smith has an excellent discussion of this point; see Smith, *ibid.*, pp. 45-9.

to attend a department meeting, and to make love must all have something in common, besides the fact that they are all desires. So even if the desire to take revenge or to make love has some phenomenological feel that the desire for one's children to prosper or to attend a department meeting lacks, it does not follow that they are not all desires proper.

In fairness, some Thomists have made clear that their claims about the nature of desire and desiring are restricted. Quinn claims that a favorable evaluation of the object of desire is “(of necessity) *typically* present in basic desire.”<sup>56</sup> Scanlon claims that having one's attention directed to considerations that count in favor of pursuing the thing is “essential *in the most common cases* of desire.”<sup>57</sup> But if Quinn and Scanlon only mean that a positive evaluation of the desired object is *typically* or *commonly* present in desiring, then they must concede that what is true of the token instances of desiring that they examine might not hold in all instances of desiring. Some Thomists have acknowledged that they are concerned with only a particular sense of ‘desire.’ Scanlon, for example, states that he is concerned with “what we ordinarily mean by” or the “commonsense notion” of desire.<sup>58</sup> Raz states that he is concerned with the “philosophical” sense and not the more “common” sense of desire.<sup>59</sup> If there is no disagreement about what the relevant intension of ‘desire’ is, then it should be no surprise that there is disagreement about what is included in the extension of ‘desire.’

There is no reason to suppose that desires for the bad, strong or otherwise, are always alike in the way that either Raz or Scanlon suppose. Following Nagel, two

---

<sup>56</sup> Quinn, *ibid.*, p. 40, emphasis added.

<sup>57</sup> Scanlon, *ibid.*, p. 38, emphasis added.

<sup>58</sup> Scanlon, *ibid.*, p. 38.

<sup>59</sup> Raz, *ibid.*, pp. 70-1.

different kinds of desires for the bad can be distinguished: desires for the bad can be motivated or unmotivated. Frankfurt's unwilling addict is perhaps an example of an agent who has an unmotivated desire for the bad since addictive desires do not typically arise from decision or deliberation.<sup>60</sup> Augustine appears to be an example of an agent who has a motivated desire for the bad. Suppose that Augustine decided to steal the pears because he wanted to perform a wicked action and believed that stealing the pears was a wicked thing to do: Augustine's desire for the bad would be motivated.

Recall Nagel's initial claim that unmotivated desires "assail us" whereas motivated desires do not, suggesting that unmotivated desires have a phenomenological feel that motivated desires lack. If there are motivated desires for the bad, then it will be difficult for the Thomist like Raz to assimilate desires for the bad to faux desires rather than desires proper given that motivated desires for the bad lack the phenomenology of unmotivated desires. Similarly, it does not seem implausible that some desires for the bad will involve one's attention being insistently drawn to the object of desire; indeed, it is not uncommon for self-mutilators like Carla to be unable to focus on anything else but the pursuit of their desire. So it will be difficult for Scanlon to claim that all desires for the bad are not desires proper since some desires for the bad do have the feel of desires proper. Nevertheless, Nagel's initial characterization of motivated and unmotivated desires is problematic: it seems that some desires that are the result of decision or deliberation are, or can be, accompanied by a phenomenological feel and that some desires that are not the result of decision or deliberation are not accompanied by any

---

<sup>60</sup> Frankfurt, "Freedom of the Will and the Concept of a Person" reprinted in *The Importance of What We Care About*, pp. 17-8. Note that Frankfurt himself provides an example of a motivated addictive desire, *ibid.*, p. 15.

phenomenological feel. Hence, appeals to the phenomenology of desire cannot be used to show that strong desires for the bad are not desires proper.

Here is another argument for assimilating strong desires for the bad with faux desires rather than desires proper: Raz argues that while we have urges and so forth, we do not normally endorse them.<sup>61</sup> Presumably, it is only desires proper that are endorsed. It is difficult to understand how an agent could endorse a strong desire for the bad just because, by definition, an agent who strongly desires the bad fails to think that the object of desire is at all good. So, if we do not endorse strong desires for the bad, then strong desires for the bad do not count as desires proper.

I am not entirely certain what Raz means by ‘endorse’ here, but Frankfurt has developed an account of endorsement in terms of identification.<sup>62</sup> Consider Frankfurt’s recent commentary on the matter:

These reflective attitudes of identification... are often based on or grounded in evaluations of desirability. However, they need not be. A person may identify himself with (or withhold himself from) a certain desire or motivation *for reasons that are unrelated to any such assessment, or for no reason at all*. . . . [t]he fact that he accepts it entails nothing, in other words, concerning what he thinks of it.<sup>63</sup>

To my ear, this suggests that Frankfurt believes an agent can identify with a strong desire for the bad; at least, for Frankfurt, agents are not precluded from identifying with strong desires for the bad even if there is no reason to identify with a strong desire for the bad. And if we can endorse strong desires for the bad, then that is some reason to suppose that strong desires for the bad are desires proper.

---

<sup>61</sup> Raz, *ibid.*, p. 71.

<sup>62</sup> Frankfurt developed his account of identification in a series of papers collected in his *The Importance of What We Care About*.

<sup>63</sup> Frankfurt, “Reply to Gary Watson” from *Contours of Agency: Essays on Themes from Harry Frankfurt*, Buss and Overton, eds., p. 160, emphasis added.

I have examined a number of Thomist arguments that strong desires for the bad are not desires proper and I have argued that Thomists have not yet succeeded in establishing their desired conclusion. However, Thomists have offered a more plausible argument that strong desires for the bad are not desires proper. Many Thomists have argued that if we accept that desires are essential to the explanation of intentional action and that the explanation of intentional action is rationalizing explanation, then it follows that desires for the bad cannot be desires proper. I shall refer to this Thomist argument as the Thomist's master argument.

### **The Master Argument**

The master argument for Thomism runs roughly as follows:

- 1) Desires have an essential role to play in the explanation of intentional action.
- 2) The explanation of intentional action is rationalizing explanation.

I have granted the truth of 1) and 2) and 3) is a straightforward consequence of 1) and 2):

- 3) Desires have an essential role to play in rationalizing explanations.

The fourth premise of the master argument constitutes the nub of Thomism about desire:

- 4) If desires are to play an essential role in rationalizing explanations, then, necessarily, an agent who desires something thinks that the object of her desire is good.

If it is part of the nature of desires proper that they should be able to produce and explain intentional action, then if a strong putative desire for the bad cannot produce and explain intentional action, then it cannot be a desire proper. 5) is a consequence of 3) and 4):

- 5) Necessarily, an agent who desires something thinks that the object of her desire is good.

5) is the conclusion of the master argument and the conclusion of the master argument just is T2). If the master argument is sound, then Thomism is vindicated and the *desirebad* thesis is false and agents cannot strongly desire the bad.

Premise 4) is the most contentious part of the master argument for Thomism. It will be useful to have a name for it in the following discussion: I shall refer to premise 4) as *the master premise*. If the master premise is false, then the master argument is not sound and its conclusion can be resisted. The remainder of this chapter is dedicated to resisting the master premise.

Typically, Thomists have offered some version of a *reductio ad absurdum* argument for the master premise: if we reject that, necessarily, an agent who desires something thinks that the object of her desire is good, then we will be unable to construct rationalizing explanations from an agent's desires and beliefs.

Quinn offers a relevant version of a *reductio* against non-Thomist accounts of desire.<sup>64</sup> Suppose we deny that, necessarily, an agent who desires something thinks that the object of her desire is good. Quinn imagines an agent who is disposed to turn on every radio he finds turned off; as noted above, Humeans may claim that Quinn's agent thereby desires to turn on radios. While Quinn's radio-turner-oner is disposed to turn on radios he finds turned off, he fails to believe there is anything good about turning on radios. But now it is unclear how the radio-turner-oner's purported desire, understood as a dispositional state, could possibly rationalize turning on radios. Quinn argues as follows:

A non-cognitive pro-attitude, conceived as a psychological state whose salient function is to dispose an agent to act, is just not the kind of thing that can rationalize. That I am psychologically set up to head in a certain way, cannot by itself rationalize my will's going along with the setup. For that I need the *thought*

---

<sup>64</sup> Quinn's target is often less than clear. At times, Quinn objects to moral non-cognitivism and moral anti-realism, to subjectivist accounts of value, and to Humean accounts of desire; see Quinn, *ibid.*, p. 31 where Quinn conflates all three.

that the direction in which I am psychologically pointed leads to something good. .  
<sup>65</sup>

Quinn intends to show that desires should not be understood simply as dispositional states to bring about some state of affairs, but Quinn also implies that strong desires for the bad cannot rationalize and therefore do not count as desires proper. Crucially, we cannot determine what it is about turning on radios that appealed to Quinn's radio-turner-  
 oner if we only know that he is disposed to turn on radios he finds turned off. Quinn suggests that what is necessary to construct rationalizing explanations in terms of an agent's desires and beliefs is just what is missing in the case of the radio-turner-  
 oner: if we are to construct rationalizing explanations from an agent's desires and beliefs, then it must be the case that, necessarily, an agent who desires something thinks that what she desires is good. But then we are committed to the truth of the master premise.

Two responses are in order. First, Quinn may be correct that crude functional states cannot rationalize action. It does not follow, however, that a desire proper is necessarily correlated with the thought that the object of desire is good. Given the assumption that desires do play a role in rationalization, Quinn leaves us to choose between two accounts of desire: Thomism and an account of desires as crude functional states underlying dispositions and tendencies, but the crude functional state account of desire fails and we are led to Thomism. But surely there are more subtle functional accounts of desire besides the crude account that Quinn considers. Thomists themselves might offer a functionalist account of desire.<sup>66</sup> Further, there are non-functional accounts

---

<sup>65</sup> Quinn, *ibid.*, p. 36.

<sup>66</sup> David Copp and David Sobel argue that Scanlon can be interpreted as being a functionalist about desire. See Copp and Sobel, "Desires, Motives, and Reasons: Scanlon's Rationalistic Moral Psychology", pp. 261-4.

of desire that are available to Thomists and non-Thomists alike. So the failure of a crude functionalist account of desire to explain how desires could rationalize action does not force us to Thomism.

Second, it does not follow that the radio-turner-oner's behavior *can* be rationalized even if the radio-turner-oner thinks that turning on radios is good. We may be unable to rationalize the radio-turner-oner's behavior because his behavior is not really *an intentional action*.<sup>67</sup> Quinn's agent appears to be obsessive or compulsive. We should not expect to be able to construct rationalizing explanations of, for example, tics and twinges, and perhaps we should also not necessarily expect to be able to construct rationalizing explanations of obsessive and compulsive behavior. If Quinn's radio-turner-oner does not perform an intentional action, then we cannot conclude that our inability to rationalize his behavior is due to a failure to accept the master premise. We should not expect rationalizing explanations of what cannot be rationalized and it is not clear that compulsively turning on radios deserves to be counted as intentional action.

Even if Quinn's particular example does not suffice to show Thomism is correct, Quinn may suggest an argument that Thomism must be correct. Recall that rationalizing explanations explain intentional action by revealing what it is about an action that appeals to the acting agent. Here is a piece of syllogistic reasoning that is, I think, implicit in Quinn's argument quoted above:

- i) An action is subject to rationalizing explanation only if there is something about that action that appeals to the acting agent.
- ii) An action appeals to the acting agent only if the acting agent thinks that there is something good about performing that action.
- iii) Therefore, an action is subject to rationalizing explanation only if the acting agent thinks that there is something good about performing that action.

---

<sup>67</sup> Copp and Sobel make a similar point, *ibid.*, p. 263.

If it were the case that an action appeals to an agent only if the agent thinks that there is something good about her action and if desires are essential to explaining actions, then there could be no case in which we cite a desire to explain an agent's action where an agent does not also think that the object of her desire is good: the consequent of the master premise would be true and the master premise would trivially follow, the master argument would be sound, and Thomism would be vindicated.

Other Thomists seem to endorse the syllogistic reasoning I attribute to Quinn. Here is Tenenbaum:

In a proper intentional explanation, the agent (or a third person) will be able to explain the point of engaging in such an activity; in other words, he will be able to explain what good he sees in the pursuit of this activity. On this view, a desire for an object as it typically appears in, for instance, an intentional explanation in the form of a belief-desire explanation must show what the agent found attractive in the choice of this action. But if the desire is not for something that one can intelligibly conceive to be good, or if it is not for something the *agent* conceives to be good, we would not know what point the agent could see in such an action and we would therefore not have made the agent intelligible.<sup>68</sup>

Unfortunately, Tenenbaum speaks of *conceiving* of something as good rather than thinking it to be good, and of finding something *attractive* rather than finding appealing. Here is what I think is a charitable reconstruction of Tenenbaum's argument, however. It does not seem implausible that the point of an agent's action is to bring about something the agent found appealing about her action. But, according to Tenenbaum, knowing the point of an action is a matter of knowing what an agent conceives of as good about the action and thus an action cannot have a point and cannot be rationalized unless the agent conceived of her action as good in some way. Thus, on this reconstruction of

---

<sup>68</sup> Tenenbaum, *ibid.*, p. 150.

Tenenbaum, if we are to explain action in terms of an agent's desires, then we must suppose that there is something about the object of her desire that she thought was good.

Raz too echoes some sympathy with the thesis that an action can appeal to an agent only if the agent thinks there is something good about what is desired:

. . . typical intentional actions are actions about which their agents have a story to tell (i.e., actions manifesting an internal viewpoint about what one is doing, or is about to do), a story that explains why one acted as one did. Moreover, and this point is crucial, the explanation makes intelligible not only why the action happened; it makes it intelligible as an action chosen. . . the "story" is of what the agent took to be facts which show the act to be good and which therefore constitute a reason for its performance. . .<sup>69</sup>

If the "story" about intentional actions is rationalizing explanation, then Raz can be read as asserting that rationalizing explanations must cite an agent's thought that there are facts which show an action to be good. If Raz supposes that desires must be part of the "story," then Raz accepts that an action can be rationalized only if the acting agent thinks that there is something good about what she desires.

The proponent of the *desirebad* thesis must resist the reasoning I attribute to Quinn, Tenenbaum, and Raz. In particular, the proponent of the *desirebad* thesis must resist ii), the thesis that an action can appeal to an agent only if the agent thought that there is something good about performing the action. In what follows, I shall offer a diagnosis of why some philosophers have supposed that only what is thought good can appeal and thus why only what is thought good can be desired. I shall then argue that the believed badness of something can be what appeals to an agent. So even if an action must appeal to the acting agent if the action is to be rationalized, it does not follow that the acting

---

<sup>69</sup> Raz, "Agency, Reason, and the Good", p. 24.

agent must think that her action is at all good. I then consider implications of my proposal for the master premise, the master argument, and for Thomism.

Unfortunately, I think Davidson is partly responsible for the popularity of the thesis that only what is thought good can appeal. Davidson claims, for example, that desiring something entails “holding” it to have “some positive characteristic,”<sup>70</sup> that desires “constitute” value judgments,<sup>71</sup> and that desiring something involves “setting a positive value” on the object of desire.<sup>72</sup> It is rather easy to conclude that Davidson must believe that the only way to hold that the object of desire is valuable or has some positive characteristic is if the object of desire is believed good. However, Davidson explicitly claims that desires “must not be taken for convictions, however temporary, that every action of a certain kind ought to be performed, is worth performing, or is, all things considered, desirable.”<sup>73</sup> For example, Davidson claims that “a man may all his life have a yen. . . to drink a can of paint, without ever, even at the moment he yields, believing that it would be worth doing.”<sup>74</sup> It is true that Davidson suggests there is some correlation between desiring *P* and making an incipient value judgment about *P*, but it does not follow that Davidson is committed to the thesis that desiring *P* implies thinking that *P* is good.<sup>75</sup> It is surprising that while Quinn and Tenenbaum and Raz appeal to

---

<sup>70</sup> Davidson, “Intending”, p. 97, n. 7.

<sup>71</sup> Davidson, *ibid.*, p. 102.

<sup>72</sup> Davidson, “How is Weakness of the Will Possible?”, p. 31.

<sup>73</sup> Davidson, “Actions, Reasons, and Causes,” p. 4; Davidson actually claims that “pro-attitudes” must not be confused with convictions that the object of desire is desirable, but again, since desires are a sub-class of pro-attitudes for Davidson, what holds for pro-attitudes should hold for desires.

<sup>74</sup> Davidson, *ibid.*, p. 4.

<sup>75</sup> I thus disagree with Velleman’s analysis of Davidson in his otherwise excellent “The Guise of the Good”, pp. 103-5.

something like Davidson's thesis that the explanation of action is rationalization, Davidson himself is not clearly a Thomist.<sup>76</sup>

Of course, Davidson might be wrong, or rather Davidson on my reading of him might be wrong, and Thomists might be right that, necessarily, desiring implies thinking good. The proponent of the *desirebad* thesis still requires an account of how what is desired can appeal to an agent when the object of desire is believed to be bad and not at all good. Surprisingly, Thomists themselves have offered an explanation of why what is believed to be bad and not at all good can appeal to an agent.

### **Desiring the Bad and Intelligibility**

Anscombe claims that what an agent who genuinely desires something must be able to provide an *intelligible* desirability characterization of the object of desire.<sup>77</sup> Some Thomists have echoed Anscombe's interest in intelligibility: Tenenbaum claims that only desires for what can intelligibly be conceived of as good can rationalize action<sup>78</sup> and Raz claims that the believed goodness of a desired action makes that action "an intelligible object of choice."<sup>79</sup> It seems false, however, that only what is believed good can intelligibly appeal to an agent or be an intelligible object of choice, at least given ordinary usage of 'intelligible.' I submit that given ordinary usage of 'intelligible,' something is intelligible just in case it is *understandable*.<sup>80</sup> Admittedly, it is hard to understand how

---

<sup>76</sup> It will not do to claim that yens are not desires, especially if 'desire' is understood in the broad sense that the Humean understands it. It is certainly plausible that yens can rationalize action, but then yens deserve to be called desires by the Thomist's own lights.

<sup>77</sup> Anscombe, *ibid.*, p. 75.

<sup>78</sup> Tenenbaum, *ibid.*, pp. 150-1.

<sup>79</sup> Raz, *ibid.*, p. 24.

<sup>80</sup> My computer's dictionary and [www.dictionary.com](http://www.dictionary.com) support this reading of 'intelligible.'

an agent could be motivated to pursue something she believed to be bad and not at all good, if all that we are told is that she believes the thing to be bad and not at all good, but it does seem understandable how an agent can desire what she believes to be bad and not at all good if we *also* know that an agent is, for example, sufficiently frustrated or depressed. As I argue above, it is not implausible to suppose that desiring the bad is constitutive of or the causal product of certain emotions or moods. As I am now, I find it difficult to understand how anyone could want to mutilate their body or to leap from a precipice onto the rocks below, but I can understand how a sufficiently depressed or frustrated agent might. Understanding how an agent could desire something is not necessarily a matter of imagining myself, as I am right now, being motivated to pursue the thing. Instead, understanding how an agent could desire something may require imagining what I would want if I were in the grip of some mood or emotion. Similarly, even if it is unintelligible to me how anyone who is in the same sort of psychological state as I am right now could desire a particular thing, it does not follow that desiring the thing is unintelligible. Given the appropriate moods and emotions, agents can be in psychological states that can rationalize actions that, absent those psychological states, could not be rationalized.

It might be objected that my appeal to emotions and moods to explain how desiring the bad is possible and how acting on a desire for the bad is intelligible is suspect. For example, I have accepted and the master argument is committed to the thesis that the explanation of intentional action is rationalizing explanation. Further, the paradigmatic schema of rationalization is to explain an action in terms of an agent's desires and beliefs. If desire-belief explanations do adequately explain their *explananda*, then it seems that

desire-belief explanations are sufficient for explaining intentional action. And insisting that desire-belief explanations are sufficient for explaining intentional action suggests that there is no need to suppose that any *further* psychological states are relevant to the explanation of action. If so, my appeal to emotions and moods to explain, for example, desiring the bad is unnecessary and *ad hoc*.

However, taking desire-belief explanations too seriously might lead us to suppose that explanations of action citing some relevant desire and belief pair are interesting and complete. Often, neither is the case. Merely citing a desire and instrumental belief need not be very interesting; it may only indicate that the agent wanted to act in a certain way, but claiming that an agent acted “because she wanted to” is an otiose explanation.<sup>81</sup>

Further, citing only a desire and belief pair in a rationalization neglects other psychological, sociological, and historical factors that normally would also be relevant, as I suggest below. It is intuitively plausible, as Michael Stocker has argued, that psychological states like moods and emotions are relevant to the explanation of action.<sup>82</sup>

Consider what it is like to act out of frustration: suppose that a sufficiently frustrated agent intentionally punches a hole in his living room wall. It is perhaps true of an agent who intentionally punches a hole in his wall that he wants to punch a hole in his wall and believes that by performing that particular action, he will punch a hole in his wall.

However, an explanation of the frustrated agent’s act of punching a hole in his wall just

---

<sup>81</sup> Following Davidson, we might claim that it is not enough to know that some action appealed to an agent if we are to explain that action; we must know what it is about the action that appealed to her and led her to act; *ibid.*, p. 3.

<sup>82</sup> See for example, Stocker, “Desiring the Bad; An Essay in Moral Psychology.” Stocker’s thesis that motivation might not follow from the judgment an action should be performed because certain moods or emotions might interfere or prevent the judgment from being motivationally efficacious is often noted. What is not always noted is Stocker’s claim that moods and emotions are *necessary* for motivation, that it is only given a host of moods and emotions that agents are *ever* motivated to act.

in terms of this desire and belief pair does not allow us to distinguish the frustrated agent from an agent who just happens to enjoy punching holes in things. The agent who enjoys punching holes in things also wants to punch a hole in his wall and believes that by performing that particular action, he will punch a hole in his wall. Explaining the two agent's actions only in terms of their desires and beliefs leaves us unable to distinguish two very different actions of two very different agents. Thus, desire-belief explanations need not be complete.

We have good reason then to suppose that desire-belief explanations of intentional actions do not necessarily adequately explain their *explananda*. But, I suggest, if we allow that various kinds of moods and emotions are relevant to the production of intentional action, then we have some reason to believe that desiring the bad is possible: it might be constitutive of certain moods and emotions that agents suffering from them desire the bad and desiring the bad might be the causal product of certain moods and emotions. First, consider Freud's discussion of "melancholia":

The distinguishing mental features of melancholia are a profoundly painful dejection, cessation of interest in the outside world, loss of the capacity to love, inhibition of all activity, a lowering of the self-regarding feelings to a degree that finds utterance in self-reproaches and self-revilings, and culminates in a delusional expectation of punishment.<sup>83</sup>

If Freud is correct that depression manifests itself in self-reproach and self-reviling, then it would not be surprising that agents suffering from depression desire things for themselves that they believe are bad and not at all good. After all, we typically desire things that are bad and not at all good for persons that we revile. Thus, it would not be

---

<sup>83</sup> From volume 14 of *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, James Strachey, ed., p. 244. I take the quote from Michael Stocker with Elizabeth Hegeman, *Valuing Emotions*, p. 245; Stocker and Hegeman equate depression and melancholia.

surprising if desiring the bad is constitutive of suffering from depression or melancholia. At least with respect to some cases of depression, wanting the bad *is* part of *what it is* to be depressed.

Second, suffering from some emotion or mood might *cause* an agent to desire what she believes is bad and not at all good. If emotions or moods like depression give rise to desires for what one believes to be bad and not at all good, then it is no surprise that agents who suffer from these moods and emotions find it so difficult to feel any better: moods and emotions like depression might be re-enforcing. For example, an agent's frustration may cause him to acquire and act upon desires whose fulfillment only furthers his frustration. An agent who is already frustrated and acts on a desire to punch a hole in his wall is left with a hole in his wall, a broken hand, and little left to show for his efforts.<sup>84</sup> Moods and emotions like frustration can perhaps give rise to desires that agents who are not in the grip of those moods and emotions do not suffer from.

For present purposes, it is important to recall that even if desire-belief explanations of actions are the paradigmatic means for explaining intentional action, there are other elements of our psychological economy that are enlightening and interesting that can be invoked in the explanation of action. Emotions and moods have considerable influence on what an agent desires and what she is motivated to do, and thus they ought not to be neglected in explanations of intentional action. I complain about desire-belief explanations because I believe that taking them *too* seriously could lead Thomists to mistaken conclusions about what desires *must* be.

---

<sup>84</sup> While discussing a similar sort of case, David Velleman remarks: "Someone who smashes crockery in order to feel better didn't feel all that bad to begin with." See Velleman, "The Guise of the Good," p. 121.

Return, then, to Carla. Again, Carla desires to mutilate herself and believes that mutilating herself is bad and not at all good. Carla may be able to offer a desirability characterization of the action she desires to perform insofar as the believed badness of self-mutilation appeals to her. Is this an intelligible desirability characterization, given Anscombe's concern? I submit that it is *understandable* how a sufficiently depressed and frustrated agent could find an action that harms herself to be appealing. Suppose that Carla acts upon her desire to mutilate herself. Is it intelligible to choose to self-mutilate one's own body, given Raz's concern? I submit that it is *understandable* how a sufficiently depressed and frustrated agent could choose to mutilate herself. Thus, if there is the relationship between intelligibility and being understandable that I have suggested, then what is believed to be bad and not at all good can intelligibly appeal to an agent and be an intelligible object of choice. Of course, the sense of 'intelligible' that Anscombe and Tenenbaum and Raz appeal to may not be the ordinary sense I make use of. But then the Thomists must offer some other sense of 'intelligible' that they mean to appeal to that does not beg the question against the proponent of the *desirebad* thesis.

But now, it seems to me, a number of Thomist theses fall apart. Given that agents suffer from the appropriate moods and emotions, agents can intelligibly desire what they believe to be bad and what they do think is not at all good, and given the appropriate moods and emotions, what is believed to be bad and thought to be not at all good can appeal to an agent. Premise ii) of Quinn's syllogism is therefore false. The master premise of the Thomist's master argument also appears to be false since desires for the bad can be a part of rationalizing explanations. The master argument is therefore not sound and we are not forced to accept its conclusion. Finally, since strong desires for the

bad are possible then T2) is false: it is not the case that, necessarily, we desire something only if we think that thing is good. We have grounds, I submit, for rejecting Thomism about desire left to right.

### Conclusion

I have argued that the *desirebad* thesis is true. I also have argued that a number of considered accounts of desire do not preclude the possibility of desiring the bad. While Thomism about desire precludes strongly desiring the bad, Thomism about desire is mistaken. The account of desire that I would favor is therefore not *Thomist*, but *Augustinian*. Indeed, Augustine appears to be a counter-example to Aquinas' thesis about desire, insofar as Augustine appears to desire, at least some things, because he thinks that they are bad and not at all good.

I say that the account of desire that I *would* favor is Augustinian and would allow that desiring the bad is possible. I shall not provide a full-blown account of desire and desiring here. Just what desires are remains an elusive question. However, we have good reason to believe that our best accounts of desire will be consistent with the possibility of desiring the bad, and thus it is a necessary condition for any account of desire and desiring to be plausible that they be consistent with the *desirebad* thesis. In subsequent chapters, I argue that the *intentionbad* and *reasonbad* theses are also true: agents can desire what they believe to be bad and not at all good, agents can intentionally pursue what they believe to be bad and not at all good, and agents can act intentionally for the reason that what they pursue is believed to be bad and not at all good. The truth of the *desirebad* thesis does not ensure the truth of the *intentionbad* or *reasonbad* theses, but we have perhaps removed one obstacle for accepting the latter two theses.

CHAPTER 3  
INTENTION AND INTENDING THE BAD

I'm behaving badly and I'm going to go on behaving badly. This is a situation where people do behave badly.

Graham Greene, *The Quiet American*<sup>1</sup>

**Introduction**

In the previous chapter, I discussed desire and desiring and I argued that the most plausible accounts of desire are consistent with the truth of the *desirebad* thesis such that agents can desire what they believe is bad and not at all good. I have suggested that the truth of the *desirebad* thesis makes it plausible to suppose that the *intentionbad* and *reasonbad* theses are also true. Yet I concede that the truth of the *desirebad* thesis does not ensure the truth of either the *intentionbad* thesis or the truth of the *reasonbad* thesis, for reasons to be discussed below. In this chapter, I begin to defend the *intentionbad* thesis that agents can intentionally pursue what they believe to be bad and not at all good.

I assumed in the previous chapter that desire underlies intentional action, insofar as I assumed that desire is essential to the production and explanation of intentional action. But this assumption is potentially misleading. While desire plausibly plays a role in the production and explanation of intentional action, intention also plausibly plays a role and it is controversial whether or not intentions are desires, as I discuss below.<sup>2</sup> It is widely

---

<sup>1</sup> Graham Greene, *The Quiet American*, p. 147.

<sup>2</sup> Humeans might count intentions as desires, in the same way that the Humean would count wishes and hopes and so forth as desires, since intentions surely have a world-to-mind direction of fit, rather than the mind-to-world direction of fit had by beliefs, for example. However, many philosophers have supposed

conceded by philosophers of action that if an agent fails to intend to perform some action, the agent cannot act intentionally.<sup>3</sup> Therefore, even if the *desirebad* thesis is true, it does not follow that the *intentionbad* thesis is true: even if agents can desire what they believe is bad and not at all good, it does not follow that agents can intentionally pursue the bad, just because it has yet to be shown that agents can *intend* the bad.

The definition of ‘intending the bad’ parallels the definition of ‘desiring the bad’ that I have worked with in previous chapters: agents who intend the bad intend to perform an action  $\phi$  that the agent believes is bad and not at all good. In this chapter, I argue our best understanding of intention is consistent with the possibility of intending the bad. Broadly, I offer two arguments in this chapter: first, I argue that there is nothing about the nature of intention that precludes us from supposing that agents can intend the bad, and second, I argue that if intending the bad is possible then we have strong reason to suppose that intentionally pursuing the bad is possible.

### **The Function of Intention and the Nature of Intention**

Many philosophers interested in understanding what intention is have focused on the functional role that intentions play in practical deliberation and action. If we are clear about the functional roles of intention, then we are in better to understand the nature of intention. It is widely conceded among philosophers of action that, for example, intentions supply motivation. However, it is a mistake to suppose that the only function of intention is to supply motivation. It has been argued that intentions also have a

---

that there is some difference between intentions and the state we typically refer to as ‘desire’, as I discuss below.

<sup>3</sup> This thesis supposes that behavior counts as an intentional action only if the behavior is suitably related to an intention. Michael Bratman calls this “the single phenomenon view”; see Bratman, *Intention, Plans and Practical Reason*, p. 112 and “Two Faces of Intention” reprinted in *The Philosophy of Action*, Mele, ed., p. 180.

*sustaining* function.<sup>4</sup> Actions are datable events and they rarely, if ever, occur in an instant or within a single-time slice; most actions, if not all, take time to unfold. Given that intentions have a motivating function, they can cause or otherwise initiate action-events, but intentions also make it the case that the action-event actually obtains after it has been initiated. Imagine a weightlifter attempting to clean-and-jerk a heavy weight: his action is complex; it proceeds in stages, and takes time to perform. Suppose that the weightlifter initially intends to lift the weight above his head. If, in the middle of his lift, he loses or abandons the intention, then, other things being equal, we would expect him to cease to his attempt to lift the weight.<sup>5</sup> Since his intention must be present if he is to actually intentionally lift the weight, it is plausible to suppose that his intention plays a role in sustaining his attempt to lift the weight. Other things being equal, an agent who intends to  $\Phi$  is disposed to bring  $\Phi$  about once the intention to  $\Phi$  has been executed, supposing  $\Phi$  has not obtained immediately after the intention is executed.

Many philosophers have also supposed that intentions have a *guiding* function.<sup>6</sup> An intention guides an intentional action only if the content of the intention to  $\Phi$  itself figures appropriately, in a sense to be explained, in the etiology of  $\Phi$ -ing.<sup>7</sup> I may intend to act and develop a complicated plan as part of my intention to act. If I come to believe that my behavior does not match the behavior I intended to perform, then, other things

---

<sup>4</sup> Al Mele, *Springs of Action*, p. 130.

<sup>5</sup> The example is mine, but the point is made by Mele, *ibid.*, pp. 130-1.

<sup>6</sup> See, for example, John Bishop, *Natural Agency*, pp. 167-72; Myles Brand, *Intending and Acting*, p. 175; Irving Thalberg, "Do Our Intentions Cause Our Intentional Actions?", *American Philosophical Quarterly* 21, pp. 249-60; Al Mele, *ibid.*, pp. 135-7.

<sup>7</sup> Mele, *ibid.*, p. 136.

being equal, if my intention performs a guiding function, I will cease my deviation and take steps to bring about the intended result in the way I had planned. If I intend to walk to my office on campus and if my intention serves a guiding function, then I will have at least an incipient plan to walk to my office and if I come to believe that I am deviating from my plan, I will take steps to correct my deviance and bring it about that I actually do walk to my office. Importantly, it is by considering my initial intention and corresponding plan to walk to my office that I determine that I have deviated from my intention. An intention serves a guiding function insofar as the intention disposes an agent to bring about the intended result in the intended way, partly because of the content of the intention itself.

It is also widely conceded by philosophers of action that intentions have a coordinative function.<sup>8</sup> Bratman, for example, claims that intentions are typically elements in coordinating plans<sup>9</sup> and that plans are intentions *writ large*.<sup>10</sup> Intending to perform an action enables an agent to coordinate her action and reasoning, as well as the actions and reasoning of others, to ensure that the intended result does actually obtain.

Bratman identifies three further features of intentions that are crucial for intentions to perform motivating, sustaining, guiding, and coordinating functions. Bratman's argues that three kinds of related dispositions are associated with intentions: intentions are "conduct-controlling," they have a characteristic "stability" or "inertia," and they play

---

<sup>8</sup>Gilbert Harman, "Practical Reasoning", *Review of Metaphysics* 79, pp. 431-63; Bratman, *Intention, Plans and Practical Reason*, "Two Faces of Intention", and the essays collected in his *Faces of Intention*; Mele, *ibid.*, pp. 137-8.

<sup>9</sup> Bratman, *Intention, Plans, and Practical Reason*, p. 111.

<sup>10</sup> Bratman, *ibid.*, p. 8.

characteristic roles in practical reasoning.<sup>11</sup> To say that intentions control conduct is to say that intentions do more than merely influence conduct. Desires might influence conduct insofar as an agent who desires something is at least somewhat motivated to pursue that thing. Yet it is consistent with desiring something that, other things being equal, an agent does not act if the time and the opportunity come and is not irrational in failing to act. However, an agent who intends to perform some action will normally at least try to act if the time and opportunity come up.<sup>12</sup> To say that intentions are stable or have a characteristic inertia is to say that agents who intend to  $\phi$  are, as a result of intending to  $\phi$ , disposed to resist reconsidering or abandoning their intention to  $\phi$ .<sup>13</sup> Normally, an agent who has intended to  $\phi$  has effectively settled the question of what is to be done, since intending to  $\phi$  disposes an agent not to reconsider the matter. To borrow from a point made by Michael Robins, once we have intended to perform some action, we have answered affirmatively the question “Shall I perform this action?”<sup>14</sup> Finally, to say that intentions play characteristic roles in practical reasoning is to say that an intention to  $\phi$  has at least two effects on an agent’s future deliberations. First, an intention to  $\phi$  disposes an agent to both reason from the intention to  $\phi$  to other distinct intentions that will lead to her  $\phi$ -ing.<sup>15</sup> If I have already intended to travel home for the holidays, my intention to travel home may, for example, lead me to intend to purchase tickets for a flight home. Second, an intention to  $\phi$  disposes an agent to constrain future

---

<sup>11</sup> Bratman, *ibid.*, p. 22.

<sup>12</sup> Bratman, *ibid.*, pp. 15-6.

<sup>13</sup> Bratman, *ibid.*, p. 17.

<sup>14</sup> See Robins, *Promising, Intending, and Moral Autonomy*, p. 21.

<sup>15</sup> Bratman, *ibid.*, pp. 16-7.

intendings in light of the intention to  $\phi$  by disposing an agent to make her intention to  $\phi$  consistent with her other intentions and with her beliefs.<sup>16</sup>

Here then are the beginnings of an account of the nature or essence of intention: an intention is a conative psychological state that motivates an agent to act but also serves sustaining and guiding and coordinating functions, such that intentions control rather than merely influence behavior, have characteristic stability and inertia, and dispose an agent to reason in particular ways. If there is some psychological state that serves all of these functions and is correlated with all of these dispositions, then we have license to claim, I suggest, that the psychological state just *is* an intention.

Note, however, that if we accept the above account of the nature of intention, then we have reason to suppose that agents *can* intend the bad. Suppose that Carla is in some psychological state, *M*, that motivates Carla to mutilate herself and she believes that mutilating herself is bad and not at all good. Suppose also that *M* sustains and guides and coordinates Carla's behavior and practical deliberation, that *M* controls rather than merely influences her conduct, that *M* is both stable and has inertia, and that *M* disposes her to form intentions in light of *M* that are also consistent with *M*. If the above account of the nature of intention is correct, then we have license to suppose that *M* just *is* an intention. But since Carla intends to perform an action that she believes is bad and not at all good, it follows that Carla intends the bad.

In this section, I have offered an account of intention in terms of the various functions that intentions perform. Further, the account of intention that I offer accords with accounts offered by other philosophers. Finally, the account of intention that I have

---

<sup>16</sup> Bratman, *ibid.*, pp. 17.

offered permits intending the bad. Yet while I have perhaps identified part of what is necessary for some psychological state to count as an intention, there might still be more to intending that precludes intending the bad. Thus, we must continue investigating the nature of intention. The investigation will continue as follows. First, I consider a popular account of intention where intentions are understood as reducible to a certain kind of desire and belief pair. Intending the bad is possible and easily explicable if desire-belief reductions of intention are correct. However, desire-belief reductions of intentions are problematic, although perhaps not fatally flawed. Worse, the arguments that lead us away from desire-belief reductions of intention seem to imply that intending the bad is not possible, for reasons I explain. Nonetheless, I consider two other attractive accounts of intention and argue that they do permit intending the bad. Thus, I suggest, for a number of popular and attractive accounts of intention, it is possible for agents do intend the bad.

### **Desire-Belief Reductions of Intention**

I intend to argue that plausible accounts of intention do allow that agents can intend the bad. There is a familiar disagreement in the literature on intention that is relevant to our investigation of the nature of intention: philosophers have often disagreed whether intentions are reducible to more basic psychological states, such as desires or beliefs, or whether intentions are irreducible to other psychological states.<sup>17</sup> In this section, I argue

---

<sup>17</sup> We might also be error theorists about intention and suppose that the English word ‘intention’ fails to refer to anything at all. Gilbert Ryle ridiculed the thesis that volitions played any role at all in the production of voluntary action; see Ryle, *The Concept of Mind*. David Velleman has supposed that Ryle similarly ridicules the thesis that intentions play a role in the production of intentional action; Velleman, *Practical Reflection*, p. 109, fn. 2. Anscombe suggests that “an action is not called ‘intentional’ in virtue of any extra feature which exists when it is performed”; *Intention*, p. 28. But then there is little reason to suppose that there are intentions at all. Donald Davidson, at one point claimed that ‘intention’ is “syncategorematic and cannot be taken to refer to an entity, state, disposition, or event”; “Actions, Reasons, and Causes”, p. 8. Davidson later retracted this view in his “Intending”. I shall simply suppose that

that intentions are not reducible to other psychological states; in particular, that intentions are not reducible to pairs of desires and beliefs. The difficulty for the proponent of the *intentionbad* thesis is that some arguments for the irreducibility of intention appear to imply that intending the bad is not possible. I resolve the apparent difficulties in later sections; for now, the proponent of the *intentionbad* thesis ought to ally herself with the best account of intention available.

According to one familiar and rather simple account of intention, an intention just is a desire paired with a belief about how to satisfy that desire: I shall refer to this account of intention as the *desire-belief* account.<sup>18</sup> For example, an agent's intention to tune a piano, according to desire-belief accounts, is reducible to a desire to tune a piano and a belief that an action will result in a piano being tuned. The proponent of the *intentionbad* thesis should be initially attracted to this desire-belief account of intention; if we accept that agents can desire the bad, as I argued in the previous chapter, then since desire-belief accounts suppose that intentions are just are suitably paired desires and beliefs, desire-belief accounts of intention can explain how agents can intend the bad. So if the *desirebad* thesis is true and if desire-belief accounts of intention are correct, then agents can intend the bad. Would that things were so simple.

While the prospect of reducing intentions to desire and instrumental belief pairs has struck an impressive number of philosophers as plausible, many philosophers have argued the project of reducing intentions to desire and instrumental belief pairs is doomed

---

'intention' does actually refer to psychological states that frequently obtain and thus I shall dismiss outright the error theorist account of intention.

<sup>18</sup> See Audi, "Intending", *Journal of Philosophy* 70: 387-402, and "Intending, Intentional Action, and Desire", in Marks, ed., *The Ways of Desire*.

to failure.<sup>19</sup> Philosophers who reject desire-belief accounts have tended to claim intentions are subject to constraints of rationality that desire and belief pairs are not subject to. For example, while there is nothing irrational about desiring two mutually exclusive ends and having beliefs about how to bring about those ends, it is not clear one could rationally intend to bring both ends about.<sup>20</sup> Further, intending to perform an action seems to commit an agent, other things being equal, to acting when the time comes. Yet an agent who has some desire and belief pair need not be committed, other things being equal, to acting when the time comes. Thus, since intentions are not identical to desire paired with beliefs about how to satisfy that desire, intentions cannot be reduced to desire and instrumental belief pairs.

I shall have more to say about rationality and intention, but consider some arguments that push us away from simple desire-belief accounts of intention. Suppose, for example, that I have a desire to remove Smith, a prominent politician, from office and I believe that if I assassinate Smith, I will thereby remove him from office. Do I thereby *intend* to assassinate Smith? I think our intuitions suggest I do not. First, note that while many of us have desires for ends that are morally innocent, we believe that we can satisfy those desires by performing actions that we believe are morally reprehensible. Many of us desire wealth and believe that we can achieve wealth by robbing a bank; many of us desire the company of an attractive partner and believe that slipping her a Mickey Finn is a means for doing so; many of us desire to rid ourselves of an annoying colleague and

---

<sup>19</sup> See Brand, *Intending and Acting*; Bratman, *Intentions, Plans, and Practical Reason*; Harman, "Practical Reasoning" in Mele, ed.; McCann, "Rationality and the Range of Intention", *Midwest Studies in Philosophy* 10:191-211; Mele, *Springs of Action*; Searle, *Intentionality*.

<sup>20</sup> See especially Bratman, *ibid.*, and Harman, *ibid.*

believe that murder is an effective means of doing so. If simple reductive accounts are correct, then any number of us have any number of morally reprehensible intentions. Yet I doubt that most of us actually intend to perform that many morally reprehensible actions, and this is some evidence that simple desire-belief accounts are flawed.

Here are some explanations of why I do not necessarily intend to assassinate Smith and why you do not necessarily intend to rob the bank, among other things. Recall the above claim that an agent who intends to perform some action has effectively answered the question about what she is going to do. Yet an agent who has some relevant desire and belief pair need not have settled the question about what she is going to do. It is possible, for example, that while I have the relevant beliefs about how to satisfy my desire, I have not “put together” by desire and belief pair and I do not recognize that my belief is relevant to the satisfaction of my desire. We know that belief is not closed under entailment, so that an agent may believe the propositions ‘A’ and ‘If A then B’ without also believing the proposition ‘B’ that is entailed by their conjunction. Similarly, an agent might desire to remove Smith from office and believe that assassination is an effective means for removing Smith from office without believing that assassinating Smith will bring about the desired result. An agent who fails to believe that assassinating Smith will bring about the desired result could not be guided by an intention to assassinate Smith, just because she does not recognize that assassinating Smith is a means for realizing the desired result. But that suggests that desire and instrumental belief pairs do not necessarily perform a guiding function. Therefore, desire and instrumental belief pairs are not intentions.

Here is another way an agent can fail to answer the question about what she is going to do, even though she has a desire and instrumental belief pair. An agent might have multiple and conflicting beliefs about how to satisfy her desire, such that she believes there is more than one way to satisfy her desire. Yet she may not be sure which way of satisfying her desire is easiest or most prudent, or she may be uncertain whether some of the options believed to be available to her ought morally not to be performed, or she may withhold final judgment on the matter or seek further counsel, and so forth. In all of these cases, an agent desires some outcome and believes that she can realize that outcome by performing a particular action, but the agent has not yet answered with sufficient finality the question of what she is going to do. An agent's desire and belief pair may influence her behavior, but since she has not yet answered with sufficient finality the question of what she is going to do, her desire and instrumental belief pair will not control her conduct. But intentions control conduct and do not merely influence conduct. Therefore, desire and instrumental belief pairs are not intentions.

Perhaps the simple reductive account of intention is too simple. More complicated reductive accounts of intention have been offered; in particular, some reductive theorists have supposed that an agent must have more than just a belief that performing some action will bring about a desired result to intend to act. Wayne Davis, for example, argues that an agent intends to  $\phi$  just in case the agent believes that she will  $\phi$  because she desires to  $\phi$  and believes that her desire will motivate her to act in such a way that she  $\phi$ s.<sup>21</sup> My suspicion is that any reductive account of intention is bound to fail and that we

---

<sup>21</sup> Davis, "A Causal Account of Intending", *American Philosophical Quarterly* 21, pp. 43-54, reprinted in *The Philosophy of Action*, Mele, ed., p. 147.

will be able to construct counter-examples to that account.<sup>22</sup> At any rate, if some reductive account of intention does escape counter-examples, we can easily accommodate the thesis that agents *can* intend the bad. A Panglossian who defends a reductive account of intention must either argue that desiring the bad is not possible or that the requisite belief requirement for intending precludes intending the bad. I have already argued that desiring the bad is possible in the previous chapter. So the Panglossian must appeal to some belief condition for intending that makes intending the bad impossible: for example, the Panglossian might argue that believing that  $\phi$ -ing is good is necessary for intending to  $\phi$  or perhaps that not believing that  $\phi$ -ing is bad and not at all good is necessary for intending to  $\phi$ . I know of no philosopher who actually defends these putative necessary conditions for intention, and such necessary conditions cannot simply be stipulated without begging the question.

It would be fortunate for the proponent of the *intentionbad* thesis if the simple desire-belief account of intention were correct, since if the simple desire-belief account seems consistent with the thesis that agents can intend the bad. We shall have to see if plausible reductive desire-belief accounts of intention can be produced that escape counter-examples. However, the alleged failures of reductive desire-belief accounts suggest different accounts of intention that suppose that intentions are psychological states irreducible to desire and instrumental belief pairs. In the case of the agent who has not put together her desire and instrumental belief and in the case of the agent who has not yet settled on any course of action, it is tempting to claim that the agent has not *decided* what she is to do. Thus, perhaps it is the case that intentions ought to be

---

<sup>22</sup> For a counter-example to Davis' proposal, see Kirk Ludwig, "Impossible Doings", *Philosophical Studies* 65, p. 258.

understood as states of being decided that end practical inquiry by answering questions about what is going to be done. Similarly, in the case of both the agent who has not put together her desire and belief and the agent who has not yet settled on any course of action, it is tempting to claim that in neither case has the agent *judged* that some action is most desirable or better than its alternatives and so forth. Thus, perhaps it is the case that intentions ought to be understood as evaluative judgments about actions. Eventually, I shall argue that both of these accounts of intention can allow that agents can intend the bad, but it may initially appear that each account rules out intending the bad. I thus consider accounts of intentions as states of being decided to act and as evaluative judgments in the following two sections.

### **Intentions as Evaluative Judgments**

It is plausible to suppose, I suggest, that intending to act is the terminus of practical reasoning and deliberation.<sup>23</sup> It is also plausible to suppose that practical reasoning and deliberation at least sometimes terminates in beliefs about action, for example, in the belief that some action is the best thing to do.<sup>24</sup> Practical reasoning and deliberation may also terminate in a belief that some action would be acceptable or permissible to perform, or is otherwise good enough, and so forth.<sup>25</sup> It is also plausible to suppose that intentions are the terminus of practical reasoning and deliberation since, normally, an agent who intends to do something need not do anything else beside execute her intention if she is to

---

<sup>23</sup> This allegedly plausible suggestion may seem to conflict with Aristotle's familiar suggestion that the conclusion of practical reasoning is an action from *Movement of Animals* 701a1, in *Aristotle: Selections*, Irwin and Fine, eds.. Aristotle's view on this matter is somewhat complicated and I suspect is ultimately not in conflict with the present proposal, but I ignore the complication here.

<sup>24</sup> Bratman, "Practical Reasoning and Weakness of Will", *Nous* 13, pp. 153-71.

<sup>25</sup> Mele, "Deciding to Act", *Philosophical Studies* 100, p. 83.

act intentionally. Thus, we perhaps have some reason to identify intentions with evaluative judgments about actions.

Indeed, at least one prominent philosopher of action has identified intentions with particular kind of evaluative judgment about an action. In his later work on the matter, Donald Davidson claims intentions are a particular kind of pro-attitude: an “all-out” or unconditional judgment, that an action is desirable, or ought to be performed, or is better than alternative actions, or some such thing.<sup>26</sup> If Davidson is correct to identify intentions with evaluative judgments, then weak and strong Panglossian challenges to the *intentionbad* thesis can be offered. Recall that strong Panglossian challenges purport that, for example, the *intentionbad* thesis is conceptually impossible while weak Panglossian challenges purport that the *intentionbad* thesis is always less plausible than standard theses of moral psychology. Explaining the precise nature of the Panglossian challenges will require a discussion of Davidson’s thesis that intentions are evaluative judgments. Briefly, however, the Panglossian challenges are that if intentions are identified evaluative judgments about an action, then it is either unclear how we could intend actions we believe to be bad and not at all good or it is impossible to intend actions we believe to be bad and not at all good. Thus, claims the Panglossian, it is either impossible to intend the bad or it is unclear how agent could intend the bad.

Davidson is led to the identify intentions with evaluative judgments about actions given certain theses he endorses concerning practical reasoning. Here is something of a truism concerning practical reasoning: for (nearly) every action, that action has

---

<sup>26</sup> See Davidson, “Intending”, p. 98-101 and his “Replies to Essays I-IX” in *Essays on Davidson: Actions and Events*, Vermazen and Hintikka, eds., (Oxford: Oxford University Press, 1985), pp. 196-7, 206, 220. Davidson hints at this thesis but does not explicitly identify intentions with all-out evaluative judgments in “How is Weakness of the Will Possible?”.

something to be said for it and something to be said against it. Alternatively, for (nearly) every action, there are reasons to perform the action and reasons to refrain from performing the action. There are perhaps always opportunity costs to performing any action: doing serious work frustrates the desire to lounge about or to see a good movie and so forth. Thus, for (nearly) any action, an agent will judge that the action is desirable to perform, given some reasons for performing the action.<sup>27</sup> Parity of reasoning suggests that for (nearly) any action, an agent will also judge that the action is not desirable to perform, given some reasons against performing the action. But knowing that an action has something to be said for it and something to be said against it is not all there is to practical reasoning: “. . . if an agent is to act in the face of conflict. . . [i]t is not enough to know the reasons on each side: he must know how they add up.”<sup>28</sup> Suppose that an agent does add up all the reasons for and against various actions, given the evidence believed to be relevant, and comes to judge that, all things considered, some particular action is most desirable or ought to be performed or is better than any alternative action. At this stage, an agent has formed a conditional judgment about an action, the judgment that, all-things-considered, the action is desirable or preferable or some such thing: for example, an agent who decides that staying at home and studying has more to be said for it and less to be said against it than going out for a few drinks judges that, all-things considered, staying home and studying is better than going out for a few drinks. The judgment is

---

<sup>27</sup> Davidson suggests that an agent who has a pro-attitude towards performing some action “must have had attitudes and beliefs from which, had he been aware of them and had the time, he *could* have reasoned that his action was desirable (or had some other positive attribute)”; “Intending”, p. 85. The relevant judgment might only be an implicit evaluative judgment, but Davidson suggests there will be a relevant judgment nonetheless that could have been part of an agent’s explicit practical reasoning or deliberation.

<sup>28</sup> Davidson, “How is Weakness of the Will Possible”, p. 36.

conditional insofar the conclusion that the action should be performed is conditioned by the evidence believed to be relevant.

However, Davidson claims that we cannot suppose that practical reasoning only issues conditional evaluative judgments about actions; we must also suppose that practical reasoning issues *unconditional* or *all-out* judgments that an action is desirable or preferable or some such thing, if practical reasoning is ever to produce action. The judgment that *staying home and studying is better than going out for a few drinks* is an unconditional all-out judgment, since the ‘all-things considered’ modifier has dropped out. Unconditional judgments may be based upon conditional judgments about an action, and if an agent is not akratic on Davidson’s view, then an agent’s unconditional and conditional judgments will roughly fall in line. Davidson’s arguments that practical reasoning must issue in unconditional judgments and not simply in conditional judgments are found in a few dark passages.<sup>29</sup> But perhaps we can understand why Davidson supposes that practical reasoning cannot issue only in conditional judgments given we recognize that Davidson identifies intentions with unconditional all-out evaluative judgments about action.<sup>30</sup> Again, we have been supposing that intending is normally the last step in practical reasoning prior to acting intentionally. While we may often come to intend to act on the basis of various reasons and evidence believed to be relevant, the content of our intention usually does not make reference to those reasons and evidence:

---

<sup>29</sup> Davidson claims, for example, that “Reasoning that stops at conditional judgments. . . is practical only in its subject, not in its issue” and that “Practical reasoning does. . . often arrive at unconditional judgments that one action is better than another- otherwise there would be no such thing as acting on a reason”; see Davidson, *ibid.*, p. 39.

<sup>30</sup> See “Intending”, p. 99 and 101 and “Replies to Essays I-IX”, pp. 196-7 and p. 220.

we simply intend to act. At any rate, Davidson identifies intentions with unconditional all-out evaluative judgments about actions.

At this point, nothing yet speaks against the possibility of intending the bad. We can, however, begin considering Panglossian challenges that might follow from Davidson's identification of intentions with all-out evaluative judgments. The Panglossian may offer a strong challenge as follows. Intentions, on Davidson's view, are all-out evaluative judgments that are arrived at on the basis of conditional judgments about the desirability of an action. But if an agent believes that an action is bad and not at all good, then an agent cannot conditionally judge that an action is desirable. Therefore, an agent cannot arrive at the unconditional judgment that an action is desirable, for example. Therefore, on Davidson's view, intending the bad is not possible.

In the previous chapter, I argued that Davidson does not explicitly reject the *desirebad* thesis and some of Davidson's remarks suggest that he does allow that desires for the bad are desires proper. Note that Davidson also claims that: "[t]he reasons an agent has for intending are basically of the same sort as the reasons an agent has for acting intentionally; they consist of both desires (and other pro-attitudes) and beliefs".<sup>31</sup> If desires and pro-attitudes can constitute reasons for intending, then Davidson is committed to claiming desires *for the bad* are reasons to intend. But then it seems that desires for the bad are relevant to practical reasoning and deliberation.

It is important not to confuse a normative reason for acting with an agent's reason for acting, and Davidson's account of intentions as evaluative judgments is a thesis about an agent's reason for acting. Moreover, while Davidson claims that intentions are pro-

---

<sup>31</sup> Davidson, "Replies to Essays I-IX", p. 213.

attitudes, he does not explicitly claim that having a pro-attitude entails believing that the object of the pro-attitude is good.<sup>32</sup> Davidson suggests that having a pro-attitude implies attributing to some state of affairs a positive attribute, but this is not the same thing as thinking that the state of affairs is at all good.<sup>33</sup> Davidson also suggests that having a pro-attitude implies judging that state of affairs has something to be said for it, but this is not the same thing as judging that the state of affairs is at all good.<sup>34</sup> I tentatively suggest that, even given Davidson's account of intention, an agent can intend the bad if she has a pro-attitude towards the believed badness of the action that she intends to perform or has a pro-attitude towards the believed badness of the state of affairs that she believes will obtain if she acts. An agent who intends the bad might intend the bad because she attributes a positive attribute to the action she believes is bad and not at all good or because she judges that the badness of her action is what the action has to be said for it.

It is difficult to shake the thought that there must be *something* irrational about intending the bad; after all, can there be sense made of taking the believed badness of an action to be a positive attribute of the action or to count in favor of performing the action? I examine this difficulty further below. Note, however, that even if it is never rational to intend the bad, irrational intentions *are* still intentions. Akratic intentions are likely irrational as well, but surely one can intend akratically and akratic intentions are intentions. Thus, even if it is not rationally possible to intend the bad, it might be nonetheless possible to intend the bad. All that I require is that intending the bad is

---

<sup>32</sup> Davidson, "Intending", p. 102.

<sup>33</sup> Davidson, *ibid.*, p. 84 and 87.

<sup>34</sup> Davidson, "Actions, Reasons, and Causes", p. 9.

possible, and if intending the bad is possible, then we have some further evidence that intentionally pursuing the bad is possible. The *intentionbad* thesis is a *modal* thesis that intending an action believed to be bad and not at all good is possible, not that it is desirable or prudent to intend the bad.

In this section, I have discussed Davidson's thesis that intentions just are a particular kind of evaluative judgment concerning actions. I suggest that Davidson's thesis is consistent with the *intentionbad* thesis and allows that agents can intend the bad. The Panglossian may object that it is a necessary condition of intending that the acting agent thinks that there is something good about her action. But the judgment that something is good is only one kind of evaluative judgment. As I suggest above and pursue below, the evaluative judgment that an action is bad and not at all good can also be part of what generates or constitutes an intention.

### **Intentions and Decisions to Act**

As discussed above, intending to act is at least one way for an agent to settle questions about what she is going to do. Another way for an agent to settle questions about what she is going to do is to decide to perform some action. Insofar as intending to act and deciding to act both normally settle questions about what an agent is going to do, we can consider the following plausible thesis: to decide to  $\phi$  just is to intend to  $\phi$ . Many philosophers have argued that to decide to perform some action  $\phi$  is at least a way of forming an intention to  $\phi$ .<sup>35</sup> Thus, we might suppose that states of being decided about acting just are intentions.

---

<sup>35</sup> See Kaufman, "Practical Decision", *Mind* 75, pp. 25-44; McCann, "Intrinsic Intentionality", *Theory and Decision* 20, pp. 247-73; Frankfurt, "Identification and Wholeheartedness", from his *The Importance of What We Care About*, pp. 174-6; Mele, *Springs of Action*, p. 156; Kane, *The Significance of Free Will*, p. 24; Pink, *The Psychology of Freedom*, p. 3; Al Mele, "Deciding to Act", p. 81-108.

It is possible for an agent to believe that some action is the thing to do but not yet intend to perform that action.<sup>36</sup> An agent who is akratic or who is suspicious of her faculty of practical reasoning or who is very timid might believe that some action is the thing to do, but not yet have committed herself to performing that action. An akratic smoker might believe that he ought to quit immediately but not yet have resolved that he will take any steps to quit.<sup>37</sup> A patron at a restaurant may believe that the mother at another table is acting badly when she raises her hand to her child and that intervening is the thing to do, but if he lacks the courage of his convictions, he might fail to intend to intervene. In the case of the akratic smoker and the timid patron, the agents believe that some action is the thing to do, but, I suggest, neither yet intends to perform that action. Neither agent has yet decided *to act*, even if they have decided that some action is the thing to do. As noted above, to intend to  $\phi$  is to be committed in the relevant sense to  $\phi$ -ing if the opportunity comes up, other things being equal. That I have decided to  $\phi$  suggests a similar sort of commitment to  $\phi$ -ing if the opportunity comes up, other things being equal. Deciding to act seems to plausibly fill the gap between believing that some action is the thing to do and actually performing that action and it can therefore seem plausible to identify intending to  $\phi$  with being decided to  $\phi$ .

Providing the details of the proposal that intentions are to be identified with states of being decided to act is beyond the scope of this project. But suppose that it is *prima facie* plausible to identify intentions with states of being decided to act. Nothing that has

---

<sup>36</sup> The discussion that follows may actually suggest that identifying intentions with evaluative judgments is mistaken, since evaluative judgments, in the relevant sense, seem to be beliefs that an action ought to be performed.

<sup>37</sup> Al Mele, *ibid.*, p. 83.

been said so far precludes the possibility of deciding to perform an action one believes is bad and not at all good and thus, for all that has been said, an account of intention that identifies intentions with states of being decided to act is consistent with the *intentionbad* thesis. The Panglossian might nonetheless challenge to the thesis that agents can decide, and thereby intend, to perform an action believed to be bad and not at all good by insisting that there is some dependency relation between deciding to  $\phi$  and believing that  $\phi$ -ing would be good.

I shall have more to say about decisions and intentional action in my fifth chapter, but briefly, my suggestion is that in the same way that an agent's moods and emotions can cause her to desire the bad, an agent's moods and emotions can cause her to decide act in ways that she believes are bad and not at all good. An agent who is in the grips of frustration might decide to punch a hole in her wall just because punching the wall is bad and not at all good. A sufficiently depressed agent might decide to mutilate herself just because the damage she inflicts is bad and not at all good. On the account of intention being considered, agents who decide to perform actions that they believe are bad and not at all good thereby intend the bad.

Suppose, however, that we grant the Panglossian that agents cannot decide to perform actions they believe are bad and not at all good. It does not follow that agents cannot intend the bad. Even if an agent initially believed there is something good about  $\phi$ -ing and decides to  $\phi$ , the agent's intention to  $\phi$  might persist even after the agent's evaluative belief there is something good about  $\phi$ -ing is gone.<sup>38</sup> Suppose that a young

---

<sup>38</sup> A suggestion by Kirk Ludwig, albeit one made in a very different context, brought this response to light. Ludwig suggested that even if intentions depend upon an agent's desire to  $\phi$  and a belief that some action will bring about  $\phi$ -ing, the intention to  $\phi$  might persist even if the relevant desire or belief is lost.

woman, Michelle, acquires an eating disorder that leads her to decide to starve herself to dangerously unhealthy degrees. Perhaps when she initially began starving herself, she believed that starving herself is good since starving herself is a means for acquiring a slim figure. On that basis, she decided to starve herself. But over time, perhaps Michelle begins to doubt that starving herself is good, since her health has began to deteriorate and she thinks that the figure that she sees in the mirror is not at all attractive. Michelle now no longer believes that starving herself is good nor does it even seem to her that starving herself is good. But her intention to starve herself is not easily abandoned, even though she no longer believes that starving herself is good. Our previous observations intention explain how agents could continue to intend to perform actions they previously decided to perform because those actions were thought to be good. We noted above that intentions are resistant to reconsideration and have a characteristic inertia and stability. If intentions are resistant to reconsideration, then if an agent intends an action she believes is bad, her intention to perform that action could overwhelm the influence of her evaluative belief, just because intentions *are* by nature resistant to reconsideration. In this case, Michelle decided to do something that she now believes is bad and now still intends the bad, even though she no longer thinks that her action is good. Thus, even if intentions are states of being decided to act and even if the Panglossian is correct that agents cannot decide to perform actions they believe are bad and not at all good, agents can still intend the bad.

Of course, intentions are not irrevocable and there are circumstances in which an agent can and should abandon an intention. Perhaps the Panglossian means to object that an agent cannot *continue* to intend the bad if she fails to believe there is something good

about her action. The Panglossian could claim that to continue to have an intention, the intention must survive further rational consideration. And if the agent's intention does not survive rational consideration, then it is simply misleading to say that the agent intends to act in ways that she believes are bad and not at all good. There are conditions under which an agent may reconsider and perhaps revise or abandon a standing intention to act. If an agent acquires new beliefs or if information available to her changes, an agent might revise or abandon an intention she previously held and be entirely reasonable or rational in doing so. Michelle initially thought that there was something good about starving herself but no longer thinks that there is anything at all good about starving herself. Perhaps Michelle will reconsider and abandon her earlier decision to starve herself.

Then again, *she might not*. To claim that there are conditions in which intentions are revocable is to claim that there are conditions in which it is rationally permissible to abandon an intention; it hardly follows that in every such case, the agent will abandon her intention, even if she may rationally do so. The same standing dispositions or moods or emotions that led Michelle to form the intention to starve herself might dispose her to continue to intend to starve herself even though she no longer thinks that starving herself is a good thing. The story of Michelle seems coherent. Thus, even if agents who presently intend the bad could rationally revoke their initial decision to act, it does not follow that they will. Therefore, it does seem possible for agents to continue to intend the bad, even if they believe that relevant facts have changed.

Could an agent's intention to perform some action she believes to be bad survive *rational* reconsideration? Even if we reject the implausible thesis that an agent's

intentions must always fall in line with her evaluative beliefs, we might still insist that an agent is irrational if she does not bring her intentions in line with her evaluative beliefs. One response is to insist that rationality is besides the point here: irrational intentions are still intentions and even if an agent could only irrationally intend the bad, she would still intend the bad.

In this section, I argued that the thesis that intentions are states of being decided to act is consistent with the *intentionbad* thesis. But recall that one of the reasons for rejecting reductive desire-belief accounts of intention was that intentions seem to be governed by norms of rationality that desire and instrumental belief pairs are not necessarily governed by. But if intentions are governed by certain norms of rationality, or rather, if part of what it is to be an intention is that intentions are governed by certain norms of rationality, then it may be a mistake to simply wave off the objection that intending the bad is irrational. In the penultimate section of this chapter, I face up to the issue whether agents can rationally intend the bad and I attempt to specify some limited circumstances in which intending the bad might be rational.

### **Whither Rationality?**

Can it be rational to intend the bad? Can it be rational to intend to perform an action that is believed to be bad and not at all good? My answers to these questions are at best tentative and speculative, but I suggest the answers to both questions are ‘yes’. To be clear, the success of the present project does not depend upon successfully demonstrating that it can be rational to intend the bad. My thesis is that intending the bad is possible, not that it is rational, but if it can be shown that intending the bad is sometimes rational, it may seem more plausible to suppose that intending the bad is

possible. And I submit, there are perhaps some instances in which intending the bad is rational.

First, consider Thomas Fowler, the protagonist in Graham Greene's *The Quiet American*, who publicly insults the man whom he believes stole his Vietnamese mistress; Fowler defends himself by explaining that he will act badly just because this is one of those situations where people do act badly.<sup>39</sup> In just what sense Fowler takes himself to have acted badly is unclear, but suppose that Fowler himself believes that publicly insulting his rival is bad and not at all good. Must it be irrational for Fowler to nonetheless intend to publicly vent his rage? Fowler believes that this is a situation where people do act badly, so perhaps on that basis he comes to intend to act badly. Similarly, I might be unfamiliar with exactly what is going on in a football game but I might intend to cheer when the home team comes on the field because I believe that cheering when the home team comes on the field is what people do in situations like this. It is unclear that I am irrational in intending to cheer and similarly it might be unclear that Fowler is irrational in intending to act badly. Alternatively, Fowler's claim that this is a situation in which people *do* act badly is perhaps a disguised way of claiming that this is a situation in which he *ought* to act badly. It would be odd to claim that it is irrational for an agent to intend to do what she believes she ought to do. Note also that intending to  $\phi$  because you believe that you ought to  $\phi$  does not entail believing that  $\phi$ -ing is good. An opponent of morality, perhaps an agent who has read too much Nietzsche, might recognize that he morally ought to act in certain ways but he does not think that acting morally is necessarily good. An opponent of convention, like Dostoyevsky's man from

---

<sup>39</sup> Graham Greene, *The Quiet American*, p. 147.

the underground, might think that etiquette demands that he ought to act in certain ways but he does not think that following the demands of convention is good. So perhaps there is room to suppose that intending the bad is at least sometimes rational.

Second, an analogy might be drawn between intending the bad and intending akratically. Akratic action is a paradigm example of irrational behavior and intending akratically is a paradigm example of irrational intending.<sup>40</sup> Recently, however, some philosophers have questioned whether or not akratic action is always irrational.<sup>41</sup> If akratic action is sometimes rational, then perhaps it is sometimes rational to intend akratically. And if it is sometimes rational to intend akratically, perhaps it is sometimes rational to intend the bad.

Finally, note that there is a gap between the rationality of intending to  $\phi$  and the rationality of  $\phi$ -ing intentionally; even if the latter would be positively irrational, the former need not be. For example, even though it would be positively lunatic to fire nuclear missiles at an enemy who has already fired theirs since doing so would only bring on more destruction, it may not be irrational to *intend* to fire at an enemy if he fires his first.<sup>42</sup> Suppose your enemy is extremely intuitive and can discern whether or not you

---

<sup>40</sup> Jon Elster, *Ulysses and the Sirens: Studies in Rationality and Irrationality*; Mele, *Irrationality: A Study on Akrasia, Self-Deception, and Self-Control*.

<sup>41</sup> See, for example, Nomy Arpaly, "On Acting Rationally Against One's Best Judgment", *Ethics* 110, pp. 488-513; Alison McIntyre, "Is Akratic Action Always Irrational?", from *Identity, Character, and Morality*, Flanagan and Rorty, eds., pp. 379-400; Robert Audi, "Weakness of Will and Rational Action", *Australasian Journal of Philosophy* 68, pp. 410-30. Thomas Hill expresses at least some doubt that akratic action is always irrational in his "Weakness of Will and Character", *Philosophical Topics* 14, pp. 93-115. Arpaly suggests, and I concur, that Harry Frankfurt is committed to rejecting the thesis that akratic action is always irrational; see Frankfurt's "Rationality and the Unthinkable", *The Importance of What We Care About*, pp. 177-90.

<sup>42</sup> David Lewis, among others, has a collection of papers on this subject reprinted in his *Papers in Ethics and Social Philosophy*; it is of course not uncontroversial that utilizing mutually assured destruction is either rational or the most rational option available here, but I gloss this point.

really will respond in kind to a nuclear attack. If he intuits that you will not, then there is little to deter him from attacking you first. Further, if he believes that you would not intend the bad, he may doubt your resolve and attack first, but if he believes that you intend to fire back and that you believe firing back is bad and not at all good, he may believe that you are a lunatic who had best not be prodded. One way to convince him that he best not attack first is to sincerely intend to respond in kind to a nuclear attack, but not because you believe deterring nuclear war is good. If you do intend what you believe is bad and not at all good, you will have successfully deterred him from attacking you first. If this strategy could be successful, then it may be rational to intend what you yourself believe is bad and not at all good, even if it would be irrational to act upon the intention.

Here is another instance in which it might be rational to intend the bad. Consider the following ironic saying: to get to heaven, you have to raise a little hell. Suppose that God has a special place in heaven reserved for genuinely repentant sinners who have genuinely committed moral transgressions. Perhaps you believe that the best chance to get into heaven is to intend to commit a series of venal sins, to execute those intentions, and later to genuinely repent for your sins. Of course, if you believe that there is something good about sinning then you won't genuinely intend the bad, you cannot genuinely repent, and He will not be persuaded to welcome you to heaven. Suppose at time  $t$ , to ensure that you will genuinely commit moral transgressions, you take steps to ensure that you will intend the bad at time  $t + 1$ . Undergoing brain surgery or securing counter-factual interveners could ensure all of this. Further, you take measures to ensure that, at time  $t + 1$ , you will have forgotten about your actions at time  $t$ , such that you

intend to act in ways that you believe are bad and not at all good. Thus, at time  $t + 1$ , you would genuinely intend the bad. Finally, suppose that you ensure that at some later time  $t + 2$ , you will come to repent in order to be forgiven for your moral transgressions. Perhaps you ensure that the persuasive priest from your childhood gets you to repent. This plan would ensure that you genuinely intend the bad and that you genuinely repent and thus that you get into heaven by raising a little hell.<sup>43</sup> If getting into heaven is a rational enterprise and if you believe this plan would be successful, then it would be rational to intend the bad.<sup>44</sup> Here too, I do not suppose that a few examples will settle the question whether intending the bad is rational, but these examples are a way to begin a long debate. At the very least, my examples suggest we cannot dismiss outright the thesis that intending the bad might sometimes be rational.

Again, the success of my overall project does not depend on showing that it is rational either to intend the bad or to intentionally pursue the bad, although I believe the above arguments suggest that at least in some local instances it might be rational to do so. Again, it is no part of my project to commend or recommend intending the bad. This is not to say that intending the bad might not be just the thing to do, however; indeed, there might be some circumstances in which rationality demands no less.

### **Conclusion**

My goal in this chapter has been to begin to defend the *intentionbad* thesis that agents can intentionally perform actions that they believe to be bad and not at all good. I discussed various functional roles that intention plays in hope of explaining the nature of

---

<sup>43</sup> This plan may suppose that He is something of a chump and won't see through your clever plan, but surely He is a man of His word and will honor His offer and let you into heaven.

<sup>44</sup> The preceding argument is inspired by Kavka's "The Toxin Puzzle" *Analysis* 43 (1983) pp. 33-6.

intention. I have also examined accounts that suppose intentions are suitably paired desires and beliefs and I have examined accounts that suppose that intentions are evaluative judgments or states of being decided to act. In any case, I have argued that however we understand intentions, it is possible for agents to intend the bad. I also suggested that in some cases, perhaps very contrived cases, it is possible to rationally intend the bad. Thus, I conclude that intending the bad is possible.

Still, it does not follow that intentionally pursuing the bad is possible. Therefore, even if it has been shown that agents can intend the bad, the *intentionbad* thesis has not been secured. Quite generally, it is not the case that if an agent can intend to  $\phi$ , an agent can  $\phi$  intentionally. We might be able to intend to perform actions that are nomic or physical impossibilities. A group of agents might be oddly struck by lightning every time one of them intends to perform a token of some act-type. Instances of causal deviance might be pervasive such that no one ever intentionally  $\phi$ s, even if everyone intends to  $\phi$ . Or agents might be very lazy. Since it is not true that intending to  $\phi$  is sufficient for  $\phi$ -ing intentionally, then it is not true that intending the bad is sufficient for intentionally pursuing the bad. But then, again, the *intentionbad* thesis has not been secured.

However, here is an argument that suggests that the *intentionbad* thesis is very likely the case, given that it is possible for agents to intend the bad:

- 1) If an agent can intend to  $\phi$  and if  $\phi$ -ing is a nomic and physical possibility, then, other things being equal, an agent can  $\phi$  intentionally
- 2) An agent can intend to pursue the bad
- 3) Intentionally pursuing the bad is a nomic and physical possibility
- 4) Therefore, other things being equal, an agent can intentionally pursue the bad

The 'other things being equal' clause is not meant to be controversial; I only mean to preclude familiar goblins that annoy action theorists and prevent us from offering a full-

blown analysis of intentional action. Suppose I intend to perform some action  $\phi$ . Suppose also that I do not abandon my intention, that contingent circumstances do not prevent me from executing my intention, that the time and opportunity to execute the intention come up, that I do not run afoul of causally deviant chains, and so forth. In those circumstances, it seems plausible to suppose that I will execute my intention and that I will  $\phi$  intentionally. But if my intention is to perform an action that I believe is bad and not at all good, then I intend the bad. And if I successfully execute my intention to perform an action I believe is bad and not at all good, then I will intentionally pursue the bad.

There are three responses to the above quick argument that the Panglossian may offer. First, the Panglossian might argue that the intentional pursuit of the bad is not a nomic or physical possibility. This first response does not seem likely to succeed. Some actions are not nomic or physical possibilities: I cannot intentionally square the circle and I cannot intentionally leap unaided over Griffin-Floyd Hall. But instances of the intentional pursuit of the bad are not guaranteed to be actions that are similarly impossible. Second, the Panglossian might object that the first premise of the above quick argument is false. Instead, the Panglossian might propose the following revised argument:

- 1) If an agent can intend to  $\phi$  and if  $\phi$ -ing is a nomic and physical and conceptual possibility, then, other things being equal, an agent can  $\phi$  intentionally
- 2) An agent can intend to pursue the bad
- 3) Intentionally pursuing the bad is a nomic and physical and conceptual possibility
- 4) Therefore, other things being equal, an agent can intentionally pursue the bad

The Panglossian would likely then deny that the above argument is sound, on the grounds that that the intentional pursuit of the bad is not conceptually possible. In the following

chapter, I shall consider a strong Panglossian challenge of this kind that both the *intentionbad* and *reasonbad* theses are false. The challenge depends on the claims that it is a conceptual truth that intentional actions are actions performed for reasons and that it is not conceptually possible that an instance of the intentional pursuit of the bad is performed for a reason. I turn to this argument in the following chapter..

CHAPTER 4  
ACTING FOR REASONS AND REASONS FOR  
INTENTIONALLY PURSUING THE BAD

You see: reason, gentleman, is a fine thing, that is unquestionable, but reason is only reason and satisfies only man's reasoning capacity, while wanting is a manifestation of the whole of life—that is, the whole of human life, including reason and various little itches. And though our life in this manifestation often turns out to be a bit of trash, it is still life and not just the extraction of a square root. . . . human nature acts as an entire whole, with everything that is in it, consciously and unconsciously, and though it lies, still it lives.

Fyodor Dostoevsky, *Notes From the Underground*<sup>1</sup>

**Introduction**

In previous chapters, I argue that the *desirebad* thesis is true and that it is possible to desire what is believed to be bad and not at all good. I also argue that agents can intend the bad and this is strong reason to suppose that the *intentionbad* thesis is true. However, I acknowledge that even if the *desirebad* thesis is true and even if agents can intend the bad, the *intentionbad* and *reasonbad* theses do not follow; it is possible that agents can desire the bad but be unable to translate those desires into action. In this chapter, I examine a strong Panglossian challenge that implies that even if the *desirebad* thesis is true, the *intentionbad* and *reasonbad* theses are nonetheless false. The strong Panglossian challenge begins from plausible theses concerning intentional action and leads to the conclusion that it is conceptually impossible to intentionally pursue the bad because it is conceptually impossible to intentionally pursue what is thought to be bad and not at all good where in doing so, one acts for a reason. I will attempt to defang this

---

<sup>1</sup> Fyodor Dostoevsky, *Notes From the Underground*, p. 28, emphasis added.

challenge and argue that it is conceptually possible to intentionally pursue the bad because it is conceptually possible to intentionally pursue the bad for a reason. I thus defend both the *intentionbad* and *reasonbad* theses.

Here are two intuitive theses: intentional actions are performed for reasons and acting intentionally just is acting for a reason. While I do not argue for their truth, note that many philosophers appear to accept them, or at least accept close relatives of them. Donald Davidson famously argues that reasons are the cause of intentional actions<sup>2</sup> and explicitly states that “it is (logically) impossible to perform an intentional action without some appropriate reason.”<sup>3</sup> Alvin Goldman claims that a piece of behavior is “an intentional action if and only if it is done for a reason.”<sup>4</sup> Robert Audi claims that “an action for a reason apparently must be intentional.”<sup>5</sup> Al Mele and Paul Moser take it to be a virtue of their account of intentional action that their account is consistent with the “popular” thesis that “intentional A-ing is coextensive with A-ing done for a reason.”<sup>6</sup> Carl Ginet suggests that there are “important connections between the concept of an action and the concept of an explanation of an action in terms of the acting agent’s reasons for doing it” and that “actions, and only actions, can have that sort of explanation.”<sup>7</sup> Rüdiger Bittner states that “people do things for reasons.”<sup>8</sup>

---

<sup>2</sup> Davidson, “Actions, Reasons, and Causes”, pp. 3-19.

<sup>3</sup> Davidson, “Hempel on Explaining Action”, p. 264.

<sup>4</sup> Goldman, *A Theory of Human Action*, p. 70.

<sup>5</sup> Audi, “Acting for Reasons”, *The Philosophical Review* 45:4 (Oct. 1986), p. 511.

<sup>6</sup> Mele and Moser, “Intentional Action” from *The Philosophy of Action*, Mele, ed., p. 254.

<sup>7</sup> Ginet, *On Action*, p. 3.

<sup>8</sup> Bittner, *Doing Things for Reasons*, p. ix.

The intuitive theses perhaps contain the seeds of the strong Panglossian challenge alluded to above that the *intentionbad* and *reasonbad* theses are conceptually impossible. The strong Panglossian challenge can be reconstructed briefly as follows. An agent who intentionally pursues the bad believes that the object of her pursuit is bad and not at all good. An agent who believes that what she pursues is bad and not at all good could not believe that there is a reason for her to act. Thus, an agent who intentionally pursues the bad could not act for a reason. But intentional actions *are* performed for reasons and acting intentionally just *is* acting for a reason. Therefore, no apparent instance of the intentional pursuit of the bad is actually an intentional action and it is simply misleading to speak of agents “intentionally” pursuing the bad. The proponent of the *intentionbad* and *reasonbad* theses thus appears to be committed to rejecting the intuitive theses.

In what follows, I consider various developed formulations of the above strong Panglossian challenge. I argue that none of them is decisive and that the Panglossian has not shown that the *intentionbad* and *reasonbad* theses are false. The Panglossian objections depend upon ambiguities with respect to our use of the term ‘reason’ and upon an overly restrictive account of what it is to believe that an action is justified.

### **The Problematic Syllogism**

Here again is the brief construction of the strong Panglossian challenge:

An agent who intentionally pursues the bad believes that the object of her pursuit is bad and not at all good. An agent who believes that what she pursues is bad and not at all good could not believe that there is a reason for her to act. Thus, an agent who intentionally pursues the bad could not act for a reason. But intentional actions *are* performed for reasons and acting intentionally just *is* acting for a reason. Therefore, no apparent instance of the intentional pursuit of the bad is actually an intentional action.

The challenge hinges on the premises that an agent must have a particular kind of belief to act for a reason and that an agent who intentionally pursues the bad lacks such a belief.

The following is a more detailed construction of the challenge. I shall refer to the following argument as *the problematic syllogism*; it is problematic precisely because it implies that the *intentionbad* and *reasonbad* theses are false. The problematic syllogism can be rendered as follows:

- 1) If an agent is acting intentionally, then she is acting for a reason;
- 2) Acting for a reason is acting with the belief that there is a reason for performing the action;
- 3) Acting with the belief that there is a reason for performing an action is acting with the belief that there is something about the action that justifies performing it;
- 4) Acting with the belief that there is something about an action that justifies performing it is acting with the belief that there is something good about the action;

The conjunction of 1) - 4) entails the following conclusion:

- 5) If an agent is acting intentionally, then she believes that there is something good about the action she performs.

Further, the conjunction of 2), 3) and 4) entails the following conclusion:

- 6) Acting for a reason is acting with the belief that there is something good about the action being performed.

5) entails that the *intentionbad* thesis is false, since one of the necessary conditions for intentionally pursuing the bad is, at least, that an agent fails to believe that her action is at all good. 6) entails that the *reasonbad* thesis is false, since it entails that an agent cannot act for a reason and fail to believe that there is something good about the action being performed.

I shall not challenge the truth of 1), but I shall examine variations of 1) below.

Premises 2), 3), and 4) are all intuitively plausible and all have philosophical arguments supporting them. Consider 2): acting for a reason is acting with the belief that there is a

reason for performing the action. An impressive number of philosophers have noted that agents regularly provide spontaneous and non-inferentially derived explanations of their actions, even while performing the action.<sup>9</sup> Call this the *spontaneity principle*. Suppose that the explanations are explanations in terms of reasons. It would be difficult to understand how agents could regularly offer spontaneous and non-inferentially derived explanations of their actions in terms of the reasons for which they acted unless they had beliefs about the reasons for which they acted while they were acting. Further, intentional action is often the product of reflective deliberation about what to do. It is plausible to suppose that agents who engage in reflective deliberation are considering and balancing reasons for and against performing various actions. Call this the *deliberative principle*. If agents do engage in reflective deliberation, then it is plausible to suppose that they must have beliefs about reasons having to do with their actions. Finally, it is plausible to suppose that intentional action is purposeful behavior, such that, whenever an agent acts intentionally, the agent has a purpose for performing that action. Call this the *teleological principle*. But if an agent has a purpose for performing an action, then it seems that she must believe that she has the purpose and thus believe that she has a reason to act, since believing an action will fulfill a purpose implies believing there is a reason for performing the action.

Consider also 3): acting with the belief that there is a reason for performing an action is acting with the belief that there is something about performing the action that justifies the action. An analogy between practical and theoretical reasoning might support 3). It is plausible to suppose, as David Velleman has, that theoretical reasoning

---

<sup>9</sup> For an impressive list of citations, see David Velleman, *Practical Reflection*, pp. 18-9.

constitutively aims at tracking the truth: Velleman claims that an agent who has no inclination toward the truth isn't in a position to believe.<sup>10</sup> Having an inclination toward the truth is a matter of being receptive and responsive to reasons for believing a proposition: roughly, an agent who actually engages in theoretical reasoning comes to believe a proposition only if she is responsive to reasons for believing that proposition. For example, if an agent believes that there is no justification for believing some proposition *P*, then given that theoretical reasoning constitutively aims at the truth, the agent would presumably not believe *P*. If there is an analogy to be drawn between theoretical and practical reasoning, then it seems plausible to suppose that agents who engage in practical reasoning must believe that their actions are justified if they are to act, just as agents who engage in theoretical reasoning must believe that a proposition *P* is justified if they are to believe that *P*. So perhaps practical reasoning constitutively aims at something analogous to truth. Call this *Velleman's principle*.<sup>11</sup> Further, recall that in previous chapters I have accepted Davidson's thesis that the explanation of intentional action is rationalizing explanation and the thesis that rationalizations explain by citing what it is about an action that appeals to an agent and leads her to perform it.<sup>12</sup> By Davidson's own lights, a rationalization provides at least a minimal justification of the performance of an action. Davidson claims there is a sense, albeit an "anemic one", in

---

<sup>10</sup> Velleman, "The Possibility of Practical Reason", reprinted in his *The Possibility of Practical Reason*, p. 188. See also "Introduction", pp. 17, from the same volume.

<sup>11</sup> Velleman also argues that there is a constitutive motive of practical reasoning that in turn provides the constitutive aim of practical reasoning. See his "Introduction", p. 17-8, fn. 26. Unfortunately, Velleman's views about what the constitutive motive and aim of practical reasoning are have evolved, and thus it is unclear what he thinks the constitutive motive and aim of practical reasoning are. Velleman offers the intentionally imprecise description of the constitutive motive as "a desire to be actuated by reasons"; *ibid.*, p. 19.

<sup>12</sup> Davidson, "Actions, Reasons, and Causes", pp. 3-19.

which every rationalization justifies: “from the agent’s point of view there was, when he acted, something to be said for the action.”<sup>13</sup> Call the thesis that rationalizing explanations express an agent’s belief that her action has something to be said for it and is therefore at least minimally justified *Davidson’s principle*.

The same principles that appear to justify 3) also appear to justify 4). Perhaps just as theoretical reasoning constitutively aims at truth, practical reasoning constitutively aims at the good. Perhaps agents who find their obsessive or compulsive behavior to be excessive and unreasonable also believe that there is nothing good about their actions. Perhaps it is simply obscure how one could believe an action is justifiably performed if one did not believe that there is something good about the action.

Given that the problematic syllogism appears to be a straightforwardly valid argument and given that 1) - 4) appear to be well justified, there are grounds for accepting 5) and 6) and for rejecting the *intentionbad* and *reasonbad* theses. I respond to the challenges to the *intentionbad* thesis and to the *reasonbad* thesis separately. I argue that while the problematic syllogism contains a formally valid argument in opposition to the *intentionbad* thesis, the argument suffers from an informal fallacy. I then argue that the argument contained within the problematic syllogism that is opposed to the *reasonbad* thesis can be resisted.

### **A Taxonomy of Reasons**

As I argue in this section, I argue that there is more than one kind of reason relevant to the explanation and justification of action and that the term ‘reason’ is often used equivocally. Nonetheless, insofar as we hope to find something consistent about the

---

<sup>13</sup> Davidson, *ibid.*, p. 9.

different senses of ‘reason’, we should, I suggest, assume the following two theses. First, our conception of a reason will be that of a consideration that is relevant to the explanation and justification of action; to be a reason just *is* to be the sort of thing that is relevant to the explanation and production and justification and evaluation of actions. This observation has led some philosophers to suppose that reasons must be *dicta*, where *dicta* are, roughly, predicates or propositions or sentences. For example, Steven Darwall claims that since reasons “must be the sort of thing that can be . . . said on behalf of an act” and “thought in favor” of an act, reasons must be *dicta*.<sup>14</sup> However, Darwall’s claim cannot be correct. Predicates and propositions and sentences are not the sort of things that explain or justify action; therefore, *dicta* are not reasons and *vice versa*. Instead of claiming with Darwall that reasons are *dicta*, we ought to assert merely that reasons must be *expressible* as *dicta*. In citing a reason to explain or justify an action, we claim *that* we have reasons; when reasoning about what we should do, we suppose *that* there are relevant reasons; if we finally decide to perform some action, we judge *that* we have sufficient reason to act. Thus, it is essential that reasons figure into explanations and justifications of action, but all that depends upon the second thesis: reasons must be capable of being represented or expressed in *dicta*.

I have already suggested that reasons can be asserted as part of an explanation or justification of behavior. There is an intuitive difference between explanation and justification. At least, it is surely the case that not every explanation of some event, *E*, justifies the occurrence of *E*; shifting weather patterns might explain a drought in the Midwest, but the drought is not justified because of shifting weather patterns. Perhaps,

---

<sup>14</sup> Darwall, *Impartial Reason*, p. 31.

therefore, we must distinguish between different kinds of reasons that are invoked in explanations of behavior, on the one hand, and in justifications on the other.<sup>15</sup>

There are, however, any number of things that might be included in an explanation of an agent's behavior, but not all of those things are reasons in the appropriate sense. A lack of serotonin might be part of what explains why an agent mopes and damage to the prefrontal lobes of the brain might be part of what explains violent behavior, but a lack of serotonin and damage to the prefrontal lobes are not reasons in the relevant sense. We are interested in a sub-class of reasons invoked in explanations of why an agent acts that we can refer to as *motivating* reasons. Following Michael Smith, we can characterize a motivating reason as a psychological state that is potentially explanatory of an agent's behavior.<sup>16</sup> Further, we can simply stipulate that motivating reasons are not simply brute causes, but must somehow be capable of rationalizing behavior. In previous chapters, we have considered various plausible candidates for motivating reasons; desires, beliefs, and intentions are all psychological states that potentially explain an agent's behavior but also are capable of rationalizing an agent's behavior and thus count as motivating reasons. Of course, not all motivating reasons will actually explain an agent's behavior since not all motivating reasons will actually produce any behavior. I might desire to read James Joyce's *Ulysses* without ever cracking its spine, much less picking up the book or purchasing it or forming the intention to find a copy or so forth. I might be physically or psychological unable to perform any relevant mental or physical behavior, I may be

---

<sup>15</sup> See, for example, Derek Parfit, *Reasons and Persons*, p. 118: "We must distinguish between two kinds of reasons; explanatory and good." See also Kurt Baier, *The Moral Point Of View*, pp. 148-56; Bernard Williams, "Internal and External Reasons" from his *Moral Luck*; Michael Smith, "The Humean Theory of Motivation", p. 38; Steven Darwall, *ibid.*, p. 27, among others.

<sup>16</sup> Smith, *ibid.*, p. 38.

struck dead before the opportunity to act comes up, and so forth. *Effective* reasons are a sub-class of motivating reasons that actually do explain an agent's behavior.

It is not necessary for something to be a motivating reason that an agent believes that she is motivated as a result of that motivating reason. An agent suffering from an Oedipus complex perhaps has a desire to sleep with his mother but insofar as he surely represses his desire, he does not believe that he is motivated to sleep with his mother. Still, his Oedipal desire might explain why he pines for mother when she is absent, why he refuses to leave mother's house, and so forth; if his Oedipal desire does explain his behavior, then his Oedipal desire is a motivating reason. This story seems coherent. If the story is coherent, then we must admit that it is possible for an agent to have a motivating reason but not believe that she has the motivating reason.

Not everything that we refer to as a 'reason' must actually explain behavior. In particular, reasons that *justify* behavior need not explain anything at all. Reasons that justify behavior are *normative* reasons. Various philosophers have suggested plausible accounts of normative reasons. Smith characterizes a normative reason as some consideration that justifies an agent's behavior "from the perspective of some normative system that generates that requirement."<sup>17</sup> Thomas Scanlon claims that a reason for something is a consideration that counts in favor of the thing. For example, a reason for performing an action just is a consideration that counts in favor of performing the action.<sup>18</sup> One thing that unites these accounts is that it is possible on all these accounts that a normative reason never explains anything that any agent does. Presumably, British

---

<sup>17</sup> Smith, *ibid.* p. 39.

<sup>18</sup> Scanlon, *What We Owe To Each Other*, p. 17.

sailors had a normative reason to eat citrus fruits to prevent scurvy and cavemen had a normative reason to brush their teeth, even if neither the sailors nor the cavemen could have intentionally eaten citrus fruits or brushed their teeth for that reason.

Note that nothing in the present characterizations of motivating and normative reasons prevents something from being both a motivating and normative reason. I have suggested that there are different kinds of reasons—motivating and normative reasons that justify and explain behavior respectively—but that does not mean a reason could not both justify and explain. Note that normative reasons are sometimes offered in explanations of intentional action. In some cases, an answer to the question “Why did you do that?” elicits both a rationalizing explanation and a justification. The most common, if not uncontroversial, example of something that is both a motivating reason and a normative reason is a morally innocent desire; for example, my desire to drink a cup of tea might both explain my act of brewing a cup and justify my act.<sup>19</sup> Admittedly, it is controversial whether or not desires can constitute normative reasons.<sup>20</sup> I do not insist that desires can or ever do constitute normative reasons. I only pause to note that, as far as I can tell, nothing that has been said so far indicates that something could not be both a normative and motivating reason.

---

<sup>19</sup> For arguments that at least some desires provide normative reasons for action, see Stephen Schiffer, “A Paradox of Desire”, *American Philosophical Quarterly* 13:3 (July 1976), pp.197-8; Donald Davidson, Davidson, “Intending”, p. 102 and “Replies to Essays I-IX”, from *Essays on Actions and Events*, Vermazen and Hintikka, eds., p. 213; Dennis Stampe, “The Authority of Desire”, *The Philosophical Review* 46:3 (July 1987), pp. 335-81; Ruth Chang, “Can Desires Provide Reasons for Action?”, from *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, R. Jay Wallace, ed., pp. 56-90.

<sup>20</sup> For arguments they cannot, see Derek Parfit, *ibid.*, p. 121; Joseph Raz, *The Morality of Freedom*, pp. 140-4; E. J. Bond, *Reason and Value*, p. 31; Stephen Darwall, *Impartial Reason*, esp. chapters 3 and 6; Michael Stocker, *Plural and Conflicting Values*, p. 191; G. F. Schuler, *Desire*, pp. 91-7; Scanlon, *ibid.*, pp. 41-9.

There is another kind of reason that demands attention that is not a normative or motivating reason; this is the sort of reason I shall refer to as an *agent's* reason. An agent's reason is the reason that an agent sincerely asserts, or would sincerely assert, is the reason for which she performed some behavior, or for which she will or would have acted. An agent who sincerely asserts her reason is providing what she thinks is at least a minimal justification of her action; at least, an agent who sincerely asserts her reason for acting appears to imply that she believes her action has something to be said for it. Further, an agent who sincerely asserts her reason for acting gives us a clue what her motivating reason is. An agent's sincere avowal of her reason for acting is supposed to help us construct, at least partially, a rationalizing explanation of her action. It is true that an agent's reasons are very often elliptical such that we cannot necessarily discern what an agent's motivating reason is even when she truly states what her reason for acting is. For example, an agent might claim that she donated money to charity because it was the right thing to do. But the rightness of an action is not a psychological state and thus the rightness of her action cannot be her motivating reason. Still, we might reasonably infer that an agent wanted to perform a right action, or that she believed that she was performing a right action, and so forth. So even if an agent's statement of her reasons does not allow us to deduce what her motivating reasons are, her statement may give some idea of what her motivating reasons are.

Still, for a reason *R* to be an agent's reason, it need not be the case that *R* is an effective or motivating reason since it is possible that an agent has mistaken, if sincere, beliefs about the reasons for which she performs some behavior. An agent who represses his Oedipal desire to sleep with his mother might falsely believe that he stays close to

home because he wants to save money. Even though his Oedipal desire is his effective motivating reason, *his* reason for acting is his desire to save money. Further, for a reason *R* to be an agent's reason, it also need not be the case that *R* is a normative reason since it is possible that an agent acts for a reason that utterly fails to justify performing the action. Thus, an agent's reason for acting might fail to either explain or justify her behavior.

It might be puzzling why an agent's reasons deserve attention if they do not necessarily explain or justify behavior. There are some reasons for philosophers to be interested in an agent's reasons, however. First, it seems that in paradigm cases of intentional action, there is overlap between an agent's reasons and her effective motivating reasons, such that in paradigm cases of intentional action, agents have true beliefs about their motivating reasons and true beliefs about why they act. I pursue this point further in Chapter 5 where I discuss cases of paradigmatic or "full-blooded" actions. Second, many of the principles discussed above suppose that an agent's reasons are relevant to the explanation and justification of action. If the deliberative principle is correct, then agents often act on the basis of reflective deliberation and it is plausible to suppose that agents who act as a result of reflective deliberation will have beliefs about what reasons they act for. If the teleological principle is correct, then agents who perform actions have goals that they mean to fulfill and it seems plausible to suppose that an agent who has a goal believes that she has a reason to act, namely, to fulfill her goal. Third, there is long tradition in moral and legal philosophy that stresses that the rightness or wrongness of an agent's action depends, at least in part, on an agent's reasons. Kantians, for example, insist that something like the following is the case: an action is morally praiseworthy only if an agent believed that she acted for moral reasons. Finally,

it is intuitive that agents who act intentionally are at least sometimes in control of what they are doing. While I do not defend any particular account of control here, it is hard to understand how an agent could be in control of what she does if she is mistaken about the reasons for which she acts. An agent who was *always* mistaken about her effective reasons, for example, could hardly be said to be in control of what she is doing since she does not even know why she is acting. But that suggests that an agent's reasons can make a difference with respect to, for example, whether or not an agent is in control of what she does. Thus, it is perhaps not that surprising that some philosophers have supposed that an agent's reasons must at least be similar to her effective reasons, if she is to act intentionally.<sup>21</sup>

In this section, I have attempted to distinguish different kinds of reasons. Failure to distinguish between these kinds of reasons leads to potential confusion, and, I suggest, the problematic syllogism conflates different kinds of reasons. Or so I argue in the following section.

### **'Reason' and Equivocation**

In this section, I discuss the problematic syllogism and the strong Panglossian challenge to the *intentionbad* and *reasonbad* theses. I have conceded that the premises of the problematic syllogism appear to be well-founded and that since the problematic syllogism is a straightforward piece of syllogistic reasoning, it appears to be sound. However, the problematic syllogism as it stands suffers from an informal fallacy: the problematic syllogism only appears plausible if we ignore that the term 'reason' is

---

<sup>21</sup> Robert Audi, for example, includes in his analysis of acting for a reason a belief condition that an agent believes something to the effect that there is a connecting relation between her action and the reason for which she acts; see Audi, "Acting for Reasons", *The Philosophical Review* 45:4, (Oct. 1986), p. 519. David Velleman has a sophisticated discussion of self-awareness in action and argues that agents at least attempt to understand what they are doing in his *Practical Reflection*.

ambiguous. Disambiguating 'reason' in the problematic syllogism makes it appear much less plausible.

Here, again, is the problematic syllogism:

- 1) If an agent is acting intentionally, then she is acting for a reason
- 2) Acting for a reason is acting with the belief that there is a reason for performing the action
- 3) Acting with the belief that there is a reason for performing an action is acting with the belief that there is something about the action that justifies performing it
- 4) Acting with the belief that there is something about an action that justifies performing it is acting with the belief that there is something good about the action

The problematic syllogism yields the following two conclusions:

- 5) If an agent is acting intentionally, then she believes that there is something good about the action she performs.
- 6) Acting for a reason is acting with the belief that there is something good about the action being performed.

Consider 1): If an agent is acting intentionally, then she is acting for a reason. I am inclined to think that 'reason', in 1), refers to a motivating reason and that it is an analytic truth that agents who act intentionally act for motivating reasons. Thus, a disambiguated 1) should read as follows:

1a) If an agent is acting intentionally, then she acts for a motivating reason

1) only seems plausible if 'reason' refers to motivating reasons.

Suppose, for example, that 1) is a thesis about normative reasons. 1) would therefore be disambiguated as follows:

1b) If an agent is acting intentionally, then she acts for a normative reason

1b) seems clearly false since there are intentional actions that have nothing to be said for them and thus that could not be supported by normative reasons. I find it plausible that there are some actions that simply have nothing at all to be said for them. Banging my

head on the desk right now yields absolutely nothing that is good and thus there is no normative reason for me to intentionally bang my head on the desk. Further, consider cases in which we claim that agents act “for no reason.” For example, I might intentionally cross my legs or crack my knuckles and claim sincerely that I do so “for no reason.” Following Ginet, we might insist that these actions are intentional if any actions are.<sup>22</sup> Clearly, however, I do not cross my legs or crack my knuckles for no *motivating* reason; *something* explains my intentionally crossing my legs or cracking my knuckles. Claiming that we act “for no reason” is a *façon de parler*, and, I suggest, is most naturally read as meaning that there is no normative reason that demands that the action is performed.

Actions performed “for no reason” also suggest that 1) is not a thesis about an agent’s reasons. Suppose that 1) were a thesis about an agent’s reasons. A disambiguated 1) would therefore read as follows:

1c) If an agent is acting intentionally then she acts for her reason

However, in cases in which an agent acts “for no reason,” it is not clear what an agent’s account of her own reasons for acting will be. Asking an agent an agent “Why did you do that?” after she has cracked her knuckles intentionally might elicit elliptical responses like “I don’t know—because I wanted to”; it is not clear that this answer implies that an agent has any belief about what her reason for acting is. An agent who claims something like “I don’t know—because I wanted to” might be trying to explain that she has a fetish for cracking her knuckles and just finds it pleasurable. In that case, perhaps she has a reason for cracking her knuckles. Alternatively, an agent who claims something like “I

---

<sup>22</sup> Ginet, *ibid.*, p. 3.

don't know—because I wanted to” might be speaking literally. Perhaps her knuckle cracking was mindless and thoughtless and although she cracked her knuckles intentionally, there was nothing about her action at the time she performed that appealed to her. Thus, *she* lacks a reason for acting. Note that, typically, claiming ‘because I wanted to’ does not quite provide an agent’s reason. An agent’s sincere avowal of her reason for acting is supposed to help us construct, at least partially, a rationalizing explanation of her action. But an agent who only claims that she acted “because she wanted to” does not tell us enough to begin constructing a rationalizing explanation; her response only tells us what we already knew—that there is something about her action that appealed to her and led her to perform it—but it gives us no idea what it is about her action that appealed to her and led her to perform it. Finally, consider cases in which agents act intentionally but have false beliefs about the reasons for which they act, such as is the case with my agent who intentionally stays near mother because of his repressed Oedipal desire. His reason for staying near mother is that he wants to save money, but he does not literally act for his reason, even if he thinks that he does. But then it appears that an agent can act intentionally without having, at the time she acted, an agent’s reason.

If all this is right, then 1) is most plausibly disambiguated as 1a) and read as a thesis about motivating reasons. If 1a) is the properly disambiguated reading of 1), then if the Panglossian means to use ‘reason’ unequivocally, then ‘reason’ in the consequent of 1) should refer to the same thing that it refers to in the antecedent of 2). Still, all of the following are potential readings of 2):

- 2a) Acting for a motivating reason is acting with the belief that there is a motivating reason for performing the action

2b) Acting for a motivating reason is acting with the belief that there is a normative reason for performing the action

2c) Acting for a motivating reason is acting with the belief that there is an agent's reason for performing the action

Rather than directly considering whether 2a) - 2c) is most plausible, consider 3). All of the following are potential readings of 3):

3a) Acting with the belief that there is a motivating reason for performing an action is acting with the belief that there is something about the action that justifies performing it

3b) Acting with the belief that there is a normative reason for performing an action is acting with the belief that there is something about the action that justifies performing it

3c) Acting with the belief that there is an agent's reason for performing an action is acting with the belief that there is something about the action that justifies performing it

I submit that 3b) must be the appropriate disambiguated reading of 3). 3a) seems false since it seems false to suppose that believing that one is acting for a motivating reason entails that one believes there is something that justifies the action being performed. At the very least, even if one believes that one's action is the product of some motivating reason, it is not clear how to infer that there is something about that action that justifies its performance. The Panglossian may find 3c) plausible, but, as I suggest below, 3c) is compatible with very un-Panglossian theses since it remains an open question at this point what can count as an agent's reason; it is an open question at this point whether an agent who believes that she has a reason for acting must also believe that her reason justifies performing her action. Thus, 3b) is the most likely reading of 3).

Since 'reason' as it appears in the consequent of 2) must refer to the same thing that 'reason' as it appears in the antecedent of 3) refers to, if the Panglossian means to use

'reason' unequivocally, then 2b) must be the appropriate disambiguated reading of 2).

The premises of the disambiguated problematic syllogism are as follows:

- 1a) If an agent is acting intentionally, then she acts for a motivating reason
- 2b) Acting for a motivating reason is acting with the belief that there is a normative reason for performing the action
- 3b) Acting with the belief that there is a normative reason for performing an action is acting with the belief that there is something about the action that justifies performing it
- 4) Acting with the belief that there is something about an action that justifies performing it is to act with the belief that there is something good about the action being performed

Whether we are forced to accept the conclusions of the problematic syllogism and deny the *desirebad* and *reasonbad* theses depends upon whether or not the disambiguated problematic syllogism is sound and therefore whether or not its premises are true.

Unfortunately, 2b) seems false. Examples of agents who act "for no reason" seem to be counter-examples to 2b), since when agents act "for no reason" it seems they do not believe there is a normative reason for acting. If 2b) is false, then not all of the premises of the disambiguated problematic syllogism are true and thus the problematic syllogism is not sound. Thus, the proponent of the *intentionbad* and *reasonbad* theses need not accept the conclusions of the problematic syllogism that purport to show the *intentionbad* and *reasonbad* theses are false.

The Panglossian who offers the problematic syllogism as an argument against the *intentionbad* and *reasonbad* theses faces the following dilemma. In its original version, the problematic syllogism appears to be plausible only because of an ambiguity. The disambiguated problematic syllogism does not appear to be sound. Thus, the Panglossian seems forced to admit that the problematic syllogism is either invalid because it suffers

from a fallacy or that it is unsound. Either way, the proponent of the *intentionbad* and *reasonbad* theses need not accept the conclusions of the problematic syllogism and thus need not abandon either the *intentionbad* and *reasonbad* theses.

### **The Problematic Syllogism Redux**

In the previous section, I argued that the problematic syllogism suffers from an ambiguity and that the problematic syllogism appears to be unsound once the ambiguity is removed. But even if this counts as a victory for the proponent of the *intentionbad* and *reasonbad* theses, it is perhaps a rather hollow victory. My argument that the disambiguated problematic syllogism is unsound depended on, for example, cases in which agents mindlessly crack their knuckles. But knuckle-cracking cases do not supply the proponent of the *intentionbad* and *reasonbad* theses with much to explain how it is possible to intentionally pursue the bad and to act for the reason that one's action is believed to be bad and not at all good. The Panglossian thesis seems to be that there is something an agent must believe, *B*, if an agent is to act intentionally and act for a reason. But even if there are knuckle-cracking cases in which an agent acts intentionally, perhaps they are not really cases in which an agent acts for a reason. If we set aside such cases—cases of actions performed “for no reason”—and focus on cases in which an agent *has* a reason for acting, perhaps the Panglossian is correct that an agent must believe *B* to act intentionally and to act for a reason. If so, the problematic syllogism can be revived.

In the discussion that follows, cases of mindless and trivial actions in which an agent acts “for no reason” will simply be ignored. For the remainder of this section, I suppose that an agent who acts for a reason must have her own reason for action. I shall refer to instances in which an agent acts intentionally and has her reason for acting cases of acting intentionally\*. I also suppose that when an agent acts intentionally\*, she not

only has a reason for acting, but her reason for acting actually does produce and explain her acting intentionally\*. I claim that an agent who acts intentionally\* and who acts for a reason, acts *in* the belief that her action is justified, acts in the belief that there is a normative reason that justifies her action, and acts in the belief that there is something good about her action. An agent who acts *in* the belief, I suggest, does not merely act *with* some belief; instead, acting *in* a belief about her reason, for example, implies that her reason actually does produce and explain her action. Thus, in cases of acting intentionally\*, an agent's reason actually tracks the effective reason that explains her action. Unless otherwise stated, the instances of acting for a reason that I discuss in this section are cases of acting intentionally\*, not simply acting intentionally.

With these stipulations in place, consider the following redux version of the disambiguated problematic syllogism:

- 7) If an agent acts for a reason, then she acts in the belief that the action is justified
- 8) If an agent acts in the belief that the action is justified is, then she acts in the belief that there is a normative reason that justifies the action
- 9) If an agent acts in the belief that there is a normative reason that justifies the action, then she acts in the belief that there is something good about the action

The conjunction of 7) - 9) yields the following conclusion:

- 10) If an agent acts for a reason, then she acts in belief that there is something good about the action

Note that the redux version of the problematic syllogism does not suffer from any obvious equivocations and, given the *inter alia* clause, it can allow that motivating reasons have a role to play in the production and explanation of action. And 10) does appear to imply that the *intentionbad* and *reasonbad* theses are false. If acting intentionally\* is acting for a reason, then it is not possible for an agent to believe that her

action is bad and not at all good and still act intentionally\*, if 10) is true; thus, the *intentionbad* thesis would be false. And it seems that 10) precludes acting for the reason that the action is believed to be bad and not at all good, since acting for a reason involves believing that there is something good about the action being performed; thus the *reasonbad* thesis appears to be false, if 10) is true. Thus, the proponent of the *intentionbad* and *reasonbad* theses must also resist the redux version of the problematic syllogism. I shall attempt to do so in the remainder of this chapter.

It is worth noting that 10) actually may be consistent with the *reasonbad* thesis, even if it is not consistent with the *intentionbad* thesis. Again, the *reasonbad* thesis states that the believed badness of an action can be the acting agent's reason for acting. Suppose that 10) is true and that it is a necessary condition of acting for a reason, and thus for acting intentionally, that an agent believes there is something good about her action; it does not follow that *her* reason for acting is that her action is believed to be good. It is possible to believe that an action would be good to perform and to act for a different reason. Imagine a variant of the Kant's shopkeeper. Suppose the shopkeeper believes that there is something good about providing exact change: perhaps he believes that it is good that people get what they deserve. Suppose the shopkeeper also believes providing exact change will lead to the ruin of his business since he believes that all of his competitors do not give exact change and that he cannot unless he too cheats his customers. So suppose that the shopkeeper also believes that there is something bad about giving exact change: it will lead to the ruin of his business. Perhaps the shopkeeper does not intentionally pursue the bad since he believes there is something good about his giving exact change. But suppose that the shopkeeper is frustrated with trying to

compete with corrupt competitors and decides to give exact change for the reason that it would lead to his own ruin. If this is possible, then the shopkeeper would intentionally give correct change and his reason for doing so would be that giving correct change is bad. The shopkeeper would act for a reason and would act intentionally, but he would not act for the reason that the Panglossian hoped.

For all that, it is not clear that the redux version of the problematic syllogism is sound and thus it is not clear that the redux version does block the *intentionbad* thesis. I suggest that the proponent of the *intentionbad* and *reasonbad* theses can accept 7), and perhaps can accept 8); if the proponent of the *intentionbad* and *reasonbad* theses can accept 7) then there is reason to suspect that 8) is false and if the proponent of the *intentionbad* and *reasonbad* theses can accept 8) then there is reason to suspect that 9) is false.

Consider 7): if an agent acts for a reason, then she acts in the belief that the action is justified. Above, I suggested that Davidson's principle supports 3); Davidson's principle also appears to support 7). Again, Davidson's principle states that rationalizing explanations indicate that from the acting agent's point of view, her action is justified because she believed her action had something to be said for it. Thus, if Davidson's principle is correct, then there is a conceptual connection between thinking an action is justified and thinking an action has something to be said for it. The proponent of the *intentionbad* and *reasonbad* theses can allow that an agent who intentionally\* pursues the bad or acts for the reason that her action is believed to be bad and not at all good thinks that her action is justified. This is because the proponent of the *intentionbad* and *reasonbad* theses can allow that an agent who intentionally\* pursues the bad or acts for

the reason that her action is believed to be bad thinks that her action has something to be said for it. Davidson's principle leaves it open just why an agent thinks that her action is justified and everything depends on what an agent thinks is sufficient for an action to have something to be said for it. Perhaps an agent believes that the believed badness of an action is what the action has to be said for it. Perhaps an agent like Carla believes that actions that lead to her being harmed are justified because she thinks that her being harmed, while not a good thing, has something to be said for it. Perhaps a sufficiently frustrated agent performs actions that she believes are utterly futile because she takes the futility of those actions to be something to be said for them. Perhaps a self-loathing agent believes that an action that harms her has something to be said for it because that action results in her being harmed. If it possible to have perverse beliefs about what an action has to be said for it, or for that matter about what sorts of considerations justify an action, then it is possible for an agent to believe that some action believed to be bad and not at all good has something to be said for it. Thus, even if Davidson's principle is correct, it may be possible for an agent to believe that some action believed to be bad and not at all good is justified.

The argument in the previous paragraph depends on the possibility of agents having perverse beliefs about, for example, what it is for an action to have something to be said for it. But it is far from clear why it is not possible for agents to have perverse beliefs of this sort. I have already argued that some moods and emotions might cause an agent to desire what she believes is bad and not at all good. Similarly, moods and emotions might cause an agent to have perverse beliefs about what an action has to be said for it and what

it is about an action that justifies it. Dostoevsky's man from the underground suggests he has perverse beliefs about what his actions have to be said for them:

“So I'm bothering you, straining your hearts, not letting anyone in the house sleep. Don't sleep, then; you, too, should feel every moment that I have a toothache. For you I'm no longer a hero, as I once wished to appear, but simply a vile little fellow, a *chenapan*. Well, so be it! I'm very glad you've gotten to the bottom of me. It's nasty for you listening to my mean little moans? Let it be nasty, then; here's an even nastier roulade for you. . .”

You're laughing? I'm very glad. To be sure, gentlemen, my jokes are bad in tone—uneven, confused, self-mistrustful. But that is simply because I don't respect myself.<sup>23</sup>

Admittedly, it is difficult to understand the underground man's reasoning here and he himself cautions against taking his ruminations too seriously. Yet it appears that the underground man thinks that moaning loudly to keep everyone up is nasty and that it is because his moaning is nasty that his action has something to be said for it. The underground man seems possible. So it seems possible that agents can have perverse beliefs about what counts in favor of an action. So long as agents can have perverse beliefs about what counts in favor of an action, then agents can intentionally\* pursue the bad and act for the reason that the action is believed to be bad and not at all good consistently with 7). Thus, the proponent of the *intentionbad* and *reasonbad* theses can accept 7).

It is unclear what we should say about 8), if 7) is consistent with the *intentionbad* and *reasonbad* theses. It is simply not clear whether or not an agent who thinks that her action is justified because, for example, she thinks that her action is bad and not at all good can also think that there is a normative reason that justifies her action. Perhaps the only thing that could be believed to count as a normative reason for an action is a

---

<sup>23</sup> Dostoevsky, *ibid.*, pp. 13-4.

consideration that implies the action is good. Raz appears to accept something like this since he argues that only facts which show an act to be good constitute a reason for performing the action.<sup>24</sup> In contrast, however, recall that Smith claims that a normative reason is some consideration that justifies an agent's behavior "from the perspective of the normative system." I am uncertain what Smith thinks constitutes a normative system, but perhaps there are perverse normative systems that generate perverse requirements. Perhaps there are perverse normative systems that imply that the fact that an action is bad and not at all good is a normative reason for performing the action. If there are perverse normative systems, then it is possible that those perverse normative systems generate perverse normative reasons. If there are perverse normative reasons, then the proponent of the *intentionbad* and *reasonbad* theses can accept 8).

Suppose, however, that perverse normative reasons are incoherent. The proponent of the *intentionbad* and *reasonbad* theses can *still* accept 8). 8) only implies that if an agent acts with a belief that her action is justified, she will believe that there is a normative reason that justifies her action. Even if there are no perverse normative reasons, it is possible for agents to *believe* that there are perverse normative reasons or to believe that, for example, the fact that an action is bad and not at all good is a normative reason. So perhaps the proponent of the *intentionbad* and *reasonbad* theses can accept 8).

Thus, it appears that if the redux version of the problematic syllogism is to make trouble for the proponent of the *intentionbad* and *reasonbad* theses, the Panglossian must establish 9): if an agent acts in the belief that there is a normative reason that justifies the

---

<sup>24</sup> Raz, "Agency, Reason, and the Good", p. 24.

action, then she acts in the belief that there is something good about the action. I am not aware of any philosopher who explicitly defends 9). Perhaps it is simply unclear how an action that is believed to be bad and not at all good could also be believed to be justified. The proponent of the *intentionbad* and *reasonbad* theses objects that, when explaining action, we do not need to suppose the agent took there to be facts which show her act to be good. In response, Raz claims:

The [objector] is, therefore, committed to the availability of an explanation of how it is that non-good making qualities make an action eligible. I find it difficult to imagine such an explanation.<sup>25</sup>

There are at least two different readings of this passage. On one reading of this passage, Raz is not committed to any distinctly Panglossian thesis that an agent who acts intentionally must believe her action is at least minimally good. Instead, Raz may only be committing himself to a negative thesis about what is required for acting intentionally; Raz may only demand that, if an agent is to act intentionally, then that agent must *not* believe that her action is bad and not at all good, or that if an agent is to act intentionally, then that agent must *not* believe that her reason for acting is the believed badness of her action. The argument for the proposed negative belief constraint appears to rest on an inability to imagine how non-good making-qualities make an action eligible. I shall return to this argument below. Note, however, that Raz's account of what it is for an action to be "eligible" suggests he does endorse something like 9). Raz claims that the explanation of action must include:

[W]hat the agent took to be facts which show the act to be good, and which therefore, constitute a reason for its performance, making it eligible.<sup>26</sup>

---

<sup>25</sup> Raz, *ibid.*, pp. 26-7.

It thus appears that it is at least sufficient for an action to be eligible that, *inter alia*, an agent took there to be facts which show the act to be good. Further, it is apparent from the above quotation that Raz does suppose there is some link between thinking that an action is good and there being a reason for performing the action. Thus, it appears that Raz does accept something very close to 9).

In fairness to Raz, he does offer arguments intended to show that only good-making qualities could make an action “eligible,” beyond merely gesturing that the *intentionbad* and *reasonbad* theses are incomprehensible. Raz argues that:

The difficulty in explaining the eligibility of actions in ways other than by reference to good-making qualities... is of finding conceptual room for an alternative. . . . It would seem to follow that those who do not hurt others, or who deny that the fact that an action would hurt others is in and of itself a reason to perform it, are irrational, or at least imperfectly rational, for they fail to acknowledge such reasons.<sup>27</sup>

In a similar vein, Raz argues that:

That an action will hurt another is no reason for performing it for if it were then everyone would have such a reason and those who did not act for them would be irrational. That is absurd. . .<sup>28</sup>

I find these remarks puzzling. In response to the first of Raz’s arguments quoted immediately above, it is not true that if an agent fails to acknowledge a normative reason for acting, the agent is irrational. Some agents are simply not in position to know about some normative reasons that apply to them: it is plausible to suppose that my British sailors and cavemen could not have been responsive to and could not have acknowledged that they had reasons to eat citrus fruits and brush their teeth, but it hardly follows that

---

<sup>26</sup> Raz, *ibid.*, p. 24.

<sup>27</sup> Raz, *ibid.*, p. 28.

<sup>28</sup> Raz, *ibid.*, p. 26.

they were irrational. Further, in response to the second of Raz's arguments, it is not clear why if something is a normative reason, then every agent must have that normative reason to act. Finally, as I argue above, the proponent of *intentionbad* and *reasonbad* theses is simply not committed to claiming that, for example, the fact that some action would hurt someone is a normative reason. The proponent of the *intentionbad* and *reasonbad* theses is only committed to claiming that the believed badness of an action can be thought to be what justifies performing an action or to constitute what an action has to be said for it.

Return to Raz's claim that:

The [objector] is, therefore, committed to the availability of an explanation of how it is that non-good making qualities make an action eligible. I find it difficult to imagine such an explanation.

Again, I am not certain what it is for an action to be eligible, but if Raz means to object to the *intentionbad* and *reasonbad* theses, he must mean to claim that:

The [objector] is, therefore, committed to the availability of an explanation of how it is that non-good making qualities make an action eligible **from the point of view of the acting agent**. I find it difficult to imagine such an explanation.<sup>29</sup>

But this is a challenge that the proponent of the *intentionbad* and *reasonbad* theses can meet. I have attempted to provide the beginnings of an explanation of how non-good making qualities can make an action eligible from the point of view of the acting agent. Given the appropriate moods or emotions, agents might come to acquire desires for what they believe to be bad and not at all good; they might be disposed to count as their reasons for acting considerations that show an action to be bad and not at all good; they might acquire evaluative principles that, for example, actions believed to be bad and not

---

<sup>29</sup> The boldface passage is my addition.

at all good are justifiable. If all of this is possible, then I submit it is possible that non-good making qualities could be thought to by an agent to make her action eligible.

If the final remarks of the preceding paragraph are correct, then we can explain why even the redux version of the problematic syllogism fails to show that the *intentionbad* and *reasonbad* theses are false. The proponent of the *intentionbad* and *reasonbad* theses can claim that 9) is false, even if 7) and 8) are true: given the appropriate moods and emotions are in place, an agent can believe there is a normative reason that justifies her action yet fail to believe there is anything good about the action. Suppose, however, that it is impossible for an agent to believe that there is a normative reason that justifies her action while also believing that her action is bad and not at all good. Then, I submit, 8) is false: given that an agent has perverse beliefs about what justifies an action, it is possible for an agent to believe that the action she believes is bad and not at all good is justified. Either way, the redux version of the problematic syllogism is not sound.

### **Conclusion**

I have argued in this chapter that the proponent of the *intentionbad* and *reasonbad* theses can resist the problematic syllogism. My arguments depend partly on the hypothesis that agents can have perverse beliefs about what justifies an action, about what counts as a reason, and so forth. Thus, what can count as an agent's reason fluctuates with the contours of an agent's psychological profile. Admittedly, agents with what we take to be standard and healthy psychological profiles will accept reasonable ethical principles, and will have reasonable judgments about what justifies an action and counts as a normative reason to act. But perhaps there is a danger in appealing to agents with abnormal and unhealthy psychological profiles in defense of the *intentionbad* and

*reasonbad* theses. I want to consider two objections to my response to the problematic syllogism that will in turn motivate the following chapter.

First, recall that I have suggested that moods and emotions can lead agents to desire what they believe to be bad and not at all good, to count the fact that an action is believed to be bad and not at all good as a reason to act, and so forth. It may be objected, however, that this does not properly or sufficiently explain how agents come to be in the position where they desire the bad or count non-good making considerations as reasons. Further, only appealing to moods and emotions does not satisfactorily explain how agents could, for example, acquire a desire for the bad as a result of deliberation or could deliberately and reflectively count the fact that an action is believed to be bad and not at all good as a reason for performing the action. I have suggested that these things are possible in previous chapters, but I have not explained why they are possible and I have only gestured and some sort of black box that somehow produces perverse desires and evaluative beliefs. But that is not an adequate explanation. Call this the *objection from the black-box*.<sup>30</sup>

I think that the objection from the black-box is overstated. I have argued, for example, that instances of the intentional pursuit of the bad admit of rationalizing explanations. The objection from the black-box appears to demand not only that intentional actions are rationalized, but that everything that contributes to explanations of intentional action must be subject to something like rationalizing explanation. Since my objector does not accept that claiming emotions and moods cause agents to acquire perverse desires and perverse evaluative principles, she must mean to demand some other

---

<sup>30</sup> A remark by Judith Hill inspired the name of this objection.

kind of explanation. The most obvious other kind of explanation that we have been considering is rationalizing explanation. But much of what produces standard kinds of intentional action is not subject to rationalizing explanation. Perhaps my desires for what I believe to be good do not *themselves* admit of rationalizing explanations. Perhaps the standard evaluative principles that I accept do not *themselves* admit of rationalizing explanations. There are surely brute facts about our psychological profiles, insofar as there are some psychological states that can contribute to rationalizing explanations of our actions that are not themselves subject to rationalizing explanations. Thus, I submit, the objection from the black-box demands too much and the fact that I must appeal to contingent facts about the psychology of some agents does not show that my defenses of the *desirebad* and *intentionbad* and *reasonbad* theses are inadequate. There is no reason that moral psychologists should not appeal to psychology for help.

Strangely, one of my most spirited Panglossian opponents appears to agree. Raz claims that when “people take what they believe to be bad-making features to be their reasons [for acting]. . . they act for *anomic reasons*.”<sup>31</sup> Raz also refers to “anomic agents” who are, I suppose, agents who act for anomic reasons.<sup>32</sup> I welcome Raz’s locution and I offer the following expanded definition: anomic reasons are an agent’s reasons and are considerations that an agent thinks show her action to be bad and not at all good. Raz also claims that “the appeal of contrariness” explains “the allure” of anomic reasons.<sup>33</sup> He also claims that “the appeal of contrariness is an established

---

<sup>31</sup> Raz, *ibid.*, p. 32, emphasis added.

<sup>32</sup> Raz, *ibid.*, p. 33.

<sup>33</sup> Raz, *ibid.*, p. 33.

psychological phenomenon,” that contrariness is “a psychological explanation” of the actions of anomic agents, and that it is through “some corruption of a psychological process” that anomic reasons could be taken to be genuine normative reasons for action.<sup>34</sup>

So I have argued.

However, it is clear that Raz still wants to resist the *desirebad* and *intentionbad* and *reasonbad* theses. Consider Raz’s following argument:

I am facing the proverbial saucer of mud, and proceed to eat it, moving my hands and mouth as I would normally do when eating. *Something has gone wrong with me.* In the absence of any good-making characteristic which I believe eating the mud possesses *I will not be able to understand what I am doing.* I will be more horrified at myself behaving in this way than other people will be. For to me this will signify *that I have lost control over myself*, that I am possessed by something which makes me act in ways *I do not understand*, ways which I disavow, protesting that it is not really me.<sup>35</sup>

Focus on Raz’s dual claims that agents who act for anomic reasons will not be able to understand themselves and what they are doing and that agents who act for anomic reasons will not be in control of themselves and what they are doing. These remarks suggest a second objection to the theses I have defended and the strategy I have used to defend them. It is a sign perhaps, of agents who are psychologically healthy that they know what they are doing and why they are doing it; agents who suffer from the kinds of psychological maladies that interest Freudians are said to suffer from mental illnesses. If Raz is right that agents who act for anomic reasons cannot understand themselves or what they are doing, then those agents are hardly paradigmatic examples of agents. Further, insofar as I claim that agents who, for example, intentionally pursue the bad, are in the

---

<sup>34</sup> Raz, *ibid.*, p. 33.

<sup>35</sup> Raz, *ibid.*, p. 32, emphasis added.

grip of various moods and emotions, I have perhaps implied that those agents are less than psychologically healthy. Perhaps agents who are less than psychologically healthy, to put the point lightly, are so far removed from what counts as a paradigmatic agent they simply do not deserve to be called agents at all. If they do not deserve to be called agents, then perhaps their behavior does not deserve to be called intentional action. Similarly, perhaps instances of intentionally pursuing the bad are so far removed from what counts as a paradigmatic instance of intentional action that they simply do not deserve to be called intentional actions at all.

In the following chapter, I want to make the case that instances of the intentional pursuit of the bad are not nearly as far removed from instances of paradigmatically intentional action, nor need they be unable to understand themselves or be out of control, as Raz's remarks suggest.

CHAPTER 5  
UNDERSTANDING, CONTROL, AND INTENTIONALLY PURSUING THE BAD

Oh, tell me, who first announced, who was the first to proclaim that man does dirty only because he doesn't know his real interests; and that were he to be enlightened, were his eyes to be opened to his real, normal interests, man would immediately stop doing dirty, would immediately become good and noble, because, being enlightened and understanding his real profit, he would see his real profit precisely in the good. . . What is to be done with the millions of facts testifying to how people *knowingly*, that is, fully understanding their real profit, would put it in second place and throw themselves onto another path. . . precisely as if they simply did not want the designated path, and stubbornly, willfully pushed off onto another one, difficult, absurd, searching for it all but in the dark.

Fyodor Dostoevsky, *Notes From the Underground*<sup>1</sup>

**Introduction**

In earlier chapters, I attempted to explain how it is possible for agents to desire the bad and to intend the bad. In the previous chapter, I argued that the believed badness of an action can be an agent's reason for acting. However, even if the Panglossian is forced to accept the truth of the *desirebad* and *reasonbad* theses, it does not follow that she must accept the *intentionbad* thesis. Agents can have desires for things they never intentionally pursue and an agent's reason is not necessarily what explains her actions. So the *intentionbad* thesis has yet to be established, even if there are grounds for supposing it is true, as I suggest at the end of Chapter 3.

Yet as I note at the end of Chapter 4, the Panglossian may insist that it is a mistake to speak of agents "intentionally" pursuing the bad and that speaking of the "intentional"

---

<sup>1</sup> Dostoevsky, *Notes From Underground*, pp. 20-1.

pursuit of the bad is a misnomer. In particular, while the Panglossian may allow that agents can pursue what they believe is bad and not at all good, the Panglossian will deny that they intentionally pursue what they believe to be bad and not at all good and that their behavior counts as intentional action. Raz is concerned that an agent who finds herself pursuing what she believes to be bad and not at all good will not be able to understand herself and what she does and will not be in control of herself and what she does.<sup>2</sup> Further, it seems plausible to suppose that an agent who does not understand what she is doing and is not in control of what she is doing cannot be acting intentionally. Since Raz challenges the proponent of the *intentionbad* thesis to explain how agents who pursue the bad understand themselves and what they do and how they can be in control of themselves and what they do, I shall call this challenge *Raz's Challenge*.

Dostoyevsky's man from the underground appears to disagree with Raz. In this chapter, I side with the man from the underground. It is intuitive that agents who act intentionally understand themselves and what they are doing and that agents who act intentionally are in control of themselves and what they are doing. In this chapter, I argue that agents who pursue the bad can understand themselves and what they are doing and be in control of themselves and what they are doing. I shall argue as follows:

- 1) Agents who act intentionally understand themselves and what they do and are in control of themselves and what they do in virtue of meeting some set of requirements, *N*, for having understanding and control
- 2) At least some agents who pursue the bad meet the requirements, *N*, for having understanding and control
- 3) Therefore, at least some agents who pursue the bad act intentionally and thus pursue the bad
- 4) Therefore, the *intentionbad* thesis is true

---

<sup>2</sup> Raz, "Agency, Reason, and the Good", p. 35.

This argument, if cogent, would rebuke Raz's challenge and establish the truth of the *intentionbad* thesis. My tasks in this chapter are, first, to provide a plausible account of what is included in *N*, and second, to defend 2) and explain how agents who pursue the bad could meet the requirements included in *N*.

### Goals and Having Goals

It is reasonably well accepted in the philosophy of action that the explanation of intentional action is teleological explanation.<sup>3</sup> A teleological explanation of action explains by citing teleological entities such as goals and purposes and ends in the *explanans* of the action.<sup>4</sup> One reason we expect intentional actions to be subject to teleological explanations is because, intuitively, intentional action is goal-directed and purposeful behavior and so forth. And surely goal-directed and purposeful behavior will be explicable in terms of goals and purposes and so forth. Since instances of the pursuit of the bad are purportedly instances of intentional action, instances of the pursuit of the bad should be subject to teleological explanation as well. If so, then instances of the

---

<sup>3</sup> Charles Taylor, *The Explanation of Behavior*, p.6; George Wilson calls the thesis that the explanation of action is teleological a "truism"; See Wilson, *The Intentionality of Human Action*, p.167; Dretske appears to accept that the explanation of action is teleological explanation since he defends the thesis that behavior can be explained in terms of reasons and goals and purposes and so forth; Dretske, *Explaining Behavior: Reasons in a World of Causes*; Michael Smith claims that reasons explanations of action are teleological explanations, "The Humean Theory of Motivation", p. 44; Donald Davidson claims that he is defending an account of explanation that is both causal and teleological, "Actions, Reasons, and Causes", p. 9; R. Jay Wallace claims that "To act intentionally, is necessarily to be in a goal-directed state", in "How to Argue About Practical Reason", *Mind* 99 (1990), pp. 359.

<sup>4</sup> To be sure, I am stipulating the senses of 'teleology' and 'teleological explanation' that I think are relevant here. It may be that these senses do not translate to other areas of philosophical discourse, such as the philosophy of biology. Perhaps explanations of human behavior are teleological but are a *sui generis* kind of teleological explanation that differs significantly from other kinds of teleological explanation, for example, explanation of plant behavior. Note however that Taylor claims that the explanation of human action is the paradigm of teleological explanation; Taylor, *ibid.*, p. 26.

pursuit of the bad must be explicable in terms of goals and they must be instances of goal-directed behavior.<sup>5</sup>

Much, then, rests on what it is to have a goal and on what it is for behavior to be goal-directed. Suppose, for example, that an agent who pursues the bad does not have a goal; her behavior would therefore not be susceptible to teleological explanation. But that suggests that her behavior, whatever it is, is not intentional action. So perhaps the Panglossian ought to be in the business of claiming that agents who pursue the bad do not have goals and therefore that their behavior is not goal-directed. Note that the thesis that instances of the pursuit of the bad are not goal-directed and are not the product of an agent's goals supports Raz's challenge. If an agent does not have a goal when she acts, it is plausible to suppose she could not understand herself or what she is doing since she herself has no idea why she is moving her body in this way or what would constitute successfully completing her behavior. For similar reasons, an agent with no goal could hardly be in control of herself or what she is doing since, by her own lights, there is no end to which she moves her body. The challenge for the proponent of the *intentionbad* thesis, then, is to make it plausible to suppose that agents who pursue the bad have goals and that their behavior is goal directed and therefore subject to teleological explanation just like other intentional actions. I shall initially discuss what it is to have a goal and then what it is for behavior to be goal-directed.

On some accounts of goals, having a goal is a relatively simple matter. Some philosophers have implied that simply desiring is tied up with having a goal. Michael

---

<sup>5</sup> I use the terms 'goal' and 'purpose' and 'end' interchangeably, since, as far as I can tell, they are synonymous; my dictionary lists them as synonyms, at any rate. Also, I use the locutions 'goal-directed' and 'purposeful' interchangeably since, as far as I can tell, they are synonymous.

Smith has argued that having a motivating reason *is, inter alia*, having a goal, and that having a motivating reason *is, inter alia*, desiring.<sup>6</sup> We have previously encountered Donald Davidson's thesis that rationalizing explanations of intentional action explain by citing what it is about an action that appealed to an agent.<sup>7</sup> It seems intuitive that an agent who performs an action because there is something about it that appeals to her has a goal—namely, to bring about that which appeals to her. Since Davidson claims that rationalizations invoke desires, it seems that Davidson will also accept that having a goal is tied up with desiring.<sup>8</sup> Dretske speaks of goals producing behavior and action and, at one point, speaks of purposes and wants interchangeably.<sup>9</sup>

The thesis that having a goal is to be identified with having a desire is consistent with a rather minimal account of what it is for something to be a goal. David Velleman claims that an end is “conceived by an agent as a potential object of his actions.”<sup>10</sup>

Consider also Taylor's account of purposes in his discussion of purposive behavior:

. . . the events productive of order in animate beings are to be explained. . . in terms of the very order which they produce. These events are held to occur because of what results from them, or, to put in a more traditional way, *they occur “for the sake of” the state of affairs which follows*. And this of course is part of what is meant by the term ‘purpose’ when it is invoked in explanation.<sup>11</sup>

---

<sup>6</sup> See Smith, *ibid.*, p. 55.

<sup>7</sup> Davidson, *ibid.*, p. 3.

<sup>8</sup> Davidson, *ibid.*, p. 4.

<sup>9</sup> Dretske, *ibid.*, pp. 109-36, and p. 114 esp.

<sup>10</sup> Velleman, “The Possibility of Practical Reason”, , p. 191.

<sup>11</sup> Taylor, *ibid.*, p. 5, my emphasis.

The “order’ that Taylor seems to be referring to here just is something like purposeful behavior. Similarly, with respect to goals, Taylor claims that:

When we say that actions are classified by their goals, we mean not only that they are classified by the result which in fact is brought about by them, *but also by that end to which they are aimed*; and that is why we speak of a ‘goal.’<sup>12</sup>

Taylor’s remarks are preliminary but they suggest a minimalist account of what goals are.

First, goals are crucially involved in the production of action and behavior. But Taylor also suggests that behavior, when it is explained by purposes and goals, occurs “for the sake of the state of affairs which follows” and that actions come about because of the “end to which they are aimed.” To borrow from Velleman’s earlier suggestion, if an end is what an agent conceives of as a potential object of her action, an agent can only believe that her action has successfully been performed if she also believes that she has attained that object. If this minimal account of goals is acceptable, then we can also develop a minimal account of what it is to have a goal: to have a goal is to be motivated to bring about some particular event or state of affairs by performing some action, partly for the sake of realizing that goal, and the realization of that goal is required for the successful performance of that action.

On this minimal account of what it is to have a goal, having a desire seems to imply that an agent has a goal. An agent who desires something is presumably motivated to pursue that thing, at least to some minimal degree. And since an agent desires to pursue something in particular, there is some criterion of success with respect to an action performed because of that desire: if I desire a mango and reach into the refrigerator and pull out a plum, then I have failed to intentionally acquire a mango. Perhaps, then, an

---

<sup>12</sup> Taylor, *ibid.*, p. 27, my emphasis.

agent who has a desire for something she believes to be bad has a goal. And if agents can act on desires for the bad, then agents can have goals when they pursue the bad.

Not all philosophers accept a minimal account of what it is to be a goal or to have a goal, however. Raz, for example, claims that having a goal is not desiring, because unlike desires:

Goals are our goals because in our actions we have set on pursuing them, because they play an important role in our emotional and imaginative life, because our success or failure in pursuing them is going to affect the quality of our life [and because]. . . they represent what matters to us in life. . .<sup>13</sup>

Some of what Raz claims about goals is, I think, mistaken. Recall that we are supposing that the explanation of intentional action is teleological and therefore that intentional actions are to be explained in terms of an agent's goals. If Raz is correct, then it would seem to follow that all of our actions must be ultimately aimed at realizing something that plays an important role in our life, that will greatly affect the quality of our life if we obtain it, and so forth. But that seems wrong. I can intentionally pursue something that is of very little value to me or something I want but have little invested in, and if I fail to successfully obtain that thing, the quality of my life will hardly be affected; acting on whims and actions performed "for no reason" seem to be actions of this kind. But given Raz's account of goals and given the thesis that intentional actions are to be explained teleologically in terms of an agent's goals, Raz is committed to claiming that we cannot act intentionally on whims and that we cannot act intentionally "for no reason."

However, focus on Raz' claim that when we have a goal, "we have set on pursuing" that goal. It does seem possible for me to have a desire that I am simply resolved never to act on or that I am alienated from, or I might have many conflicting

---

<sup>13</sup> Raz, "Incommensurability and Agency", from his *Engaging Reason*, pp. 63-4.

desires and have not settled on what to do, and so forth. In those cases, it seems plausible that I might have no goal corresponding to the realization of the object of my desire, just because I am not “set on pursuing” the object of my desire.

Suppose, then, that desiring is not always sufficient for having a goal because having a goal implies being set on pursuing something. If not being settled on pursuing something implies that an agent does not have a goal, perhaps being settled on pursuing something implies that an agent has a goal of pursuing that thing. Call this *the settlement account* of having a goal. Recall that one way of settling on what is to be done is to intend to pursue it, as discussed in Chapter 3. Perhaps, then, intending to perform some action or intending to pursue something implies having a goal. This result is consistent, however, with the thesis that agents who pursue the bad have the goal of doing so. If intending implies having a goal, then given that agents can intend the bad as I argued in Chapter 3, then agents can have the goal of obtaining something they believe to be bad and not at all good. Settlement accounts preclude having the goal of obtaining what is believed to be bad and not at all good only if the only way of being settled on pursuing something is to think that the thing is good. But there are other ways of being and becoming settled on pursuing something. Being settled amounts to ending deliberation and accepting the outcome of one’s deliberation; an agent might decide to end deliberation because she is happy with the result or she might simply end deliberation and accept the result out of frustration or exasperation. In situations where an agent ends deliberation because she is frustrated or exasperated, there is no guarantee that the outcome of her deliberation will direct her to something that she thinks is good.

On both minimal and settlement accounts of having a goal, it is possible to have the goal of pursuing the bad. But perhaps there are other accounts of having a goal that rule out having the goal of pursuing what is believed to be bad and not at all good. Such accounts would imply that to have a goal of pursuing something, that thing must be thought to be at least minimally good. I know of no account of having a goal that implies this result, but call an account of this kind an *honorific account* of having a goal.

I find honorific accounts of having a goal implausible. Again, suppose that the explanation of action is teleological explanation and that actions are to be explained in terms of an agent's goals. Recall the discussion of actions performed "for no reason" from Chapter 4. I claimed that agents who act for no reason lack the belief that their action is justified by any normative reason; this is rather close to claiming that agents who act for no reason do not believe that there is anything good about the action they perform. Proponents of honorific accounts are left with a dilemma: they must either claim that agents who "act for no reason" do not really perform intentional actions since they do not have a goal or they must deny that the explanation of action is teleological.<sup>14</sup>

Unless there is an adequate defense of some honorific account of what it is to have a goal, it seems plausible that agents who desire and intend the bad can have goals corresponding to their desires and intentions for the bad. Further, if it possible for an agent to have the goal of pursuing the bad, her pursuit of what she believes to be bad and not at all good might count as goal-directed behavior. Given the above suggestion that a

---

<sup>14</sup> It may be objected that anyone who accepts that the explanation of intentional action is teleological must be concerned about actions performed for no reason since it is not clear that agents who act for no reason have any goal or end or purpose. This result need not follow. An agent who sticks her finger in a puddle of goo for no reason might have the goal of sticking her finger in a puddle of good or for the purpose of acting for no reason.

goal both produces intentional action and counts as the criterion for the successful performance of the action, perhaps goal-directed behavior is simply behavior that is produced that, at least sometimes leads to the realization of that goal. Things are unfortunately not so simple, as I suggest below. Supposing that agents can have the goal or obtaining what they believe to be bad and not at all good, not just any behavior produced by that goal realizes that goal will necessarily count as goal-directed behavior, nor will it necessarily count as intentional action.

### **Goal-Directed Behavior and Intentional Action**

Suppose that an agent is in a mental state, *M* and that being in *M* suffices for having a goal of bringing about some state of affairs, *G*. Suppose also that *M* produces some behavior, *B*. Here is one account of goal-directed behavior adapted from a proposal of Taylor and Wright:

Some behavior, *B*, is goal-directed just in case *B* is produced by *M*, partly because *B*-ing tends to bring about *G*.<sup>15</sup>

In a simple case, an agent calls out to his beloved who happens to be across the street in order to catch her attention; his calling out is perhaps goal-directed behavior because it is produced by a desire to catch her attention and because calling out to her tends to bring it about that he catches her attention.

The account of goal-directed behavior I attribute to Taylor and Wright may not be adequate as a teleological account of goal-directed behavior. First, *M* might produce *B*, partly because *B*-ing tends to bring about *G*, but do so as a result of a deviant causal chain. To borrow a familiar example, a mountain climber might have the goal of killing

---

<sup>15</sup> Taylor, *ibid.*, p. 5 and Wright, *Teleological Explanations*, p. 39. In what follows, I borrow from Dretske's discussion and criticisms of Taylor and Wright's proposal; see Dretske, *ibid.*, pp. 111-2.

his climbing companion and releasing a climber from his support rope does tend to bring about killing mountain climbers. Further, the mountain climber's goal might lead him to release his climbing companion, but only because his goal makes him nervous and causes his hand to shake involuntarily. This may be goal-directed behavior but it certainly is not a case of intentional action since involuntary bodily movements are not intentional actions.<sup>16</sup> For behavior to be both goal-directed and intentional, it must be the case that the agent's goal is *appropriately* related to her behavior. Second, it is unclear that *B*-ing must tend to bring about *G* if *B*-ing is goal-directed. Counterfactually, *B*-ing now might produce *G*, even if no one has ever *Bed* previously and thereby brought about *G*. Relatedly, it might be that *B*-ing used to tend to produce *G* but no longer does. Perhaps *B*-ing never did produce *G*. Still, an agent who believes falsely that *B*-ing produces *G* might perform *B* and although she fails to produce *G*, her *B*-ing is still goal-directed. Finally, and relatedly, even if *B*-ing does or will tend to bring *G*, it does not follow that *B*-ing that realizes *G* produced by *M*, will be goal-directed behavior; unbeknownst to me, some behavior of mine might bring about one of my goals. Suppose I go to the home of my rival who I want to defeat and I ring his doorbell to get his attention and begin my plan of defeating him. Suppose that he is fiddling with the wiring of his doorbell inside just as I push his doorbell I unknowingly kill him and thus defeat him. My goal partly explains my killing him and electrocuting your rivals does tend to defeat them, but if I do not know that I am killing him, then my behavior could hardly count as goal-directed intentional action.

---

<sup>16</sup> The example is adapted from Davidson, "Freedom to Act", reprinted in his *Essays*, p. 79.

In light of the above criticisms, I propose the following emendations of Taylor and Wright's account of goal-directed behavior:

Some behavior, *B*, is goal-directed just in case *B* is appropriately produced by *M*, partly because the agent who performs the behavior believes that *B*-ing has tended to bring about *G* or that *B*-ing will tend to bring about *G*.

This account of goal-directed behavior makes it clear that an agent who acts intentionally acts partly because she believes that her action will bring about the end that she aims to realize. This account also remedies the other difficulties with Taylor and Wright's proposal noted above.

If my emended proposal is an adequate account of goal-directed intentional action, then it is not difficult to demonstrate that at least some instances of the pursuit of the bad are goal-directed. Suppose that John desires to punch a hole in the wall and he believes that punching a hole in his wall is bad and not at all good. Suppose that John's desire to punch a hole in his wall leads him to swing his fist, partly because he believes that swinging his fist tends to bring it about that one punches a hole in the wall, and suppose that John swings his fist in exactly the way that he believes will produce a hole in his wall. If John succeeds, then he pursues the bad since John believes that punching a hole in his wall is bad and not at all good and he has the goal of punching a hole in his wall. Thus, given the present account of goal-directed behavior, some instances of the pursuit of the bad are instances of goal-directed behavior.

However, even if my emended account of Taylor and Wright's proposal suffices as an account of goal-directed behavior, it may still fail as an account of intentional action. We have been supposing that all intentional action is subject to teleological explanation and that seems to imply that all intentional action is goal-directed behavior. However,

not all goal-directed behavior is intentional action; the emended account of goal-directed behavior is missing something that is needed for an account of intentional action.

### **Full-Blooded Action, Control, and Conscious Direction**

David Velleman claims that “what makes for [intentional] action is not simply being goal-directed.”<sup>17</sup> Velleman offers the following example to illustrate:

Say, a child accidentally brushes a glass off of the table, and your hand shoots out to catch it. Everything happens so fast that you see your hand catching the glass *before you realize that the glass is falling*. Now suppose, finally, that another child—an older and sassier child—hefts the glass with a smirk and calls, “Here, catch!” You then undertake the same behavior, but as a fully intentional action.<sup>18</sup>

In both cases, it is plausible to suppose that the glass-catcher wants to save the falling glassware. Further, it is plausible to suppose that thrusting out one’s hand tends to save falling glassware, that both glass-catchers believe that thrusting out their hands tends to save falling glassware, and that the belief of each glass-catcher is part of what explains why each thrusts out her hand. So both instances of glass-catching appear to count as instances of goal-directed behavior. Yet Velleman insists that only the second instance of glass-catching behavior, and not the first, is an intentional action. Interestingly, Velleman claims that the behavior of the second glass-catcher, but not the first, is under the agent’s “conscious control” and “conscious direction.”<sup>19</sup> Similarly, Velleman claims that intentional action just is goal-directed behavior that is “executed under conscious control.”<sup>20</sup> Since we are already interested in understanding what it is for an agent to be in control of herself and what she does, it may be useful to determine what the difference

---

<sup>17</sup> Velleman, *ibid.*, p. 190.

<sup>18</sup> Velleman, *ibid.*, p. 189, emphasis added.

<sup>19</sup> Velleman, *ibid.*, pp. 191-2.

<sup>20</sup> Velleman, *ibid.*, p. 191.

between Velleman's glass-catcher is in order to determine what the difference is between goal-directed behavior and intentional action.

To be sure, the cases of the two glass-catchers are perhaps under-described. The first instance of glass-catching might appear to be a reflex and, intuitively, reflexes are not intentional actions; if a bee stings me and I quickly withdraw my arm, it seems plausible to claim that I did not withdraw my arm intentionally. But some instances of reflexive actions might count as intentional actions. An agile baseball player might reflexively catch a line-drive coming at his head. Out of habit, I might intentionally pick up the phone receiver when it rings without thinking about what I am doing. Similarly, a parent with a number of clumsy children might routinely and habitually catch falling glasses, and thus catch the glasses intentionally even if she catches them reflexively. The agile baseball player has been trained to catch line-drives and has done so a number of times. I answer the phone out of habit as does the parent who catches glasses knocked off tables by her clumsy children. In all these cases, it is plausible to suppose that the agents who perform reflexive actions have an intention or standing policy to perform that behavior in those circumstances. It is unclear if the first glass-catcher does intend to catch falling glasses, but if he does have a standing intention to catch falling glasses, we should not be so quick to insist that he does not act intentionally.

It might be the case that Velleman is wrong and that the first instance of glass-catching is an instance of intentional behavior, but it does not follow that it is a *fully* intentional action like the second instance of glass-catching. Velleman distinguishes between "full-blooded actions" which are paradigm instances of intentional action in which "human agency is exercised to its fullest" and actions that are less than full-

blooded in which human agency is exercised only “partially or imperfectly”<sup>21</sup> Full-blooded action, claims Velleman, is human action *par excellence*.<sup>22</sup> Actions that are less than full-blooded include instances in which an agent acts “halfheartedly, or unwittingly, or in some equally defective way.”<sup>23</sup> Consider what else Velleman claims about full-blooded action and the two glass-catchers:

What’s missing from the reflexive case is conscious direction on your part, which is something other than eye-hand coordination. When goal-directed behavior proceeds under this conscious control, it becomes a full-blooded action, rather than a well-coordinated reflex.<sup>24</sup>

Here, Velleman equates conscious direction with conscious control, such that an agent who consciously directs her behavior will also have conscious control of it. Perhaps the behavior of the first glass-catcher counts as intentional action, perhaps it does not, but there is a strong case to be made that the behavior of the first glass-catcher does not count as full-blooded action just because there is a strong case to be made that the first glass-catcher fails to consciously direct his behavior, and thus, following Velleman, he would not be in control of his behavior and would therefore not perform a full-blooded action.

Recall that Velleman’s first-glass catcher thrusts out his arm before he realizes that the glass is falling. If the first-glass catcher does not realize that the glass is falling, he surely cannot direct his behavior in order to save falling the falling glass. Thus, even if the first glass-catcher has an intention to save falling glasses, it does not follow that he directs his behavior in accord with his intention when he intentionally catches the falling

---

<sup>21</sup> Velleman, *ibid.*, p. 189.

<sup>22</sup> Velleman, “What Happens When Someone Acts?”, p. 124.

<sup>23</sup> Velleman, *ibid.*, p. 124.

<sup>24</sup> Velleman, “The Possibility of Practical Reason”, p. 191.

glass. If he does not, then while he may or may not save the falling glass intentionally, he does not perform a full-blooded action insofar as he lacks the requisite control over his behavior that agents exhibit when they perform full-blooded actions.

It will be useful to focus on Velleman's account of full-blooded action for two reasons. First, while I have argued that pursuing the bad can be goal-directed behavior, I have been trying to demonstrate that pursuing the bad can also be intentional action. If it can be shown that at least some instances of pursuing the bad are full-blooded actions, or at least that they more closely resemble full-blooded actions than defective actions, then we have strong grounds for claiming that at least some agents who pursue the bad have the same sort of control and understanding that agents who perform uncontroversial instances of intentional action do. This just is the argumentative strategy that I proposed at the beginning of this chapter. Second, we are attempting to discern what it is for an agent to control herself and what she does and to understand herself and what she does. Velleman offers a plausible account of what it is for an agent to control her action that suggests an account of what it is for an agent to understand herself and what she does in terms of "conscious direction." I propose to consider Velleman's account of control in terms of consciously directing one's action.<sup>25</sup> If it should turn out that agents who pursue the bad can also consciously direct their behavior, then we have a direct argument that agents who pursue the bad are in control of themselves and what they do in response to Raz's challenge.

Consider two claims that Velleman advances with respect to control and conscious direction of one's behavior. First, Velleman claims that in instances of full-blooded

---

<sup>25</sup> Velleman's account of control is not the only account of control that is available in the literature. For a rather different account of control, see John Martin Fisher and Mark Ravizza, *Responsibility and Control* and Fischer, *The Metaphysics of Free Will: An Essay on Control*.

action, the acting agent both forms an intention and deliberately executes that intention.<sup>26</sup> Second, Velleman claims that “practical reasoning is the process by which you exercise conscious control.”<sup>27</sup> These claims are intertwined: part of practical reasoning, as I understand it, includes the deliberate formation and execution of intentions, and if practical reasoning is the process of exercising conscious control over what one does, then deliberately forming and executing intentions will also be part of the process of exercising conscious control. Above, I have suggested that simply having and behaving as a result of an intention is not sufficient for acting intentionally; again, the first glass-catcher might have an intention to save falling glasses but he might not consciously direct his behavior. Velleman claims that “an agent’s desires and beliefs can cause a corresponding intention despite him, and hence without his participation” and that those same desires can cause an agent to execute that intention, again, despite him and without his participation.<sup>28</sup> A reluctant smoker’s desire to smoke might cause him to intend to smoke and to execute his intention to smoke, but even though he smokes intentionally, he does so reluctantly and half-heartedly and unwittingly. His smoking behavior is certainly goal-directed and appears to be intentional action, but it does not count as a case of full-blooded action, nor does it seem to be a case in which he controls and consciously directs his smoking behavior.

We are now in better position to explain what it is to consciously direct one’s behavior: an agent consciously directs her behavior only if she deliberately forms and

---

<sup>26</sup> Velleman, “What Happens When Someone Acts?”, p. 124.

<sup>27</sup> Velleman, “The Possibility of Practical Reason”, p. 198.

<sup>28</sup> Velleman, “What Happens When Someone Acts?”, p. 125. See also Harry Frankfurt, “The Problem of Action”, reprinted in his *The Importance of What We Care About*, p. 72.

executes an intention to act; at the very least, both the first glass-catcher and the reluctant smoker do not deliberately form and execute the intentions that produce and explain their behavior and this is part of the explanation of why their behavior fails to count as full-blooded action. In Chapter 3, I argued that agents can deliberately form intentions to pursue the bad and that agents can deliberately execute those intentions by deciding to act and by deciding to execute their intentions. I shall have more to say about decisions, including decisions to act below. However, this is likely not all there is to Velleman's account of control; again, Velleman claims that practical reasoning is a means by which agents exercise conscious control over their behavior, but there is more to practical reasoning than just forming and executing intentions, especially in instances of full-blooded action. Velleman, again, following Frankfurt, suggests that agents who perform full-blooded actions reflect on possible motives and reasons for acting and adjudicate between those various motives and reasons for acting.<sup>29</sup> Very often, agents engage in deliberation about what to do before they form intentions to act and this deliberation will also be a part of the means by which agents exercise conscious control over their behavior on Velleman's account. We can also stipulate that agents who engage in the appropriate kind of deliberation about reasons and motives prior to forming intentions to act are sufficiently reflective and attentive in their deliberations, that counter-factual interveners are not interfering with them or coercing them, and so forth.

We now have a rather complicated account of what is required for agents to perform full-blooded actions and thus to consciously direct their behavior: agents who act intentionally must deliberately form and execute intentions to act, and prior to forming

---

<sup>29</sup> Velleman, *ibid.*, pp. 138-9. See also Frankfurt, "Freedom of the Will and the Concept of a Person", p. 18.

their intentions, agents must sufficiently and reflectively consider various possible motives and reasons for acting and then adjudicate between those motives and reasons absent coercion or interference.<sup>30</sup> Note that demanding that agents who perform full-blooded actions deliberately form an intention to act after having engaged in conscious deliberation about what to do is just the thing to ensure that an agent understands what she does and why she is doing it. And deliberately executing that intention is just the thing to ensure that an agent who is engaging in goal-directed behavior is in control of what she is doing. Thus, we have a plausible account of what it is for an agent to consciously direct her behavior and a plausible account of what is required for an agent to be in control of herself and what she does and to understand herself and what she does. The question, then, is whether or not agents who pursue the bad can similarly exercise control and understanding by considering and adjudicating between various reasons and motives for acting and whether they can deliberately and consciously form and execute intentions to pursue what they believe to be bad and not at all good.

I suggest that agents can engage in conscious and deliberate reflection and adjudication and deliberation and deliberately and consciously form and execute their intentions while also pursuing the bad. Note that we can construct different cases of pursuing the bad that differ in how conscious and deliberate they are; for example, we can imagine two different Carlas, Carla<sup>1</sup> and Carla<sup>2</sup>, both of whom engage in self-mutilating behavior. Suppose that Carla<sup>1</sup> has a deep-seeded desire to harm herself based upon feelings of a lack of self-worth, one that is engaged when she is made to feel bad about herself. Upon being demeaned yet again by her rather uncaring parents, her desire

---

<sup>30</sup> For sympathetic proposals, see Michael Bratman, "Practical Reasoning and Weakness of Will", *Nous* 13:2 (May, 1979), p. 154, and his remarks on "full-blown actions" and Al Mele, *Springs of Action*, p. 13.

to harm herself is engaged, and she locks herself in her room, procures her favored cutting instrument, and proceeds to mutilate herself. It is only after she has finished her cutting and has begun to bleed that she realizes what she has been doing. Contrast Carla<sup>1</sup> with Carla<sup>2</sup> who has a similar deep-seeded desire to harm herself. Upon being demeaned yet again by her rather uncaring parents, her desire to harm herself is engaged. Unlike Carla<sup>1</sup>, Carla<sup>2</sup> reflects upon her desire to harm herself and decides to act on her desire. Carla<sup>2</sup> then consciously and deliberately forms an intention to mutilate herself. She then locks herself in her room and consciously and deliberately executes her intention and harms herself. It seems that Carla<sup>2</sup> performs a full-blooded action and thus counts as being in control of herself and what she does and understands herself and what she does.

Of course, the Panglossian is not going to accept all of this. In particular, I suspect the Panglossian will claim that I have simply helped myself to the claim that an agent who pursues the bad can consciously and deliberately reflect on a desire to harm herself, decide to act on that desire, and then form and execute an intention to pursue the bad. Further, claims the Panglossian, while I have suggested in previous chapters that various moods and emotions can explain why agents desire and intend the bad and why agents regard the believed badness of an action to be a reason to act, I have not explained in any detail how deliberation that produces those results is supposed to proceed; recall the objection from the black box noted at the end of the previous chapter. The Panglossian might agree that practical reasoning precedes full-blooded action and agree with Velleman that practical reasoning is the means by which agents exercise conscious control over what they do, but deny that I have adequately explained how agents could engage in practical reasoning and then pursue the bad. Thus, the Panglossian would deny

that I have earned the claim that some instances of the pursuit of the bad are instances of full-blooded action. Further, if instances of full-blooded action and instances of the pursuit of the bad are sufficiently disanalogous, then the argument from analogy in response to Raz's challenge is undermined.

I am suggesting that some instances of the pursuit of the bad do proceed from practical reasoning of the kind discussed above. I concede that I have perhaps not said enough about how practical reasoning is supposed to produce instances of the pursuit of the bad. In the following sections of this chapter, I attempt to provide a more detailed account of how agents might engage in practical reasoning yet still pursue the bad. If it can be shown that agents can engage in practical reasoning just like agents who perform instances of full-blooded actions yet still pursue the bad, then given I have already shown that agents who pursue the bad can have goals and that their behavior is goal-directed, we ought to conclude that at least some instances of the pursuit of the bad are sufficiently analogous to full-blooded actions and are therefore intentional. We will then have grounds for rejecting Raz's challenge. We will then be in position to insist the truth of the *intentionbad* thesis.

### **Practical Reasoning**

Velleman's account of full-blooded action implies that agents who perform full-blooded action engage in practical reasoning about what they will do; in particular, they consider and adjudicate between various reasons for acting. In this section and in the sections that follow, I want to consider various ways agents might consider and adjudicate between various reasons for acting that might lead them to pursue the bad.

Here are some plausible thoughts regarding practical reasoning. Practical reasoning involves, among other things, thinking about what considerations are relevant

to deciding what to do, weighing those considerations and balancing them against conflicting and competing considerations and so forth. Failing to treat some consideration as a reason is to fail to treat that consideration as relevant, or to treat it as irrelevant, to determining what to do. Treating some consideration as a reason involves more than treating that consideration *as if* it were a reason; I might imagine what it would be like to be a very different person who is moved by reasons that do not at all move me. In that case, I might treat some consideration as if it were a reason but not treat it as a reason. Treating some consideration as a reason thus appears to be treating that consideration as relevant to determining what to do. Thus, insofar as we are interested in understanding what it is to engage in practical reasoning and in determining if agents who pursue the good can do so on the basis of practical reasoning, we ought to take an interest in what it is to treat some consideration as a reason and as relevant to determining what to do.

Other philosophers have endorsed accounts of what it is to treat some consideration as a reason that are at least consistent with the above proposal. Michael Bratman suggests that treating a desire as a reason involves treating that desire as justifying to some extent the performance of some relevant action.<sup>31</sup> Thomas Scanlon claims that taking a consideration to be a reason for acting is to take that consideration as counting in favor of that action.<sup>32</sup> Steven Darwall suggests that when an agent counts some consideration as her reason, that reason “engage[s] considerations that seemed

---

<sup>31</sup> Bratman, “Identification, Decision, and Treating as a Reason” reprinted in his *Faces of Intention*, p. 198.

<sup>32</sup> Scanlon, *What We Owe To Each Other*, p. 17.

recommendatory in the agent's own view."<sup>33</sup> Treating some consideration as a reason thus appears to involve a certain kind of psychological attitude; in particular, treating some consideration as a reason appears to involve a belief that something like a justificatory relation holds between that consideration and some relevant action.

An agent may come to treat some consideration as a reason on the basis of some decision or deliberation, but not necessarily. An agent may make a decision to treat some consideration as a reason, where absent that decision, the agent would not have included that consideration in her deliberation. I shall refer to this first kind of case of treating as a reason as *active*. An agent might also treat a consideration as a reason for acting, but not on the basis of any prior psychological activity. At no point must an agent *decide*, for example, to treat some consideration as a reason, even if she does treat that consideration as a reason. She may simply be sensitive or responsive to reasons of that kind or disposed to treat considerations of that kind as a reason.<sup>34</sup> Hoping not to confuse the issue, I shall refer to this kind of treating as a reason as *passive*. In cases of passive treating as a reason, the agent is still reflective to a degree, just as agents are who come to believe a proposition without drawing any inferences.

We are supposing that practical reasoning precedes full-blooded intentional actions, that agents who perform full-blooded actions have understanding and control, and that engaging practical reasoning is a means of exercising control. Since I have suggested that part of practical reasoning involves both actively and passively treating considerations as reasons and therefore as relevant to determining what to do, it is worthwhile to determine

---

<sup>33</sup> Darwall, *Impartial Reason*, p. 32.

<sup>34</sup> For discussion of reasons-responsiveness, see John Martin Fischer and Mark Ravizza, *Responsibility and Control*.

whether or not agents who pursue the bad can actively or passively treat as reasons considerations that favor or recommend or prescribe pursuing the bad. If so, we will be that much closer to demonstrating that agents who pursue the bad can perform full-blooded actions, that they have control and understanding over themselves and what they do, and that the *intentionbad* thesis is true.

### **Passive Cases of Treating as a Reason and Pursuing the Bad**

In earlier chapters, I have argued that agents might have perverse evaluative beliefs and might accept perverse principles concerning what justifies an action and that these evaluative beliefs and principles might be perverse by an agent's own lights. If this is right, then perhaps agents can passively treat, for example, the fact that an action would cause her to be harmed or to suffer needlessly as a reason to act if she already accepts perverse principles with respect to what considerations should be treated as reasons.

Most of us are disposed to passively fail to treat, or to passively treat as irrelevant, considerations that show an action to be bad and not at all good. Typically, I do not need to deliberate or decide to fail to treat the fact that an action would cause me to be harmed or to suffer needlessly as a reason; I simply don't treat that consideration as a reason to act. We should already be willing to accept the more general thesis that agents can be disposed to passively treat some considerations, but not others as reasons. We attempt to habituate our children to share, to avoid being cruel, and so forth in the hopes that these habits are internalized and continue to be practiced in adult life. We don't necessarily teach children to decide that the fact that an action is honest or kind is a reason to perform that action; we might only try to dispose them to passively treat those considerations as reasons. Finally, as I have suggested in earlier chapters, an agent's moods or emotions might dispose her to treat some considerations as reasons that, absent that mood or

emotion, she would not have treated as a reason. Love makes a man do crazy things, perhaps, but if I love you, I might also be disposed to sacrifice for your well-being, to help you where I can and so forth.

While we can be disposed to treat, for example, the fact that another person is suffering as a reason, we know all too well that people can fail to be disposed to treat facts of that kind as reasons. Psychopaths seem to be both unwilling and unable to treat the fact that another person is suffering as a reason to act. It is not simply that psychopaths lack a desire to help other people or to tend to their feelings, but that even if a psychopath had such a desire he would be unsure about or oblivious to taking that desire as a reason to do anything that did not also serve his own interests.<sup>35</sup> There is some evidence that some psychopaths fail to treat the suffering of others as reasons because of damage to particular regions of the brain near the frontal lobes.<sup>36</sup> Thus, the psychopath can be disposed to fail to treat some consideration as a reason absent any decision or deliberation.

In the same way I might be disposed to passively treat some consideration as a reason or to fail to be disposed to passively treat some consideration as a reason, I might be disposed to passively treat something I believe to be a bad-making feature of an action as a reason for me to act. If I am raised by sadists, I might be disposed to treat the suffering of others as a reason to intervene to end their suffering, and I thus treat their suffering or my desire to help as a reason for me to act, even if I also believe that their suffering is bad and not at all good. A woman who is raised in an abusive household

---

<sup>35</sup> See Hare, *Without Conscience* for authoritative discussions of this point.

<sup>36</sup> See, for example, Antonio R. Damasio, *Descartes' Error*.

might be disposed to seek relationships with abusive men, even if she sees nothing good about being abused. Psychologist Lorna Smith Benjamin discusses what she calls “attachment behavior” and, according to Claudia Card, “attachment to an important other person can elucidate behavior that otherwise appears simply perverse, irrational, perhaps even diabolical.”<sup>37</sup> Carla might be disposed to treat the fact that mutilating herself will cause her to be harmed as a reason because she has been raised in an abusive and neglectful household and because she has been trained to think of herself as worthless. Depressed agents might fail to be attracted to actions that would promote their happiness or well-being and frustrated agents might be disposed to perform actions they believe they will fail at.

To be clear, I am arguing for two distinct theses here. First, I suggest that agents can passively treat considerations as reasons, insofar as agents can be disposed to treat some considerations as reasons absent any decision or deliberation to treat those considerations as reasons. Second, I suggest that dispositions to passively treat certain considerations as reasons can arise absent any decision or deliberation. Keeping these theses distinct is important because it allows us to tend to a Panglossian objection that has been lingering.

I have been claiming that if an agent pursues the bad, we can provide a rationalizing explanation of that action by citing the fact that the believed badness of the action appealed to the acting agent. In the current language, I claim that an agent who pursues the bad might be disposed to treat the believed bad-making feature of her action as a reason for her to perform the action. The Panglossian objects that an agent cannot

---

<sup>37</sup> Claudia Card, *The Atrocity Paradigm*, p. 90; Laura Smith Benjamin, *Interpersonal Reconstructive Therapy*.

treat the believed badness of an action as a reason to act. I respond that an agent can treat the believed badness of an action as a reason to act, if she is, for example, disposed to treat that particular bad-making feature as a reason to act; if she is so disposed, then she might passively treat the bad-making feature of her action as a reason for her to act.

It is here, I think, that a Panglossian objection emerges. The Panglossian objects that we are interested in providing rationalizing explanations of actions. To appeal to factors like an agent's upbringing or social conditioning or biology or psychology might explain why she treats some bad-making consideration as a reason, but *that* kind of explanation is the wrong kind of explanation; that kind of explanation is a merely causal explanation, but merely causal explanations are not rationalizations.<sup>38</sup> But this objection is misguided. Rationalizing explanations explain why an agent acted as she did. But not everything that is part of a rationalizing explanation can itself be rationalized; as I have suggested in this section, agents can be passively disposed to treat a consideration as a reason, and there is no guarantee that there will be a rationalizing explanation of why an agent is disposed to treat that consideration as a reason, although there will be non-rationalizing explanations of why she treats that consideration as a reason. The Panglossian cannot object that an agent who passively treats some bad-making consideration as a reason is unable to understand herself and what she does or that she is not in control of herself and what she does. For one thing, an agent who learns about her habituation and upbringing and biology and psychology might come to understand why she does what she does. For another, agents often passively treat good-making

---

<sup>38</sup> I say that *merely* causal explanations are not rationalizations to make clear that I do not accept the thesis that rationalizations cannot also be causal explanations. Davidson himself, of course, defends the thesis that rationalizing explanation is a variety of causal explanation in his classic "Actions, Reasons, and Causes."

considerations as reasons for them to act. It is not plausible that they must fail to understand themselves and what they do and must fail to be in control of themselves and what they do. If we claim that we lack understanding and control whenever we cannot provide a rationalizing explanation of why we are disposed to treat some consideration as a reason, then I submit we almost always lack understanding and control. And that seems to be a *reductio* of the present objection.

In this section, I have discussed cases of passively treating a consideration as a reason. I have argued that agents might passively treat bad-making considerations as reasons as a result of perverse dispositions and evaluative tendencies that are not subject to rationalizing explanations. Since agents who perform full-blooded actions sometimes passively treat considerations as reasons for them to act, there is no disanalogy here between paradigmatic instances of action and instances of pursuing the bad.

### **Active Treating as a Reason and Pursuing the Bad**

In the previous section, I argued that agents can be disposed, for various reasons, to passively treat bad-making features of an action as reasons to act. Practical reasoning must begin somewhere, and at least sometimes, practical reason begins with agents passively treating some considerations as reasons. But practical reasoning also involves making decisions about which considerations to take into account when deciding what to do and it involves investigating and discerning what reasons there are to act as well. Scanlon discusses an agent who attempts to decide whether or not he will play to win while believing there are reasons for and against playing to win; the agent may decide, for example, to treat the fact that playing to win will be more enjoyable as a reason but not to treat other considerations believed relevant, such as the potential disappointment of

one's opponent, as reasons to act.<sup>39</sup> We might deliberately ignore or disregard some considerations we believe are relevant to avoid the costs of including those considerations in our reasoning. I might disregard the possibility that my future beloved will be at the bar I am considering going to if I believe that if I do start thinking about her, I will be unable to think about anything else. I might look at only part of a menu because I know that I will not make up my mind in time if I have too many options for my entrée. I might decide to disregard some consideration I believe is relevant to my deliberation on moral grounds as well.

As I suggested above, making a decision about what considerations to treat as reasons is a way of actively treating some consideration as a reason. It is not the only way, however. I might actively treat some consideration as a reason by engaging in something more akin to theoretical deliberations about whether some consideration really is a reason and thereby coming to believe that it is, or is not, a reason for me to act. I might wonder about whether or not some desire of mine really is a reason to act and engage in thoughtful patterns of reasoning, not just to decide what my reasons are, but to discern what my reasons are; I might reason about what a fully informed and benevolent advisor would advise me to do or I might consider whether my initial intuitions are correct by attempting to reach a reflective equilibrium. So practical reasoning involves not just weighing considerations for and against action; practical reasoning also involves decisions about what to count as relevant in determining what to do and it involves deliberation that leads to new beliefs about what to do.

---

<sup>39</sup> Scanlon, *ibid.*, pp. 51-2.

In the same way that an agent's dispositions and evaluative tendencies and emotions can explain why she passively treats some consideration as a reason, an agent's dispositions and evaluative tendencies and emotions can explain why she actively treats some consideration as a reason. I want to examine different kinds of decisions to treat a consideration as a reason and then argue we can plausibly explain decisions to treat bad-making properties of an action as a reason to act. What I have to say about decisions to treat considerations as reasons can also largely be applied, I think, to the ways in which we acquire beliefs about our reasons and I suggest that we can also plausibly explain cases in which agents come to believe that bad-making properties of an action are reasons to act.

We can identify different sorts of decisions to treat a consideration as a reason. A decision can be made on a principled basis, based upon evaluative criteria, or a decision can be made that is unprincipled and not based upon any evaluative criteria. An agent might make a decision based upon what she takes to be good grounds. Alternatively, an agent might make a capricious or erratic decision, knowing that there are no grounds for deciding in that way rather than some other way, with real indifference to what is at stake.<sup>40</sup> I shall refer to decisions of the first kind, decisions that are based upon evaluative criteria believed by the agent to be relevant, as *principled* decisions, and I shall refer to decisions of the second kind, decisions that are believed by the agent not to be supported by some further evaluative criteria, as *existential* decisions. An agent might make a principled decision to treat, for example, her desire to help a friend who is in need

---

<sup>40</sup> Frankfurt, *ibid.*, p. 19, fn. 6. However, in other places, Frankfurt suggests that what I am calling existential decisions are not really decisions but are rather *choices*; see Frankfurt, "Identification and Wholeheartedness", p. 172 and Bratman, "Identification, Decision, and Treating as a Reason", pp. 191-2.

as a reason. An agent might make an existential decision to treat, for example, her desire for tea rather than her desire for coffee as a reason while failing to believe there is any justification for preferring tea to coffee. Perhaps she simply decides to be the sort of person who drinks tea rather than coffee, even though she does not believe there is anything that actually justifies doing so.

Decisions can have another feature that is relevant. For example, an addict might grudgingly or half-heartedly decide to treat her desire for her favored drug as a reason simply because it takes too much effort to resist the drug and the struggle is too painful; in such a case, it is not clear the agent endorses or identifies with her desire in any strong sense, even though she treats her addictive desire as a reason for her to act.<sup>41</sup> It is consistent with making a grudging decision that the agent has real reservations about making the decision; nonetheless, the agent is treating his desire as a reason, insofar as he is willing to act upon the desire and form intentions on that basis. I shall refer to decisions of this sort as *grudging* decisions.

With these distinctions in tow, I begin by examining principled decisions; I shall argue it is possible for an agent to actively make a principled decision to treat the believed badness of an action as a reason. In the previous section, I argued that agents can be disposed to passively treat a bad-making feature of an action as a reason to act. It is equally possible, I suggest for an agent to make a principled decision to treat a bad-making feature of an action as a reason to act.

Some agents acquire or inherit perverse beliefs or principles about what justifies an action; such an agent can perhaps make a principled decision to treat a bad-making

---

<sup>41</sup> Bratman notes this possibility in his *ibid.*, p. 199 and 205.

feature of an action as a reason. Consider Carla again. Carla recognizes there are reasons to perform any number of actions that do not constitute self-mutilating behavior. Carla believes there is a reason for her to confront her parents and tell them how terribly they act towards her; she believes there is a reason for her to go for a walk and to blow off some steam and relax; she believes she has a reason to call her close friends so they can console her. In all these cases, Carla believes she has a reason to do something that she knows will result in her feeling better and in her present state improving. But because of her emotional state and because she has acquired the dispositions and habits of a self-mutilator, Carla is disposed to disregard those considerations and not to treat them as reason. Instead, she is disposed to treat the bad-making features of self-mutilating actions as reasons, and when she is sufficiently distressed or upset, she is disposed to only see and actively treat the bad-making features of self-mutilating actions as reasons.

Similarly, an agent who attempts to arrive at some reflective equilibrium about what is to be done might arrive at perverse conclusions about what to do if she begins the reflective process with perverse intuitions or principles about what to do. Carla may have the intuition that she should perform actions that will make her feel better but also have the intuition that she is to be made to suffer. Perhaps to avoid cognitive dissonance and arrive at a reflective equilibrium, she abandons one of those intuitions; there is no guarantee that she will abandon her intuition that she is to be made to suffer, rather than retain that intuition. Garbage in, garbage out. If Carla arrives at a reflective equilibrium and believes that her intuition that she is to be made to suffer suggests there is a reason for her to act, then Carla will actively be treating her intuition that she is to be made to suffer as a reason to act.

### **Principled Decisions to Pursue the Bad**

Can a decision to treat the bad-making features of an action as reason to act be principled? Principled decision to treat a consideration as a reason must be made on the basis of some evaluative criteria, but it is unclear just what sort of evaluative criteria could be appealed to that would favor or recommend or prescribe performing an action believed to be bad and not at all good. But it does seem possible for agent to inherit or acquire perverse evaluative criteria that they are disposed to appeal to when they are actively treating some consideration as a reason. Rachel Cohon notes that it is possible that agents have “imperfect” standards of practical rationality, such that the agent will “see reasons to act where there are none.”<sup>42</sup> Carla, for example, might have imperfect standards of practical rationality such that she comes to believe that the fact that a self-mutilating action would harm her is a reason for her to act, even though there is really no such reason. Admittedly, the evaluative criteria Carla appeals to that implies she the fact that her self-mutilating action would harm her is a reason for her to act might be perverse, and perverse by her own lights. But if Carla already accepts that evaluative criteria and is disposed to treat it as relevant when she decides what considerations to count as reasons or when she tries to discern what reasons she has, she can actively treat a bad-making feature of an action as a reason for her to act. Again, garbage in, garbage out.

Various philosophers have suggested or left it open that agents can make principled decisions based upon evaluative criteria, even if the agents believe that the evaluative criteria is perverse. Michael Bratman has focused on the role of policies in practical reasoning and deliberation; a policy, for Bratman, is a suitably general future-directed

---

<sup>42</sup> Cohon, “Internalism About Reasons for Action”, *Pacific Philosophical Quarterly* 74:4 (Dec., 1993), p. 274.

intention that has the distinctive function of organizing and coordinating and directing an agent's action and activity over time. Policies amount to commitments to perform a certain kind of action on certain potentially recurrent occasions.<sup>43</sup> Bratman stops short of claiming that for some suitable intention to count as a policy, the agent must identify or endorse or be satisfied *with that policy*; Bratman explicitly considers the possibility that an agent could be estranged from her self-governing policies.<sup>44</sup> This strikes me as plausible. I may have a policy to support leftist causes, but perhaps my youthful idealism has withered over time and I am now estranged from my leftist policy. Still, if I am disposed to appeal to my leftist policy, a policy that I am now estranged from, when I attempt to determine what reasons I have to act, then my decision to treat my desire to promote leftist causes as a reason to act can be a principled decision.

Gary Watson has suggested that an agent's values, when combined with factual beliefs and probability estimates, yield judgments about what is to be done.<sup>45</sup> Interestingly, Watson later claims that agents can be estranged from their values, such that it is a mistake to conflate "valuing with judging good" and allows that we can be alienated from our values: what one values "in a particular case may not be sanctioned by

---

<sup>43</sup> Bratman has developed this account in various places, but see especially "Reflection, Planning, and Temporally Extended Agency", p. 41-8; "Intention and Personal Policies", pp. 455-61; and "Identification, Decision, and Treating as a Reason" and his *Intention, Plans, and Practical Reason*.

<sup>44</sup> Bratman takes this concern seriously enough to modify an initial proposal that identification with a desire is a matter of whether the desire coheres with self-governing policies an agent is satisfied with; Bratman, "Reflection, Planning, and Temporally Extended Agency", pp. 48-9. Note that Bratman also does not claim that the fact that an agent decided to treat, for example, some desire as a reason implies the agent identifies with that desire. Since Bratman's account of identification with a desire depends upon making a decision to treat that desire as a reason, it cannot be a necessary condition for deciding to treat a desire as a reason that an agent *already* identifies with that desire; that would reverse the order of explanation. See Bratman, "Identification, Decision, and Treating as a Reason", p. 197 and pp. 200-1.

<sup>45</sup> Watson, "Free Agency", *The Journal of Philosophy* 72:8 (April, 1975), p. 215.

a more general evaluational standpoint that one would be prepared to accept.”<sup>46</sup>

Velleman suggests that an agent might be alienated from materialism or sense of sin.<sup>47</sup>

Similarly, my values might be rooted in a rigorous Catholic upbringing that I now think is ill-advised and silly and horribly misguided. Still, if Watson’s initial thought that values contribute to judgments about what is to be done is correct, then perhaps values that an agent is alienated from can be appealed to as evaluative criteria when she is deciding to treat some consideration as a reason; her decision would then be a principled decision.

Allan Gibbard contrasts *accepting* a norm with *being in the grip* of a norm; an agent is in the grip of some norm,  $N_1$ , if she does not believe that  $N_1$  overrides some other norm,  $N_2$ , but still governs her behavior in accord with  $N_1$  rather than  $N_2$ .<sup>48</sup> Even when an agent is only in the grip of a norm, however, the agent has a motivational tendency to act as prescribed by that norm.<sup>49</sup> For example, the subjects of the infamous Millgram experiments followed norms prescribing cooperativeness and on that basis they complied with orders to administer increasingly painful- or what they thought were increasingly painful-electric shocks.<sup>50</sup> But at least some of the subjects protested vigorously and vehemently. It strikes me as plausible to suppose that at least some of the subjects in the Millgram experiments genuinely believed that torturing victims is bad and not at all good, but were in the grip of norms of cooperativeness and complied with their orders to shock.

---

<sup>46</sup> Watson, “Free Action and Free Will”, *Mind* 96 (April, 1987), p. 150.

<sup>47</sup> Velleman, “What Happens When Someone Acts?”, p. 134. Velleman attributes the examples to Elizabeth Anderson.

<sup>48</sup> Gibbard, *Wise Choices, Apt Feelings*, p. 60.

<sup>49</sup> Gibbard, *ibid.*, p. 60 and pp. 68-71.

<sup>50</sup> See Gibbard, *ibid.*, pp. 58-60 for discussion.

*Their* reasons for acting will make reference norms of cooperativeness, and thus they might use norms they are in the grip of when deciding what to do; their decisions would thus be principled decisions. Agents who are in the grip of a norm may not think that performing the relevant action is good or that it is good to be in the grip of that norm; again, agents who are in the grip of a norm do not necessarily accept that norm.

I have suggested that insofar as agents can appeal to policies and values and norms that are perverse by their own lights as evaluative criteria, a decision to treat some consideration as a reason will be a principled decision. If so, there is no disanalogy here between instances of full-blooded action and at least some instances of the pursuit of the bad. Perhaps, however, there is no room for the thesis that agents who pursue the bad make principled decisions to treat bad-making features of their actions as reasons. Even if agents cannot make principled decisions to treat some bad-making feature of an action as a reason, it does not follow that agents cannot make decisions to pursue the bad. Note that even if Carla's decision to treat the fact that her action instantiates some bad-making property as a reason is not a principled decision, it still might be an existential decision. If existential decisions are possible, then it seems possible for Carla to make a decision to treat the fact that her action will harm herself as a reason to perform it and her decision to treat that fact as a reason will be a case of actively treating that fact as a reason.

A decision to treat a bad-making feature of an action as a reason to act may or may not be a grudging decision. It is possible, I am supposing, for an agent to wholeheartedly decide to focus only upon the bad-making properties of the actions she is considering performing and to have no qualms or reservations. But it is also possible for an agent to make such a decision only grudgingly. Even if Carla only grudgingly decides to treat the

fact that her self-mutilating action will harm her as a reason to act, she still actively treats that consideration as a reason.

In this section, I have argued that agents can make decisions to treat a bad-making feature of an action as a reason to act. If so, then agents can actively treat a bad-making feature of an action as a reason to act. I pause to note that agents who actively treat some consideration as a reason do appear to understand themselves and what they do and do appear to be in control of what they do. An agent who decides to  $\phi$  because of some reason  $R$  perhaps understands that she is the sort of person who acts in that sort of way for that sort of reason. An agent who decides to  $\phi$  because she actively treats some consideration  $C$  as a reason  $R$  takes charge of what reasons she will act for, in some interesting sense.

It is time, then, to return to Raz's challenge. In previous sections, I have already suggested how the various arguments will go, but in the penultimate section of this chapter, I want to make the arguments explicit. I provide the relevant arguments below.

### **Resisting Raz's Challenge**

Again, Raz challenges the proponent of the *intentionbad* thesis to explain how agents who pursue the bad understand themselves and what they do and be in control of themselves and what they do. I promised to argue as follows:

- 1) Agents who act intentionally understand themselves and what they do and are in control of themselves and what they do in virtue of meeting some set of requirements,  $N$ , for having understanding and control.
- 2) At least some agents who pursue the bad meet the requirements,  $N$ , for having understanding and control.
- 3) Therefore, at least some agents who pursue the bad act intentionally and thus pursue the bad.
- 4) Therefore, the *intentionbad* thesis is true.

In earlier sections of this chapter, I have identified some of the requirements included in *N* that suggest that agents who act intentionally understand themselves and what they do and are in control of themselves and what they do. Agents who act intentionally have goals and purposes and ends, and their actions are subject to teleological explanations in terms of those goals and purposes and ends. Intentional action is goal-directed behavior insofar as agents who act intentionally act because of and for the sake of realizing their goals and purposes and ends. Agents who act intentionally often engage in practical reasoning and acting as a result of practical reasoning is a means of exhibiting control by considering and adjudicating between various reasons for acting. Agents can consider and adjudicate between various reasons for acting by forming and executing intentions. Further, practical reasoning involves treating some considerations rather than others as reasons for acting. Agents can passively or actively treat some considerations rather than others as reasons and agents who actively treat some considerations rather than others as reasons make decisions to treat those considerations as reasons. At least some of the decisions to treat considerations as reasons are principled decisions based upon evaluative criteria believed to be relevant to determining what to do. I find it intuitive that these are all ways of understanding oneself and one's actions or of exhibiting control over oneself and what one does.

In defense of 2), I have argued that agents who pursue the bad can also exhibit understanding and control in precisely the same ways that agents who perform uncontroversial instances of intentional action exhibit understanding and control. Agents who pursue the bad can have goals and purposes and ends. Their behavior counts as goal-directed behavior when they act because of and for the sake of realizing those goals

and purposes and ends, and thus, their behavior is subject to teleological explanation. Agents who pursue the bad can act on the basis of practical reasoning and can consider and adjudicate between various reasons for acting. Agents who pursue the bad can form and execute intentions to pursue the bad. They can both actively and passively treat bad-making features of their actions as reasons for pursuing the bad. Agents who pursue the bad might be disposed to passively treat bad-making features of their actions as reasons. They also might actively treat bad-making features of their actions as reasons by deciding to treat those bad-making features as reasons. Their decisions might only be existential or grudging decisions, but there is a case to be made that agents who pursue the bad can make principled decisions to treat bad-making features of their actions as reasons. If all this is right, or if very much of it is right, then 2) of my argument is true. But if 1) and 2) are true, then we have good grounds for accepting 3) and 4), and thus, for accepting the *intentionbad* thesis.

Of course, the Panglossian might accept 1) but offer a different account of understanding and control and claim that *N* includes rather different requirements than those I have proposed. I confess I do not know what the account would look like or how it would not avoid my arguments in this chapter. Surely understanding what one does is, at least partly, a matter of understanding the reasons for which one acts. Surely understanding oneself is, at least partly, a matter of knowing the kinds of reasons for which one acts. Surely having control over what one does is a matter, at least partly, having some say with respect to the reasons for which one acts. Similarly, having control over oneself is, at least partly a matter of having some control of what one treats as reasons. Admittedly, we may very often lack significant control over what we treat as

reasons, but in that respect, agents who pursue the bad are no worse off than agents who perform paradigmatic instances of intentional action.

I take it, therefore, that I have responded to Raz's challenge. I also take it that I have supplied sufficient positive arguments that the *intentionbad* thesis is true. Note also that some of my arguments in this chapter suggest that the *reasonbad* thesis is true; for example, if agents can treat the believed badness of an action as a reason for acting, it seems plausible to suppose that the believed badness of an action could be an agent's reason for acting. And that just is the *reasonbad* thesis.

It is odd to be in agreement with the man from the underground. But he is conceptually possible as are the actions he imagines. Our best accounts of moral psychology ought to permit these possibilities.

### **Conclusion**

I have argued for and defended a trio of theses in this and earlier chapters: I have defended the *desirebad* thesis that agents can desire what they believe to be bad and not at all good, the *intentionbad* thesis that agents can intend and intentionally pursue what they believe to be bad and not at all good, and the *reasonbad* thesis that the believed badness of an action can be an agent's reason for acting. If my three theses are correct, then so much the worse for standard accounts.

## LIST OF REFERENCES

- Anscombe, Elizabeth (2000), *Intention* (Cambridge: Harvard University Press), 2<sup>nd</sup> edition.
- Aquinas, Thomas (1993), *Selected Philosophical Writings* (New York: Oxford University Press), translated by Timothy McDermott.
- Aristotle (1995), *Aristotle: Selections* (Indianapolis: Hackett Publishing), translated by Terence Irwin and Gail Fine.
- \_\_\_\_\_ (1999), *The Nicomachean Ethics* (Indianapolis: Hackett Publishing), 2<sup>nd</sup> edition, translated by Terence Irwin.
- Arpaly, Nomy (2000), "On Acting Rationally Against One's Best Judgment", *Ethics* 110, pp. 488-513.
- \_\_\_\_\_ (2003), *Unprincipled Virtue: An Inquiry Into Moral Agency* (Oxford: Oxford University Press).
- Audi, Robert (1973), "Intending", *The Journal of Philosophy*, Vol. 70, No. 13, pp. 387-403.
- \_\_\_\_\_ (1979), "Weakness of Will and Practical Judgment", *Nous* Vol. 13, pp. 173-96.
- \_\_\_\_\_ (1986a), "Acting For Reasons", *The Philosophical Review*, Vol. 95, No. 4, pp. 511-46. Reprinted in Mele (1997), pp. 75-105.
- \_\_\_\_\_ (1986b), "Intending, Intentional Action, and Desire", in Marks (1986), pp. 17-38.
- \_\_\_\_\_ (1990), "Weakness of Will and Rational Action", *Australasian Journal of Philosophy*, Vol. 68, No. 3, pp. 410-30.
- Augustine (2001), *The Confessions* (New York: Everyman Publishing), translated by Philip Burton.
- Austin, J. L. (1954), "How to Talk—Some Simple Ways", *Proceedings of the Aristotelian Society*, 53, pp. 227-46.

- Baer, Lee (2002), *The Imp of the Mind: Exploring the Silent Epidemic of Obsessive Bad Thoughts* (New York: Penguin Publishing).
- Baier, Kurt (1958), *The Moral Point of View: A Rational Basis of Ethics* (Ithaca: Cornell University Press).
- Benjamin, Laura Smith (2003), *Interpersonal Reconstructive Therapy: Promoting Change in Nonresponders* (New York: The Guilford Press).
- Benn, Stanley I. (1985), "Wickedness", Vol. 95, No. 4, pp. 795-810. Reprinted in Deigh (1992), pp. 191-206.
- Bishop, John (1989), *Natural Agency* (Cambridge: Cambridge University Press).
- Bittner, Rüdiger (2001), *Doing Things for Reasons* (New York: Oxford University Press).
- Brand, Myles (1984), *Intending and Acting* (Cambridge: The MIT Press).
- Bratman, Michael E. (1979), "Practical Reasoning and Weakness of the Will", *Nous*, Vol. 13, pp. 153-72.
- \_\_\_\_\_ (1984), "Two Faces of Intention", *The Philosophical Review*, Vol. 93, pp. 375-405. Reprinted in Mele (1997), pp. 178-203.
- \_\_\_\_\_ (1987), *Intention, Plans, and Practical Reason* (Cambridge: Cambridge University Press).
- \_\_\_\_\_ (1989), "Intention and Personal Policies", *Philosophical Perspectives*, Vol. 3, pp. 443-69.
- \_\_\_\_\_ (1996), "Identification, Decision, and Treating as a Reason", *Philosophical Topics*, Vol. 24, pp. 1-18.
- \_\_\_\_\_ (1999), *Faces of Intention: Selected Essays on Intention and Agency* (Cambridge: Cambridge University Press).
- \_\_\_\_\_ (2000), "Reflection, Planning, and Temporally Extended Agency", *The Philosophical Review*, Vol. 109, No. 1, pp. 35-61.
- Bond, Edward Jarvis (1983), *Reason and Value* (New York: Cambridge University Press).
- Buss, Sarah and Overton, Lee (2002), *Contours of Agency* (Cambridge: MIT Press).
- Card, Claudia (2002), *The Atrocity Paradigm: A Theory of Evil* (Cambridge: Oxford University Press).

- Cavanaugh, Robert (2002), "Self-Mutilation as a Manifestation of Sexual Abuse in Adolescent Girls", *Journal of Pediatric and Adolescent Gynecology*, Vol. 15, No. 97, pp. 97-100.
- Chang, Ruth (2004), "Can Desires Provide Reasons for Action?", in Wallace (2004), pp. 56-90.
- Cohon, Rachel (1993), "Internalism About Reasons for Action", *Pacific Philosophical Quarterly*, Vol. 74, No. 4, p. 274.
- Copp, David and Sobel, David (2002), "Desires, Motives, and Reasons: Scanlon's Rationalistic Moral Psychology", *Social Theory and Practice*, Vol. 28, No. 2, pp. 243-76.
- Damasio, Antonio R., *Descartes' Error: Emotion, Reason, and the Human Brain* (1994: Putnam Berkeley Group).
- Dancy, Jonathan (2000), *Practical Reality* (Oxford: Oxford University Press).
- Darwall, Stephen L. (1983), *Impartial Reason* (Ithaca: Cornell University Press).
- Davidson, Donald (1963), "Actions, Reasons, and Causes", *Journal of Philosophy*, Vol. 60, pp. 685-700. Reprinted in Davidson (1980), pp. 3-19.
- \_\_\_\_\_ (1970), "How is Weakness of the Will Possible?", in Davidson (1980), pp. 21-42.
- \_\_\_\_\_ (1973), "Freedom to Act", in *Essays on Freedom of Action* (London: Routledge and Kegan Paul). Reprinted in Davidson (1980), pp. 63-81.
- \_\_\_\_\_ (1978), "Intending", in *Philosophy of History and Action*, Yovel and Reidel, eds. (1978). Reprinted in Davidson (1980), pp. 83-102.
- \_\_\_\_\_ (1975), "Hempel on Explaining Action", *Erkenntnis*, Vol. 10, pp. 239-53. Reprinted in Davidson (1980), pp. 261-75.
- \_\_\_\_\_ (1980), *Essays on Actions and Events* (Oxford: Oxford University Press).
- \_\_\_\_\_ (1985), "Replies to Essays I-IX", in Vermazen and Hintikka (1985), pp. 195-229.
- Davis, Wayne A. (1984), "A Causal Theory of Intending", *American Philosophical Quarterly*, Vol. 21, pp. 43-54. Reprinted in Mele (1997), pp. 131-48.
- Deigh, John, ed. (1992), *Ethics and Personality: Essays in Moral Psychology* (Chicago: University of Chicago Press).

- De Sousa, Ronald (1974), "The Good and the True", *Mind*, Vol. 83, No. 332, pp. 534-51.
- Dostoevsky, Fyodor (1993), *Notes from Underground* (New York: Vintage Books), translated by Richard Pevear and Larissa Volokhonsky.
- Dretske, Fred (1988), *Explaining Behavior: Reasons in a World of Causes* (Cambridge: MIT Press).
- Elster, Jon (1979), *Ulysses and the Sirens: Studies in Rationality and Irrationality* (New York: Cambridge University Press)
- Erikson, Erik H. (1962), *Young Man Luther: A Study in Psychoanalysis and History* (New York: W. W. Norton and Company).
- Fischer, John Martin (1994), *The Metaphysics of Free Will: An Essay on Control* (Malden, Massachusetts: Blackwell Publishing).
- Fischer, John Martin and Ravizza, Mark, eds. (1998), *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press).
- Flanagan, Owen and Rorty, Amelie Oksenberg, eds., (1990), *Identity, Character and Morality: Essays in Moral Psychology* (Cambridge: The MIT Press).
- Frankfurt, Harry (1971), "Freedom of the Will and the Concept of a Person", *Journal of Philosophy*, Vol. 68, No. 1, pp. 5-28. Reprinted in Frankfurt (1988), pp. 11-25.
- \_\_\_\_\_ (1982), "The Problem of Action", *Synthese*, Vol. 53, No. 2, pp. 257-72. Reprinted in Frankfurt (1988), pp. 69-79 and in Mele (1997), pp. 42-52.
- \_\_\_\_\_ (1987), "Identification and Wholeheartedness: in Frankfurt (1988), pp. 159-76.
- \_\_\_\_\_ (1988a), *The Importance of What We Care About* (New York: Cambridge University Press).
- \_\_\_\_\_ (1998b), "Rationality and the Unthinkable", in Frankfurt (1988a), pp. 177-90.
- \_\_\_\_\_ (2002), "Reply to Gary Watson", in Buss and Overton (2002), pp. 160-4.
- Frey, Ray and Morris, Christopher, eds. (1993), *Value, Welfare, and Morality* (New York: Cambridge University Press).
- Gibbard, Allan (1990), *Wise Choices, Apt Feelings* (Cambridge: Harvard University Press).

- Ginet, Carl (1990), *On Action* (Cambridge: Cambridge University Press).
- Goldman, Alvin I. (1970), *A Theory of Human Action* (New Jersey: Prentice Hall).
- Greene, Graham (1955), *The Quiet American* (New York: Penguin Publishing).
- Griffin, James (1996), *Value Judgment: Improving our Ethical Beliefs* (New York: Oxford University Press).
- Griffiths, A. Phillips (1958), "Acting with Reason", *Philosophical Quarterly* Vol. 8, pp. 289-99.
- Hare, Robert D. (1993), *Without Conscience: The Disturbing World of Psychopaths Among Us* (New York: Guilford Press).
- Harman, Gilbert (1976), "Practical Reasoning", *Review of Metaphysics*, Vol. 79, pp. 431-63. Reprinted in Mele (1997), pp. 149-77.
- Hill, Thomas (1986), "Weakness of Will and Character", *Philosophical Topics*, Vol. 14, pp. 93-115.
- Humberstone, Lloyd (1992), "Direction of Fit", *Mind*, Vol. 101, pp. 59-83.
- Hume, David (1992), *Treatise of Human Nature* (Buffalo: Prometheus Books).
- Kane, Robert (1996), *The Significance of Free Will* (New York: Oxford University Press).
- Kant Immanuel (1997), *Critique of Practical Reason* (Cambridge: Cambridge University Press), translated by Mary Gregor.
- Kaufman, Arnold S. (1966), "Practical Decision", *Mind*, Vol. 75, No. 297, pp. 25-44.
- Kavka, Gregory (1983), "The Toxin Puzzle", *Analysis*, Vol. 43, No. 1, pp. 33-6.
- Leibniz, Gottfried W. (1995), *Discourse on Metaphysics* (Peru, Illinois: Open Court Publishing), translated by George Mongomery.
- Levenkron, Steven (1998), *Cutting: Understanding and Overcoming Self-Mutilation* (New York: W. W. Norton and Company).
- Lewis, David (1998), "Desire as Belief", *Mind*, Vol. 97, pp. 323-32. Reprinted in Lewis (2000), pp. 42-54.
- \_\_\_\_\_ (2000), *Papers in Ethics and Social Philosophy* (Cambridge: Cambridge University Press).

- Locke, John (1964), *An Essay Concerning Human Understanding* (New York: Penguin Publishing), edited by A. D. Woozley.
- Ludwig, Kirk (1992), "Impossible Doings", *Philosophical Studies*, Vol. 65, pp. 257-81.
- Marks, Joel (1986), *The Ways of Desire: New Essays in Philosophical Psychology on the Concept of Wanting* (Chicago: Precedent Publishing).
- Matthews, Gareth B. (2005), *Augustine* (Malden, Massachusetts: Blackwell Publishing).
- McCann, Hugh (1986a), "Intrinsic Intentionality", *Theory and Decision*, Vol. 20, pp. 247-73.
- \_\_\_\_\_ (1986b), "Rationality and the Range of Intention", *Midwest Studies in Philosophy*, Vol. 10, pp. 191-211.
- McDowell, John (1978), "Are Moral Requirements Hypothetical Imperatives?", *Proceedings of the Aristotelian Society*, supplementary volume, pp. 13-29. Reprinted in McDowell (1998), pp. 77-94.
- \_\_\_\_\_ (1998), *Mind, Value, and Reality* (Cambridge: Harvard University Press).
- McIntyre, Alison (1990), "Is Akratic Action Always Irrational?", in Flanagan and Rorty (1990), pp. 379-400.
- Mele, Alfred R. (1987), *Irrationality: A Study on Akrasia, Self-Deception, and Self-Control* (New York: Oxford University Press).
- \_\_\_\_\_ (1992), *Springs of Action: Understanding Intentional Behavior* (New York: Oxford University Press).
- \_\_\_\_\_ (1995), "Motivation: Essentially Motivation-Constituting Attitudes", *The Philosophical Review*, Vol.104, No. 3, pp. 394-5.
- \_\_\_\_\_, ed., (1997), *The Philosophy of Action* (Oxford: Oxford University Press).
- \_\_\_\_\_ (2000), "Deciding to Act", *Philosophical Studies*, Vol. 100, pp. 81-108.
- Mele, Alfred R. and Moser, Paul K. "Intentional Action", *Nous*, Vol. 28, pp. 39-68. Reprinted in Mele (1997), pp. 223-55.
- Mill, John Stuart (1979), *Utilitarianism*, (Indianapolis: Hackett Publishing), George Sher, ed.
- Milton, John (1998), *Paradise Lost* (New York: Addison Wesley Longman, Inc.), 2<sup>nd</sup> edition, Alastair Fowler, ed.

- Nagel, Thomas (1970), *The Possibility of Altruism* (Princeton, Princeton University Press).
- Parfit, Derek (1984), *Reasons and Persons* (Oxford: Clarendon Press).
- Pears, David (1998), *Motivated Irrationality* (South Bend: St. Augustine's Press).
- Pink, Thomas (1996), *The Psychology of Freedom* (New York: Cambridge University Press).
- Plato (1997), *Complete Works* (Cambridge: Hackett Publishing), John M. Cooper, ed.
- Poe, Edgar Allan, (1975), *The Complete Tales and Poems of Edgar Allan Poe* (New York: Random House).
- Price, Huw (1989), "Defending Desire-as-Belief", *Mind*, Vol. 98, No. 389, pp.119-27.
- Quinn, Warren (1993), "Putting Rationality in its Place", in Frey and Morris (1993), pp. 26-50.
- Raz, Joseph (1986), *The Morality of Freedom* (New York; Oxford University Press).
- \_\_\_\_\_ (1995), "On the Moral Point of View", in Schneewind (1995), pp. 58-83.
- \_\_\_\_\_ (1999a), *Engaging Reason* (New York: Oxford University Press).
- \_\_\_\_\_ (1999b), "Agency, Reason, and the Good", in Raz (1999a), pp. 22-45.
- \_\_\_\_\_ (1999c), "Incommensurability and Agency", in Raz (1999a), pp. 46-67.
- Robins, Michael H. (1984), *Promising, Intending, and Moral Autonomy* (Cambridge: Cambridge University Press).
- Ryle, Gilbert (1949), *The Concept of Mind* (Chicago: The University of Chicago Press).
- Saavedra, Miguel de Cervantes (1995), *La Galatea* (Madrid: Cátedra).
- Santas, Gerasimos (1966), "Plato's *Protagoras* and Explanations of Weakness", *Philosophical Review* Vol. 65, pp. 3-33;
- Scanlon, Thomas (1998), *What We Owe to Each Other* (Cambridge: Harvard University Press).
- Schopenhauer, Arthur (1928) *The Philosophy of Schopenhauer* (New York: The Modern Library), Edman, Irwin, ed.

- Schiffer, Steven (1976), "A Paradox of Desire", *American Philosophical Quarterly* Vol. 13, No. 3, pp.197-8
- Schneewind, J. B. (1995), *Reason, Ethics, and Society: Essays in Honor of Kurt Baier* (Chicago: Open Court Publishing).
- Schroeder, Timothy (2004), *Three Faces of Desire* (Oxford: Oxford University Press).
- Schuler, G. F. (1991), "Pro-attitudes and Direction of Fit", *Mind*, Vol. 100, pp. 277-81.
- \_\_\_\_\_ (1995), *Desire: Its Role in Practical Reason and the Explanation of Action* (Cambridge: The MIT Press).
- Searle, John (1983), *Intentionality: an Essay in the Philosophy of Mind* (New York: Cambridge University Press).
- Sidgwick, Henry (1981), *The Methods of Ethics* (Cambridge: Hackett Publishing), 7<sup>th</sup> edition.
- Smith, Michael (1987), "The Humean Theory of Motivation", *Mind*, Vol. 96, pp. 36-61.
- \_\_\_\_\_ (1994), *The Moral Problem* (Oxford: Blackwell Publishing).
- Spinoza, Baruch (1992), *The Ethics: Treatise on the Emendation of the Intellect* (New York: Hackett Publishing), Seymour Feldman, ed.
- Stalnaker, Robert (1984), *Inquiry* (Cambridge: The MIT Press).
- Stampe, Dennis W. (1987), "The Authority of Desire", *The Philosophical Review*, Vol. 96, pp. 335-81.
- Stocker, Michael (1979), "Desiring the Bad: An Essay in Moral Psychology", *The Journal of Philosophy*, Vol. 76, pp. 738-53.
- \_\_\_\_\_ (1990), *Plural and Conflicting Values* (New York: Oxford University Press).
- \_\_\_\_\_ (2004), "Raz on the Intelligibility of Bad Acts", in Wallace (2004), pp. 303-332.
- Stocker, Michael with Hegeman, Elizabeth (1996), *Valuing Emotions* (Cambridge: Cambridge University Press).
- Strachey, James, ed. (1986), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, (London: Hogarth Press).

- Strawson, Galen (1994), *Mental Reality* (Cambridge: The MIT Press).
- Stroud, Sarah and Tappolet, Christine, eds. (2003), *Weakness of the Will and Practical Irrationality* (Oxford: Oxford University Press).
- Suyemoto, Karen L. (1998), "The Functions of Self-Mutilation", *Clinical Psychology Review*, Vol. 18, No. 5, pp. 531-54.
- Taylor, Richard (1964), *The Explanation of Behavior* (New York: The Humanities Press).
- Tenenbaum, Sergio (2003), "Accidie, Evaluation, and Motivation", in Stroud and Tappolet (2003), pp. 147-71.
- Thalberg, Irving (1984), "Do Our Intentions Cause Our Intentional Actions?", *American Philosophical Quarterly*, Vol. 21, No. 3, pp. 249-60.
- Velleman, J. David (1989), *Practical Reflection* (Princeton; Princeton University Press).
- \_\_\_\_\_ (1992a), "The Guise of the Good", *Nous*, Vol. 26, pp. 3-26. Reprinted in Velleman (2000), pp. 99-122.
- \_\_\_\_\_ (1992b), "What Happens When Someone Acts?", *Mind*, Vol. 101, pp. 461-81. Reprinted in Velleman (2000), pp. 123-43.
- \_\_\_\_\_ (1996), "The Possibility of Practical Reason", *Ethics*, Vol. 106, pp. 694-726. Reprinted in Velleman (2000), pp. 170-99.
- \_\_\_\_\_ (2000a), *The Possibility of Practical Reason* (Oxford: Oxford University Press).
- \_\_\_\_\_ (2000b), "Introduction", in Velleman (2000a), pp. 1-31.
- Vermazen, Bruce and Hintikka, Verrill B., eds. (1985), *Essays on Davidson: Actions and Events* (New York: Oxford University Press).
- Voltaire (2002), *Candide*, (New York: Modern Library Edition).
- Wallace, R. Jay (1990), "How to Argue About Practical Reason", *Mind*, Vol. 99, pp. 355-85.
- \_\_\_\_\_ ed., (2004), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* (Oxford: Oxford University Press).
- Watson, Gary (1975), "Free Agency", *The Journal of Philosophy* Vol. 72, No. 8 (April, 1975), pp. 205-20. Reprinted in Watson (2003), pp. 337-351.

\_\_\_\_\_ (1987), "Free Action and Free Will", *Mind* Vol. 96, No. 382, pp. 145-72.

\_\_\_\_\_, ed., (2003), *Free Will* (New York: Oxford University Press), 2<sup>nd</sup> edition.

Williams, Bernard (1993a), *Moral Luck* (Albany: State University of New York Press).

\_\_\_\_\_ (1993b), "Internal and External Reasons", in Williams (1993a), pp. 101-13.

Wilson, George (1989), *The Intentionality of Human Action* (Stanford: Stanford University Press).

Wright, Larry (1976), *Teleological Explanations: An Etiological Analysis of Goals and Functions* (Berkeley: University of California Press)..

## BIOGRAPHICAL SKETCH

Peter Brian Barry is a graduate of the University of Florida Department of Philosophy. He has also earned master's degrees in applied philosophy and philosophy from Bowling Green State University (Ohio) and the University of Wisconsin-Milwaukee, respectively. He has also earned a bachelor's degree from the University of Wisconsin-Milwaukee after majoring in philosophy and political science. He is the son of Barbara Jean Barry and the twin brother of Robert James Barry.