

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Philosophy Dissertations, Theses, & Student  
Research

Philosophy, Department of

---

Summer 7-20-2011

# Doing Without Desiring

Steven E. Swartzter

University of Nebraska-Lincoln, [swartzter@unlserve.unl.edu](mailto:swartzter@unlserve.unl.edu)

Follow this and additional works at: <http://digitalcommons.unl.edu/philosophydiss>



Part of the [Ethics and Political Philosophy Commons](#), and the [Philosophy of Mind Commons](#)

---

Swartzter, Steven E., "Doing Without Desiring" (2011). *Philosophy Dissertations, Theses, & Student Research*. 5.  
<http://digitalcommons.unl.edu/philosophydiss/5>

This Article is brought to you for free and open access by the Philosophy, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Philosophy Dissertations, Theses, & Student Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

DOING WITHOUT DESIRING

by

Steven E. Swartzner

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Philosophy

Under the Supervision of Professor Mark van Roojen

Lincoln, Nebraska

July, 2011

# DOING WITHOUT DESIRING

Steven Swartzter, Ph.D.

University of Nebraska, 2011

Adviser: Mark van Roojen

This dissertation defends a cognitivist alternative to the Humean belief-desire theory of motivation against standard philosophical arguments.

Moral judgments influence our action. For instance, someone might donate to charity because she believes she has a duty to give back to her community. According to the Humean orthodoxy, some additional state—some passion or desire—is needed to explain her action. She may *want* to donate the money, to give back to her community, or to fulfill her duty. Yet there must be something she wants, the Humean insists, because only desires are capable of moving us. Even moral judgment is no more than desire's slave.

This dissertation explores the possibility that cognitive states are capable of playing a directly motivational role. I argue that the standard philosophical arguments against this possibility do not survive close scrutiny. Instead of proceeding from assumptions about rationality, morality, and agency that frequently drive motivational cognitivists, my arguments are distinctive in that they are built largely out of Humean materials; these arguments show how cognitivism is compatible with many of the considerations Humeans have used to make their account seem compelling.

For instance, agents who are unmoved by their moral judgments are often taken as evidence for the Humean Theory. This is odd, since agents are not uniformly moved by their desires either. Moral beliefs and desires seem as though they may be closely analogous in this respect.

I also try to show that desire-based motivation might serve as a useful model for cognitive motivation by arguing that cognitivists can explain motivated action in ways that parallel desire-based explanations. While these cognitivist explanations are committed to the existence of *besires*, I argue that this is no problem for the view. Humean *a priori* proclamations that besires would be incoherent or absurd notwithstanding, the arguments of this dissertation suggest that besires are not so bizarre. Indeed, I argue that their existence would follow from plausible empirical hypotheses.

To the Anderson, Brown, Ertl, and Swartzer clans,  
and to my fellow graduate students at UNL,  
because it sometimes takes two very different villages

## Acknowledgments

This dissertation would have been much worse, and would likely have never been finished, were it not for the help and support of a large number of people.

My committee—composed of Mark van Roojen, Joe Mendola, David Sobel, and John Gibbons—influenced this project in countless ways. In addition to the feedback they provided on several drafts and the countless discussions, their work (especially their work on this topic) taught me more than they might realize. Indeed, many of the ideas in this project were inspired by courses taught by Joe and John (and also, I should note, by Jennifer McKittrick).

I would like to give special thanks to Mark and Joe for knowing when (and how) to push me, and when to let me do my own thing.

I would also like to thank the other members of the UNL faculty for their instruction and encouragement throughout the years, especially Aaron Bronfman, Al Casullo, Janice Dowell, Reina Hayaki, and David Henderson. I also owe a great deal of gratitude to Donette Petersen and Margaret Skean for making things much easier here than they might have otherwise been.

Parts of Chapter 2 were presented under the title “Humeanism and Amoralism” at the Central States Philosophical Association meeting in October 2009 and the Central Division Meeting of the American Philosophical Association in February 2010. I thank Robyn Gaier and Melissa Barry for their challenging and helpful comments. Some of the ideas of Chapter 3 were presented under the title “Unmanifested Desires Unmasked” at the Central Division Meeting of the American Philosophical Association in April 2011.

Many thanks to my commentator Naomi Reshotko and to the audience for the helpful (and immensely enjoyable) conversation.

Almost all of these ideas were presented at some point or another at UNL's Philosophy Graduate Student Colloquium. I must apologize to my colleagues who were subjected to talk after talk about desires, besires, dispositions, and the rest. I must also thank them. The questions and comments I received at these talks were by far the most important tool for improving this project, as well as for my philosophical development more generally. These Colloquia helped me realize where I needed to be clearer, where I needed a more complete or more convincing argument, where I was making bad moves, and when I was simply heading in the wrong direction. These discussions also convinced me that I was mostly on the right track.

Given the nearly complete overlap, I find it hard to separate the philosophical feedback and encouragement from the personal encouragement and friendship that I have received from the graduate students throughout my years at UNL. So I will not try. Many thanks to Brent Braga, David Chavez, Mark Decker, Matt Dee, Luke Elwonger, Allison Fritz, Cullen Gatten, Megan Herington, Leo Iacono, Clare LaFrance, Errol Lord, Christopher McCammon, Justin Moss, Tim Loughlin, Sruthi Rothenfluch, Adam Thompson, Virendra Tripathi, Preston Werner, and also to three of my oldest friends—Bre Fordahl Ratnayake, Garreth McMaster, and Melissa Montang—who don't exactly belong on this list, but belong somewhere. All of these people have, in their own ways, helped me make it through. That is why this dissertation is dedicated (in part) to them.

I would most of all like to thank my family, to whom this dissertation is also dedicated. Thanks to Mary, Joe, Theresa, Dawn, Brian, Kevin, Susan, Cindy, Katie,

Tracy, Karen, Tony, Alex, and all of my nieces and nephews for their support, encouragement, and love. I thank them for believing that I could do this and for being proud of me now that I have. And, of course, I owe most to my parents—Joseph and Dorothy Swartzler—and the encouraging words I have heard from them so many times: “Have a good day, do your best, and I love you.”

# DOING WITHOUT DESIRING

Steven Swartz

## Table of Contents

ABSTRACT.....	ii
ACKNOWLEDGMENTS .....	v
TABLE OF CONTENTS.....	viii
CHAPTER 1: THE HUMEAN ORTHODOXY.....	1
1. Introduction.....	1
2. Clarifying the Target.....	4
2.1 An Extremely Rough Initial Characterization .....	4
2.2 The Strength of ‘Must’ .....	4
2.3 Humean Desires .....	9
2.4 A Note on Terminology.....	15
3. Why We Cannot Ignore Motivational Cognitivism.....	17
3.1 Instrumental Reasoning: The Canonical Cases.....	17
3.2 Instrumental Reasoning: Less Paradigmatic Cases.....	19
3.3 Reasoning.....	22
4. Things to Come.....	29
CHAPTER 2: DEPRESSION, AMORALISM, AND THE HUMEAN THEORY ....	33
1. Introduction.....	33
2. The Externalist Cases.....	34
2.1 Internalism: A First Pass.....	35

2.2 Depression.....	36
2.3 Amoralism.....	38
3. Motivational Cognitivism and Depression .....	44
4. Motivational Cognitivism and Amoralism .....	52
5. Humean Explanations .....	57
CHAPTER 3: DISPOSITIONS, MOTIVATION, AND DESIRE .....	65
1. Being Moved.....	66
1.1 Two Senses of Motivation .....	66
1.2 Motivation and Dispositions .....	67
1.3 Motivation and Masking .....	71
2. The Dispositionalist Framework.....	73
2.1 Some Preliminaries .....	73
2.2 Behavioral Accounts .....	74
2.3 Non-Behavioral Accounts.....	78
2.4 Hybrid Accounts .....	81
2.5 The Structure of Dispositionalism .....	83
3. Refining the Dispositionalist Account.....	85
3.1 The Standard Move.....	85
3.2 Wanting More (or Less).....	87
3.3 Wanting <i>Simpliciter</i> .....	91
3.4 Why Desires Motivate only Contingently .....	94
4. Silencing, Continenence, and Motivational Variation.....	97

CHAPTER 4: COGNITIVISM AS AN EMPIRICAL HYPOTHESIS .....	102
1. A Preliminary Account of Desire .....	102
1.1 Characteristic Manifestations of Desire Redux .....	102
1.2 The Hybrid Account of Desire.....	107
2. Two Anti-Humean Empirical Hypotheses.....	115
2.1 The Negative Hypothesis.....	115
2.2 The Positive Hypothesis .....	118
2.3 Are the Cognitivist Hypotheses False? .....	121
2.4 The Dialectic to Come .....	125
3. A Humean Objection .....	126
4. Purpose-Giving Desires vs. Desires Proper.....	129
4.1 Nagel's Complaint .....	129
4.2 Why Nagel's Complaint (As Is) Won't Quite Work .....	131
4.3 Refining Nagel's Complaint I: The Model .....	133
4.4 Refining Nagel's Complaint II: Extending the Model.....	135
5. The Guise of the Good.....	136
5.1 The <i>Purpose-Value Link</i> .....	136
5.2 Refining <i>The Purpose-Value Link</i> .....	139
5.3 The <i>Purpose-Value Link</i> and Motivational Cognitivism.....	143
6. Review and Preview .....	145
CHAPTER 5: WHAT'S SO BIZARRE ABOUT BESIRES?.....	147

1. Introduction.....	147
2. The Stalemate.....	149
3. The Fuss about Fit.....	151
4. Normative Characterizations .....	157
5. Causal and Counterfactual Characterizations .....	162
6. Appetites, Needs, and Besires.....	170
7. Motivated Desires, Unmotivated Desires, and Besires.....	177
8. Concluding Remarks.....	180
REFERENCES .....	184

## Chapter 1

### The Humean Orthodoxy

#### 1. Introduction

Moral judgments have effects. Our thoughts about what morality demands often influence what we do. To take a toy example, suppose that Anne volunteers to build low-cost housing for those in need. When asked why she does this, Anne might answer that successful people have an obligation to give back to the community and that volunteering was the best way for her to accomplish this. Alternatively, she might say that it's wrong to sit back and do nothing when one could easily help someone in need. Or, were she feeling less wordy, Anne may simply respond, "It's the right thing to do."

At first blush, any of the potential responses would render Anne's action suitably intelligible and thus, if true, would explain why she volunteers. I will simply take it for granted that moral judgments, such as Anne's, are at least partly explanatory of our actions and that they are so in virtue of some causal role that they play in the production of those actions. The question of interest to me is *what* causal role do they play? *How* do these beliefs shape our behavior?

According to the current philosophical orthodoxy, there are two kinds of roles that a given mental state could play in the production of action. One role can be described as *motivational*. States that play the motivational role are, in effect, the sources of intentional action—they provide a goal and give the "oomph" that "pushes" us to achieve them. Motivational states can be contrasted with what we might call *informational* states, which play a distinct role. Informational states are supposed to describe how things are,

so as to illuminate the path by which our motivationally-determined goals can be achieved.

A second piece of the philosophical orthodoxy is that the distinction between informational and motivational states maps closely on to the distinction between reason and passion—between the cognitive side of the mind that involves belief, perception, and theoretical rationality and the non-cognitive side of the mind that houses desire and emotion. A corollary of this is that all behavior ultimately has its source in goals that are determined by our desires and other non-cognitive passionate states; all cognitive states, including moral beliefs, influence behavior only by playing the informational role. According to this Humean belief-desire theory of motivation, cognitive states are no more than desire’s slaves, and cannot move us to act without desire’s command.<sup>1</sup>

In this dissertation, I argue that the Humean orthodoxy is not well-supported, and that cognitivist theories of motivation merit a closer look.<sup>2</sup> The view I defend here is that

---

<sup>1</sup> This metaphor of reason being desire’s slave is, of course, from David Hume ([1888], 415). It is worth noting that there is some controversy with regard to how best to understand Hume’s views. Several philosophers have noted that Hume may not endorse many of the claims that are claimed to be Humean. See, for instance, Bernard Williams (1980, 101-102), Michael Smith (1994, 113), and Elijah Millgram (1995). The label ‘Humean’, then, should be understood to indicate not that these views are Hume’s, but that they are in some important sense inspired by the views Hume forwarded.

<sup>2</sup> The label ‘motivational cognitivism’ has become a standard label for anti-Humean theories claiming that cognitive states are capable of being directly motivational in the sense specified in the text. I accept this usage here, and for the sake of convenience will often abbreviate this as ‘cognitivism’. This gives rise to a potential confusion, as ‘cognitivism’ is also the standard label in metaethics for the claims that moral assertions express beliefs and that moral claims are truth-evaluable. This view—we can call it ‘metaethical cognitivism’—is not the topic of this dissertation. Motivational cognitivism and metaethical cognitivism are somewhat independent of one another, in that many metaethical cognitivists reject motivational cognitivism (see, for instance, David Brink (1986), Peter Railton (1986), Richard Boyd (1988), and Michael Smith (1994), among others). At the same time, I know of no major philosophers who reject metaethical cognitivism while endorsing motivational cognitivism. This is likely due to the fact that one of the historically dominant motivations for rejecting metaethical cognitivism stems from the Humean Theory (for a discussion of this line of reasoning, see Russ Shafer-Landau (2003,

it is plausible that some cognitive states—including, but not limited to moral judgments—might be well-suited to shape our behavior by playing a directly motivational role. I will not ultimately argue that these cognitive states *do* play this role; instead I contend that the philosophical arguments against this possibility cannot survive close scrutiny. This is because, as I will try to argue, this is ultimately an *a posteriori* question. That motivation could sometimes be the result of wholly cognitive forces remains a plausible, but somewhat speculative, possibility.

While my goals are relatively modest, my overall strategy for achieving them departs in a number of crucial ways from standard attempts to upset the Humean belief-desire orthodoxy. My approach is distinctive in that it does not proceed from lofty assumptions about rationality, morality, and agency that often drive the motivational cognitivists. Instead, the arguments I develop are built largely out of Humean resources. I try to show that the possibility of wholly cognitive motivation is actually compatible with—and in some ways even supported by—the very considerations that have made the belief-desire picture seem so attractive.

In this introductory chapter I have three goals. Section 2 clarifies the thesis at which the arguments of this dissertation are directed. Section 3 provides an initial argument for the view I intend to defend; this argument is meant to establish that we need to take motivational cognitivism seriously. Section 4 offers an overview of the rest of the project.

---

121), Connie Rosati (2008), and Mark van Roojen (2011)). In this dissertation, I will be simply assuming the truth of metaethical cognitivism in order to show that it is compatible with a plausible anti-Humean Theory of Motivation.

## 2. Clarifying the Target

### 2.1 An Extremely Rough Initial Characterization

In the broad, largely metaphorical terms used in Section 1, the Humean picture should sound familiar enough. This loose idea, however, is not yet sufficiently precise to be useful, in that it is subject to a number of incompatible interpretations. We must therefore begin by clarifying the position at issue in this dissertation. We can get a better grip on our target if we reflect on and refine a rough, initial characterization of the Humean view:

***Desires are Necessary:***

In order for an agent  $S$  to be motivated to perform some action,  $\phi$ ,  $S$  must have a desire appropriately related to  $\phi$ -ing.

### 2.2 The Strength of ‘Must’

*Desires are Necessary*, and its refinements that follow, are explicitly modal—they provide a condition that *must* be met for motivation. The Humean Theory is frequently defended as a conceptually necessary truth about the relationship between motivation and desire. As a case in point, Michael Smith’s (1987; 1994, Ch. 4) main argument for the claim that desires are required for goal-directed action proceeds on the conceptual grounds that in order to constitute or provide a goal, an attitude must be such that the world is to “fit” it and that this world-to-mind “direction of fit” is the mark of desire.

To shore up a key assumption of the argument—namely, that no beliefs also have that world-to-mind “direction of fit”—Smith offers two arguments. The first argument is clearly intended to be conceptual; it is that it would be *incoherent* for an attitude to have

both belief's and desire's directions of fit (Smith 1994, 118).<sup>3</sup> Thus Smith's argument promises to support the Humean Theory as a conceptually necessary truth.

Smith's (1994, 119-121) second supporting argument is that the belief-like aspect of moral judgment (or any other purportedly motivational cognitive state) is "modally separable" from its desire-like aspect, in that it is possible to have the former without the latter. If these two aspects can be "pulled apart" in this way, Smith thinks, they are not a single, unitary thing. Thus the desire-like direction of fit is not embodied by the belief but by some distinct state, like desire. The motivation that seems to originate in the moral judgment really originates in that distinct desire. On a natural interpretation, this argument rests on two claims about conceptual possibilities: (1) If an agent could conceivably have some mental state (e.g. some moral judgment) and not be motivated accordingly, then that mental state lacks desire's direction of fit and (2) It is conceivable that agents may judge some course of action to be morally right or wrong without being motivated to act accordingly.

I will have a great deal more to say about Smith's arguments in later chapters; here I am content to point out that they are naturally interpreted as attempts to support the Humean Theory as a conceptually necessary truth. This is not, however, the only way to interpret the modality involved in our statement of the Humean Theory. It is widely acknowledged that there are *a posteriori* necessities that cannot be arrived at by armchair analysis of our concepts. Water's chemical structure is essentially H<sub>2</sub>O; some philosophers may also find it plausible that an individual's parentage or genetic makeup

---

<sup>3</sup> David Lewis (1988; 1996) is also often interpreted as arguing that the anti-Humean Theory is incoherent. Lewis's argument is that a particular interpretation of the theory—what he calls the "Desire as Belief Thesis"—runs afoul of the constraints set forth by decision theory. For a related decision-theoretic argument, see John Collins (1988).

and an artwork's creator are essential to them. However, these supposedly necessary truths cannot be found out without serious empirical work.<sup>4</sup> For this reason some philosophers may wish to defend the Humean Theory as an *a posteriori* metaphysically necessary truth. Alternatively, and perhaps more plausibly, it may be argued that the Humean Theory is a law of nature about human psychology, rather than a matter of conceptual or metaphysical necessity. In either case, it may be that motivation involves desire essentially or in some law-like way, but that this can be known only by drawing on empirical support of some kind.

Indeed, there is a way of interpreting Smith's second supporting argument to be compatible with this. Instead of insisting on the conceptual claims above, Smith might instead opt for an alternative: If actual agents have some mental state (e.g. some moral judgment) and are not motivated accordingly, then that mental state lacks desire's direction of fit. If the argument rests on only actual agents, and our *a posteriori* knowledge of them, then there is no worry that it relies on conceivable agents that are not metaphysically or nomologically possible. Indeed, insofar as Smith (1994, 119-121) uses actually-occurring phenomena like weakness of will and severe depression to establish that moral belief and motivation are sometimes "pulled apart", his argument can be reasonably interpreted as an attempt to provide partly *a posteriori* grounds for thinking that beliefs cannot have desire's direction of fit. Interpreted in this way, the argument consists of a conceptual part (concluding in the claim that the desire-like direction of fit is

---

<sup>4</sup> The insight that there can be claims that are metaphysically necessary that are not conceptually necessary or knowable *a priori* is due largely to Saul Kripke (1980).

required for motivation) and a partly *a posteriori* part (concluding in the claim that moral judgments do not have that desire-like direction of fit).<sup>5</sup>

A number of other philosophical arguments may also be used to support the Humean Theory as an *a posteriori* claim. For instance, Sigrún Svavarsdóttir (1999) and Neil Sinhababu (2009) support the Humean Theory by arguing that it best explains various phenomena. Svavarsdóttir, for instance, argues that Humeanism best explains the variation in moral motivation exhibited by typical human agents; Sinhababu is concerned with, among other things, the way it feels to be motivated from a feeling of duty. It is clear that the variation in moral motivation exhibited by actual (rather than merely conceivable) agents is itself an observable phenomenon—one that requires an explanation that is at least partly empirical.<sup>6</sup> The phenomenology of a feeling of obligation should also be explainable empirically. If it were true that these phenomena are best explained by a picture of the mind that posits two essentially different types of mental state, this would lend some *a posteriori* support for that picture. Moreover, we should expect that the Humean explanatory hypotheses Svavarsdóttir and Sinhababu provide could be supported or upset by further empirical investigation into human psychology.

It is not always clear whether such inferences to the best explanation are meant to support the claim that belief and desire are *essentially* different in the way the Humean

---

<sup>5</sup> I say ‘partly *a posteriori*’ here because in addition to the claim about what actual agents are (not) motivated to do, the present understanding of Smith’s argument would require two further assumptions: (i) that if an agent is not actually motivated by a particular mental state, then that state does not have the desire-like direction of fit, and (ii) mental states have their direction of fit essentially. Both of these assumptions are presumably *a priori*.

<sup>6</sup> Indeed, this fact plays a fairly large role in her argument, as will be clear later.

orthodoxy understood as a claim of metaphysical necessity would suggest. Sinhababu, for one, seems to resist this.<sup>7</sup> Sinhababu (2009, 467) is explicit about the fact that he is defending the Humean Theory “merely as a truth about actual actions performed by human beings and about all the actions that humans are psychologically capable of performing.” On his view, then, the Humean Theory should be seen as a law of human psychology, rather than a statement of metaphysical necessity.

My primary target in this dissertation is the Humean Theory conceived of as a conceptually necessary claim about the relationship between motivated action and desire. I argue that so-conceived, the Humean Theory is false—it is at least conceivable that agents are sometimes moved to act by their moral judgments or other cognitive states, without the aid of any distinct conative attitude.

Other versions of the Humean Theory will serve as secondary targets. My conclusions regarding these views are, by necessity, weaker and largely defensive. I contend that the philosophical arguments for the Humean Theory, conceived of as a metaphysically or nomologically necessary claim about the relationship between motivated action and desire (in human agents), carry no weight. In order to do this, I will try to make it clear why cognitivists can offer explanations of motivated action that mirror Humean explanations in key ways. I am in no position to settle whether the cognitivist or Humean explanations are correct. Instead, I try to show that given what we currently know, cognitivist explanations are at least as plausible as Humean explanations.

---

<sup>7</sup> Although this is not completely clear. In making this claim, Sinhababu (2009, 467) contrasts his view only with the claim that the Humean Theory is a *conceptually* necessary truth. Sinhababu fails to recognize that the view might be forwarded as a metaphysically necessary truth about the *a posteriori* essences of belief, desire, and motivation.

In doing so, I will also try to draw out some empirical hypotheses that may help decide the issue.

### 2.3 Humean Desires

According to *Desires are Necessary*, agents who are motivated to act in some way “have a desire”. On at least some interpretations, cognitivists can safely accept this condition.

Consider the following deflationary view: to say that *S* desires that *p* is just another way to restate the claim that *S* is doing something that promotes (or seems to promote) *p*. This view straightforwardly entails *Desires are Necessary*. However, it is compatible with a number of quite un-Humean—even anti-realist—views of desire. For instance, one might hold that the desires we attribute to others for the purposes of explaining their behavior are no different from the “desires” we are talking about with respect to inanimate objects—as when (for instance) we say that the robotic vacuum cleaner wants to find the charging station, that the car wants to pull to the left, or that the vending machine doesn’t want to take our money. This is not an account of desire on which Humeans should rely when interpreting *Desires are Necessary*—these “desires” are clearly not Humean *passions*.

Less radically, a philosopher may think that talking about human desire is just a way of talking about certain causal processes that are initiated by psychologically real cognitive states, like beliefs.<sup>8</sup> *S* desiring that *p* (and also *S* being motivated to pursue *p*), on this view, need not involve any more than *S* being caused to act in some fashion by the belief that acting in that fashion would promote *p*, on the condition that such causation

---

<sup>8</sup> A view that is similar to this in some respects has been defended by Jonathan Dancy (1993, Ch. 3).

proceeds by way of some non-deviant process. Or, consider a weaker version of this story: there are some psychologically real states worthy of being called “desires”, and these states might sometimes be implicated in the explanation of motivated action, but these states needn’t be implicated in every case in which we are willing to explain *S*’s behavior by reference to the fact that she desires that *p*. On this weaker story, there is a sense of “desire” according to which *Desires are Necessary* is true, but this sense of “desire” does not always refer to the psychologically real desires; sometimes the fact that the action is caused along the lines indicated by the stronger story is sufficient grounds for attributing a desire to the agent.<sup>9</sup> On these stories, it would be difficult to agree with Hume’s claim that reason is desire’s slave—for, both of these stories portray reason (that is, belief) as capable of being its own boss in the practical domain.

I do not mean to suggest here that any of these accounts of desire are true (although later in this dissertation I will defend the plausibility of the last one). The point is just that *Desires are Necessary* does not adequately capture what is supposed to be the distinctive claim behind the Humean Theory. What is missing is, in part, the claim that the desires in question are “psychologically real” states—to “have a desire” in the relevant sense, is to have a psychologically real, causally relevant, attitude and that attitude must be worthy of the label “desire”.<sup>10</sup> We can thus refine our target:

---

<sup>9</sup> Thomas Nagel (1970) seems to defend a view along these lines; however, it is not always clear how to best interpret his discussion of these matters. I will say more about Nagel in Chapters 4 and 5.

<sup>10</sup> To say that these desires are “psychologically real” is not to say that they are conscious—for unconscious desires can potentially explain our behavior too. I only mean to be indicating that it would be a *bona fide* attitude of the agent (as opposed to, for instance, the state of the robotic vacuum cleaner).

***Real Desires are Necessary:***

In order for an agent  $S$  to be motivated to perform some action,  $\phi$ ,  $S$  must have a psychologically real, causally relevant desire, appropriately related to  $\phi$ -ing.

While *Real Desires are Necessary* is an improvement over our original characterization, it too must be refined if it is to capture the interesting sense in which action is supposed to have its source in the passionate side of the mind. For, the cognitivist may argue, some (perhaps even all) such desires might themselves be the result of wholly cognitive forces.

The use of the word “result” here is deliberately vague. There are at least two different ways in which a desire might rightly be said to result from things wholly on the side of reason. One possibility is that some or all desires can be *reduced to* cognitive states—if, for example, desires are really a species of belief or perception.<sup>11</sup> Alternatively, some desires might be the *causal* result of purely cognitive forces. It should be clear that an agent can revise her existing extrinsic or instrumental desires in light of new information. If an agent initially desires that  $q$  because she believes that  $q$  is a necessary condition for something else that she wants—say, that  $p$ —and then comes to discover that she is wrong about this, she may come to abandon that desire; or if she desires that  $p$  and comes to find that  $\phi$ -ing is an efficient way to bring it about that  $p$ , the agent may form a new desire to  $\phi$ . In such cases, an agent’s body of desires is altered as the result of cognitive forces, but not as the result of *wholly* cognitive forces. In both cases, the more fundamental desire that  $p$  plays a key explanatory role in the desire

---

<sup>11</sup> A variety of views that hold that things are desired only under the “guise of the good” fall under this first branch. For further discussion of the guise of the good, see, Michael Stocker (1979), J. David Velleman (1992), Sergio Tenenbaum (2007), and the various essays in Tenenbaum (2010). I say more about this in Chapters 4 and 5.

revision.  $S$ 's desire to  $\phi$  would be the causal result of wholly cognitive forces if, for instance, her belief that it would be morally good that  $p$  and her belief that  $\phi$ -ing would promote  $p$  together caused the desire, and if no other antecedent desire had a hand in this.<sup>12</sup> If an agent's psychologically real, causally relevant desire were the result of wholly cognitive forces either in our earlier constitutive sense or in this causal sense, then the cognitive side of the mind would still be calling the shots.

Some self-described Humeans would balk at the inclusion of this last condition. Smith's (1994, 92) view is that the Humean Theory is, at bottom, an account of motivating reasons:

R at t constitutes a motivating reason of agent A to  $\Phi$  iff there is some  $\psi$  such that R at t consists of an appropriately related desire of A to  $\psi$  and a belief that were she to  $\Phi$  she would  $\psi$ .

It is clear, on Smith's view, that the desires that are constitutive of motivating reasons are not themselves results of rational forces in our first sense—he clearly holds that these desires are not themselves constituted by beliefs.<sup>13</sup> Nor are they, presumably, reducible to any other type of cognitive state. However, Smith's (1994, 193) considered view is that beliefs about what is rationally required can give rise to new desires:<sup>14</sup>

---

<sup>12</sup> Nagel's (1970) view is sometimes interpreted along these lines. See, for instance, Shafer-Landau (2003, 137-138). For further discussion, see Dancy (1993, Ch. 1). I will discuss this briefly in Chapter 5.

<sup>13</sup> See, for instance, Smith's (1988b) response to Philip Pettit (1987).

<sup>14</sup> I take this to be a departure from his seminal work defending the Humean Theory. There is one argument in "The Humean Theory of Motivation" that is notably absent from Smith's later restatement of the view in Chapter 4 of *The Moral Problem*. In this initial argument, Smith (1987, 58-60) rejects the idea that some desires can be "motivated" by states other than desires, because any state that could motivate a desire must have the appropriate world-to-mind direction of fit. It is natural to think that this is supposed to mean that any process by which a new desire is produced must have, among its antecedents, some attitude that has desire's direction of fit. Given Smith's commitment to the claim that no beliefs have that

[I]f we believe that  $\Phi$ -ing in certain circumstances  $C$  has the feature that we would want acts to have in  $C$  if we were fully rational then, on the one hand, to the extent that we are rational either we will already want to  $\Phi$  in  $C$  or our belief will cause us to have this want .... [Emphasis added.]

Moreover, the process by which these desires are produced need not draw on any prior desire of the usual sort—Smith (1994, 71-76) is adamant that rational agents need not be motivated by any desire *to do the morally right thing*, where this desire is understood *de dicto*.

Insofar as Smith attempts to rule out the first (i.e. reductive) way in which desires might “result” from entirely cognitive factors, he offers a substantive departure from what many self-described cognitivists wish to argue and this view has as much a claim to the label “Humean” as any.<sup>15</sup> For my purposes, it will be safe to skate over these issues. It is safe for us to land on the following admittedly vague characterization of the Humean Theory:

***The Humean Theory of Motivation:***

In order for an agent  $S$  to be motivated to perform some action,  $\phi$ ,  $S$  must have a psychologically real, causally relevant desire that is appropriately related to  $\phi$ -ing and suitably independent of her cognitive states.

I will, for the most part, leave it open whether or not desires that are the causal product of wholly causal forces are “suitably independent” of an agent’s cognitive states in the relevant sense. With one caveat, this terminological point need not detain us

---

direction of fit, the earlier argument is incompatible with the claim that beliefs about what rationality demands can produce new desires without the help of antecedent desires.

<sup>15</sup> Some philosophers are not so willing to demur here. Sinhababu (2009, 465) describes the Humean as committed to the “Desire-Out? Desire In!” principle: “Desires can be changed as the conclusion of reasoning only if a desire is among the premises of reasoning.” For an interesting discussion of the various implications of different ways of spelling out the main Humean idea, see Melissa Barry (2010)

further, since I will be focusing on a version of motivational cognitivism according to which some desires may not be independent of the agent's cognitive states in the first, non-causal sense.<sup>16</sup> The caveat arises from the supposition that intentions are either a species of desire or are partly constituted by desire. If an agent forms an intention wholly on the basis of her cognitive states, we should not allow these "desires" to count as suitably-independent of those cognitive states; for if we did, the Humean Theory would follow trivially from the claim that intentions are necessary for motivated action.<sup>17</sup> Humeans should not be pleased if it turned out that the only desire that is in the neighborhood of some action is an intention that was itself produced as the result of wholly cognitive forces.

There is another potential condition that is relevant to this one. Hume ([1888], 416) thought that an agent's conative attitudes are open to rational criticism only to the extent that they rely on false beliefs or initiate actions that are known to leave them unfulfilled. When it comes to desires that are suitably informed by the facts, all bets are off:

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter. (Hume [1888], 416)

---

<sup>16</sup> In Chapter 5, I briefly discuss a view I characterize as a version of cognitivism according to which some desires are not suitably-independent in the causal sense. However, I argue there that this version of cognitivism does not provide a genuine alternative to the view I develop.

<sup>17</sup> A similar point is made by Mark van Roojen (2002, 209).

One might interpret “suitably independent” to entail this rational independence too. I prefer not to build this condition in to the definition of the Humean Theory since, once again, a notable contingent of self-described Humeans are open to the idea that some desires are rationally evaluable or even rationally required.<sup>18</sup> Moreover, this claim is deeply connected to the question of whether an agent’s fundamental or intrinsic desires are revisable wholly on the basis of her beliefs. Hume’s defense of rational independence is based entirely on his argument for the claim that desires *cannot* be revised wholly on the basis of reason.<sup>19</sup> As we indicated a moment ago, Smith (1994, 193) disagrees, and I am not interested in arguing about this here. So I will assume that rational independence is a distinct claim—one that might or might not fall out of the Humean’s central commitments.

#### **2.4 A Note on Terminology:**

Before moving on to the main argument, it is worth highlighting a couple of ways in which this view of the target differs from some other discussions of the Humean Theory. Philosophers frequently employ a number of concepts when talking about motivation, including “motive”, “motivate”, “reason”, “motivating reason”, “motivated action”, “feeling motivated”, “being (or feeling) moved”, “inclination”, and so on.

The thesis I will be discussing is spelled out in terms of an agent *being motivated*. In addition to this way of talking, I will, in general, only use a relatively narrow range of terms. I will often talk in terms of an agent being moved to do something, of something

---

<sup>18</sup> For a discussion of some of these interesting issues, see Williams (1980), Christine M. Korsgaard (1986), R. Jay Wallace (1990), Smith (1994), van Roojen (2002), Barry (2010).

<sup>19</sup> See Hume ([1888], 413-418).

motivating an agent to act, and of the conditions required for agents to engage in motivated action. I will not, in general, talk about motives, reasons, or motivating reasons. My preferences for avoiding these notions are based in the fact that these terms are often used in technical or quasi-technical ways, with stipulated senses; that their non-technical uses are embroiled in a number of controversial issues that I would prefer to avoid; and that the best understanding of the ordinary sense of these terms is plausibly given by first understanding the issues that interest me.

Many philosophers who are interested in these same issues do not share my concerns. For this reason, they often present their views as being about reasons or motives. As I mentioned moments ago, Smith (1994, 92), contends that the Humean belief-desire picture is ultimately an account of what constitutes a motivating reason. I am not sure if there is such a thing as the ordinary concept of a motivating reason. Perhaps it is a reason that motivates an agent to act in some way. But might not someone have a motivating reason to act in some way without actually being motivated to do so? Perhaps instead we might want to say that a motivating reason is something that is *capable* of motivating an agent to act (even if it does not actually do so). I have no clear intuitions about this. But in any event, it is plausible that the ordinary concept of a motivating reason (if there is one) is conceptually dependant on the concepts in which I am interested.

Several philosophers—Hume included—prefer to talk of motives. While I am more confident that there is an ordinary concept of a motive, I think that this too is dependant on the concepts that interest me. Motives are plausibly considerations that

could (or do) move an agent to act in a particular way. If so, the concept of being moved is prior to the concept of a motive.

### **3. Why We Cannot Ignore Motivational Cognitivism**

#### **3.1 Instrumental Reasoning: The Canonical Cases**

In this section, I argue that we need to take the truth of motivational cognitivism as a serious possibility. In order to defend her view as a necessary truth (whether conceptually, metaphysically, or nomologically), the Humean bears the burden of showing why cognitivism is untenable. While Humeans have offered a number of philosophical arguments in favor of their view, I think that many of these arguments are far less compelling when one fully appreciates how radical the Humean Theory is.

The argument takes as its starting point a form of practical reasoning that philosophers often take, either implicitly or explicitly, as central to our understanding of rational action. This canonical form of reasoning is embodied in situations like the following: Theresa wants to take her children to see the new *Harry Potter* movie on opening night. Knowing that advance tickets go on sale at noon today and are going to sell out quickly, she decides to purchase the tickets over her lunch hour. When the time comes, she goes on the movie theater's website and purchases the tickets.

This type of reasoning is instrumental in the sense that it involves reasoning about how to act so as to achieve some end. In its canonical form, such reasoning is commonly represented as an argument (or series of arguments), consisting of two types of "premises": motivational premises that determine the thing to be achieved and

informational premises that lay out the path by which that thing can be had.<sup>20</sup> The basic idea is that Theresa puts a motivational element (her desire) and a factual or informational element (her knowledge about what it would take to achieve her desire) together as a way of reasoning about what to do in something approximating the following “arguments”:

***Reasoning 1:***

**Desire:** I take my children to see the new *Harry Potter* movie on opening night. (*Motivational Premise*)

**Belief:** If I do not buy advance tickets relatively close to noon today, I will not be able to take my children to see the movie on opening night. (*Factual or Informational Premise*)

---

**∴Intention:** I will buy tickets over my lunch hour.

There is room to fuss about the details. Can all cases of practical reasoning be captured in this way? Is the conclusion of such reasoning really an intention? Or is it an action? Or is it something else? Do the belief and the desire serve as “premises”? Or is it the facts that one has the belief and that one has the desire that are the premises? These are no doubt important questions, but I will set them aside. Assuming that something in the ballpark of this story is correct for an interesting range of cases, I will focus on the question of what kind of attitudes can be represented as motivational premises in such a structure.

---

<sup>20</sup> This picture is often traced back to Aristotle (see, for instance, *Nicomachean Ethics* at 1147b). For further discussions of the so-called “practical syllogism”, see Elizabeth Anscombe (1957, 57-66), Donald Davidson (1970), several of the essays in Joseph Raz (ed.) (1978), Robert Audi (1989, 17-19; Ch. 4), and G.F. Schueler (1995, 97-108).

### 3.2 Instrumental Reasoning: Less Paradigmatic Cases

On the face of it, the instrumental reasoning occurring in the paradigmatic case is not so different from the reasoning embodied by less paradigmatic cases. Suppose that Pete is at a high-end retail boutique, looking at belts. After inspecting the price tag of a belt he is thinking about buying, he judges that it would be immoral to spend that kind of money on something so trifling while so many other people in the world are in dire need. Pete then puts back the belt and leaves the boutique empty-handed. On the face of it, it seems like Pete's reasoning can be represented by an argument that is quite similar to Theresa's:

***Reasoning 2:***

**[Moral] Belief:** I ought to avoid spending substantial amounts of money on trivial things if that money could be used by others to avoid great suffering.

**Belief:** If I purchase this belt, I would be spending substantial amounts of money on trivial things when others could use it to avoid great suffering.

---

**∴ Intention:** I will not purchase the belt.

The reasoning embodied in *Reasoning 1* bears important similarities to two more bits of reasoning as well. If Sasha thinks that she wants to go to earn an M.B.A. from a top business school, we can imagine her going through a bit of reasoning that can be reconstructed in the following argument:

***Reasoning 3:***

**Belief [about her desire]:** I want to go to a top M.B.A program.

**Belief:** My chances of getting into a top M.B.A program will be improved if I score high marks on the Graduate Record Exam.

**Belief:** Taking a prep course could help me improve my score on the GRE.

**∴ Intention:** I will take the GRE prep course I saw advertised on the bulletin board in the Advising office.

Or suppose that Jackie is sitting in her office one afternoon, drowsy and unable to concentrate. She thinks that she needs some caffeine. Jackie might then run through the following reasoning:<sup>21</sup>

***Reasoning 4:***

**Belief [about her Needs]:** I need some caffeine.

**Belief:** By going to the café on the first floor and purchasing a cup of coffee, I could secure some caffeine.

**∴ Intention:** I will go down and grab a cup.

The thoughts that run through Pete's, Sasha's, and Jackie's heads are similar to Theresa's reasoning in at least two important respects. First, they approximate arguments leading (in part) from thoughts about how one would promote some state of affairs, to an intention to act in a way that would do so. Second, all three pass through some further attitude that is part of the story of why it would be more-or-less reasonable for the agent to promote such a state of affairs in the first place. Given Pete's moral belief, it would

<sup>21</sup> For a similar example involving a felt need, see Audi (1989, 87).

make sense for him to put the belt back, given its price. Sasha's thoughts about what she wants seems to make it reasonable for her to be moved as she is by the thoughts about what it would take for her to get into a top M.B.A program. Jackie's intention to get some coffee is a reasonable response to her thoughts about what she needs. Theresa's desire, Pete's moral belief, Sasha's thoughts about what she wants, and Jackie's thoughts about what she needs all seem to play a similar role in this respect: all four seem to determine the (or an) end to be achieved. In other words, certain *beliefs* seem capable of playing the end-determining role that desires play in the canonical form of practical reasoning.

Before moving on to defend these intuitions, the following point is important. In the remainder of this chapter—indeed, the remainder of this dissertation—I will focus primarily on the case involving moral judgments. I have brought up Sasha's and Jackie's cases in order to highlight that the debate is much broader. Sasha's case raises the question of whether one's beliefs about one's desires are capable of motivating independently of her actual desires. We can, presumably, be wrong about what we want. If so, can false beliefs about what we want have motivational "oomph"?<sup>22</sup> If so, can true beliefs about our desires have "oomph" independently of their truthmakers? Jackie's case raises the issue of whether one's recognition of a need could move her to act without some independent desire. Must we have a desire to have our needs satisfied in order to be moved to satisfy them? Might it not be enough that we lack a self-destructive desire to have our needs remain unfulfilled?

---

<sup>22</sup> This issue is discussed by, among others, Smith (1994, 208 n. 6) and Schueler (2009).

Given that these questions are broader than the issue of whether or not moral judgments are capable of motivating, there is a way in which my focus is somewhat artificial. In fact, this artificiality is a deep part of the contemporary debate between cognitivists and Humeans because this debate is largely driven by philosophers who discuss this issue in the context of larger puzzles related to ethics.<sup>23</sup> So I will largely follow the focus of the literature. At the same time, the arguments on which I rely do not draw on any features that are unique to moral judgment, and for this reason the general approach I offer can easily be extended to other cases of cognitive motivation as well.

### **3.3 Reasoning**

Many cognitivists find it overwhelmingly plausible that agents can reason in the way that we just described in Section 3.2. More precisely, they find it compelling that the practical “inferences” articulated in Section 3.2 provide a relatively exhaustive characterization of the practical reasoning that is occurring in these cases, in the sense that the reconstructed arguments need not be missing relevant non-cognitive desires that serves as independent motivational premises.

We could try to defend this thought by appealing to lofty claims about reasons or rationality—say, by establishing that the capacity to reason in such ways is a precondition for rational agency or the ability to act for reasons.<sup>24</sup> While there certainly may be good arguments in this direction, there are at least three pragmatic reasons for looking for an

---

<sup>23</sup> See, for instance, David McNaughton (1988), Dancy (1993), Smith (1994), Scanlon (1998), and Shafer-Landau (2003).

<sup>24</sup> For arguments along these lines, see Sarah Buss (1999) and Christian Miller (2008).

alternative argumentative approach that does not rest on any controversial assumptions about reasons or rationality.

The first is purely dialectical: cognitivist arguments that rely on anything more than a thin, instrumentalist picture of rationality or agency are not likely to convince those with Humean sympathies.

Second, such arguments are necessarily incomplete. Even if some varieties of cognitive motivation can be vindicated in this way—say those involving some highly specific, all-things-considered practical “ought” or “reason” claim—they leave us no better able to understand other, equally plausible examples. It would be better if we had a more general argument that could also cover the vast panoply of potentially motivational cognitive states—including beliefs about what we *morally* ought to do, about what would be morally *quite good* (but not necessarily required) for us to do, about what would promote what we think we want, or about what would get us what we think we need.

Third, these arguments are simply unnecessary. The cognitivist needn’t think that the “inferences” described in the cases of interest would be rationally required (or even rationally acceptable) in order to establish that they are possible. Suppose that someone wanted to deny that people could engage in theoretical reasoning along the following lines:

***Reasoning 5:***

**Belief:** I have lost three straight hands.

---

**∴ Belief:** I will win the next hand.

*Reasoning 5* is clearly bad reasoning, but couldn't someone make such an inference anyway? That something would be bad reasoning does not undermine the fact that people could reason in that way. One could try to resist this by making the argument somewhat better. One could insist that, say, no one could reason in this way unless the transition were mediated by some further mental state—perhaps by a conditional belief with an antecedent that matches the content above the lines and a consequent that matches the content below the lines. If an individual did not have such a belief, our fictional objector may continue, she could not make this inference. But fully generalized, this strategy would commit us to the claim that people only reason in valid ways. This is clearly false. And it really doesn't help anyway—surely the conditional belief we have posited is itself irrational to have in most circumstances. We must be willing to attribute bad moves to reasoners at some point.

As I said, these are entirely pragmatic reasons for finding a more general argument for cognitivism that does not rest on controversial claims regarding reasons or rational agency. To develop such an argument, let us focus on the following question: What would it take to show that (human) agents are incapable of reasoning in the ways discussed in Section 3.2?

The answer to this question depends on what reasoning is. At an absolute minimum, reasoning—both theoretical and practical—involves transitioning from one set of mental states to another. Therefore, the Humean could proceed by trying to show that in the absence of a relevant non-cognitive desire, it would be impossible for an agent to

make the transitions from the belief states represented above the line to the intention represented below it.<sup>25</sup>

If the claim is that it is conceptually impossible for individuals to make these transitions, it is hard to see this as anything other than a non-starter. Even if the claim is that this is metaphysically necessary or a law of human psychology, then the Humean still faces a high burden. The philosophical literature on just about everything has pretty lax standards for admitting counter-possibilities to purported conceptually and metaphysically necessary truths. Even the most far-out of these possibilities are taken by serious philosophers to pose at least *prima facie* challenges to serious philosophical theses. The possibility of far-off planets that possess substances that are functionally, but not chemically, indistinguishable from water; the possibility of molecule-for-molecule doppelgangers materializing out of swamp gas; and the possibility of nearly omniscient deities living on mountaintops must be taken seriously by theories regarding the contents of thought and language. Reliable clairvoyants, brains in vats, trickster demons, and special brain anomalies are all taken by epistemologists to pose serious challenges for certain views regarding knowledge or epistemic justification. Cases involving people seeds, Henry Fonda's curative touch, drugs that make for kittens with human intelligence, and forced choices between living the life of a clam and living the life of a great artist are the means by which philosophers advance debates in applied and normative ethics. Fantastical counter-possibilities are the common coin of philosophical trade.

With respect to the purported conceptual or metaphysical truths, Humeans must have a principled reason for ruling out the possibility that even agents who (by

---

<sup>25</sup> This way of thinking about the issues is borrowed (in part) from John Gibbons (2009).

stipulation) have atypical psychologies could make the transitions in question. Suppose we offer a new version of Pete's case: suppose Pete has a brain anomaly such that whenever he has beliefs that entail that some particular action would be morally wrong, he automatically forms the intention to avoid that action if he considers whether or not to do it. Suppose further that, as Pete is standing at the boutique thinking about the belt and its price, he forms the beliefs that entail that buying it would be wrong. As the result of these beliefs and his brain anomaly, Pete forms the intention not to buy the belt and subsequently puts it back. He does this (we stipulate) without the presence of any distinct, causally-relevant desire appropriately related to this intention. In this version of the case, Pete puts the belt back as the result of forming an intention not to buy it, which was itself the result of the beliefs and the brain anomaly, but not any distinct, appropriately-related desire. Surely this is a conceptually coherent possibility. And we should take claims that it is not metaphysically possible with a healthy dose of skepticism.

The reason that Humeans tend to be unmoved by putative counterexamples to their view is that it is very easy to posit hidden desires. In the version of the case discussed in Section 3.2, if Pete was moved by his moral judgment to put the belt back, it is easy to say that he had a hidden desire—say, a desire to avoid wrongdoing. Though cognitivists think that this move is ad hoc, there is not much more that they can say; the Humean sincerely disputes the substance of that charge. Can Humeans respond to our new version of the case in the same way? One might be tempted by the thought that this brain anomaly would, in effect, give rise to a type of desire—a desire to avoid wrongdoing. Here, the criticism of ad hocness has more bite. It must be admitted that

someone could argue for such an interpretation; however, they would need to *argue* for this.

Remember, we are wondering here whether it is conceptually or metaphysically possible for agents lacking desire to transition from one set of mental states to another. If the Humean wants to avoid the counterexample by insisting that Pete's brain anomaly gives rise to a desire, they must defend this as a necessary claim too. However, there is good reason to resist this. We can stipulate that this particular anomaly is distinct from the structures and processes in the brain that undergird typical human motivation. We can also stipulate that it has a different etiology. We can even stipulate that this is a completely novel intention-forming process that is entirely unique to Pete. If we can stipulate all of this, the Humean must explain to us in a non-question-begging way why this anomaly would involve desire in any normal sense.<sup>26</sup>

It is good practice to insist that unless we have good reason to rule out the possibility that agents could make certain mental transitions, we should take the purported possibility seriously. This argument does not depend on any controversial assumptions about reasons or rationality or anything else the Humean is likely to reject. It relies only on a general methodological principle to which Humeans should have no

---

<sup>26</sup> This version of Pete's case gives us good reason to reflect on whether the Humean's claim has been subject to sufficient scrutiny—especially when this claim is forwarded as a conceptual or metaphysically necessary truth. This much is true even if one believes that the Humean story is approximately correct when it comes to motivation in typical agents. Yet even this more modest quasi-Humean claim should be open to critical examination. Instead of a novel anomaly in Pete's brain, couldn't he have inherited some brain structure that plays this role—a bundle of neurons that was passed down from his ancient progenitors? Couldn't the vast majority of Peter's contemporaries possess a similar trait? Couldn't *we*? Couldn't there be multiple neurological pathways that lead to motivated action? If Pete's case is a possibility, then I see no reason why it couldn't provide a range of intermediate cases that make salient the possibility that some significant range of human motivation is actually the product of wholly rational forces.

complaints. In fact, Svavarsdóttir (1999, 179) deploys basically the same methodological principle in support of the Humean Theory:

But when there is a conflict of intuitions (among intelligent and sensible people) about which hypotheses are in the running as an explanation of some observable phenomenon, the burden of argument is on those who insist on a more restrictive class of explanations. This seems to me entirely reasonable as a methodological principle governing empirical explanation.

Given this burden, few philosophers seem to appreciate how difficult it is to show that it is impossible for agents to make these transitions without the existence of a non-cognitive desire. Yet showing that these transitions are impossible is just one way that the Humean might go about arguing that agents cannot reason in the ways suggested in Section 3.2. Not all transitions between mental states constitute reasoning. One can mindlessly watch the scenery change through a train window, overhear nearby conversations, dream, and fantasize a drawn-out scene. Or even more radically, a subject could be hooked up to a remote control device that manipulates her thoughts and actions from one moment to the next. These phenomena also involve moving from one set of mental states to another, yet the transitions involved in these cases do not always involve reasoning in any obvious way.

It may then be open for the Humean to argue that even if it is possible for subjects to make the transitions mentioned in Section 3.2, the cases in which this happens would not count as cases of practical reasoning. It may be open for the Humean to claim that if Pete forms the intention as a result of the brain abnormality, he would not be arriving at it by reasoning.

This objection strikes me as too quick. In order for the objection to succeed, the Humean needs this to be a necessary claim. This too stands in need of defense. Even if

Pete forms the intention partly as the result of a brain abnormality, he might still be reasoning—at least potentially. We should at least allow that brain anomalies can sometimes give rise to anomalous forms of reasoning. But anomalous forms of reasoning are still forms of reasoning. Consider a non-practical case. If a particular neurological anomaly enables an individual to quickly form any mathematical belief entailed by other mathematical beliefs she holds (at least when she asks herself whether the mathematical claim in question is true), we would not automatically insist that this new belief was not formed by reasoning. It is at least potentially formed by reasoning. (Indeed, we may even consider such an agent to be a superior mathematical reasoner when compared to the rest of us.) Once again, the Humean faces a high burden here. She must tell us which mental transitions count as reasoning and which do not. Her account must make it clear (in a non-question-begging way) why transitions mediated by desire qualify as reasoning while Pete's anomaly and every other possible analogue would disqualify his mental transitions from being so counted. Humean assurances that this is the case should carry no weight unless they can be supported with solid evidence.

#### **4. Things to Come**

With this argument in favor of taking cognitivism seriously in hand, much of the rest of this dissertation examines the evidence on which Humeans have traditionally relied. It may be that the case we just described as conceptually possible is actually incoherent for some subtle reason, and that the *a priori* arguments offered by the Humean can bring this incoherence out from the shadows. Or, it could turn out that some of the philosophical arguments that incorporate *a posteriori* considerations show that this conceptual

possibility is metaphysically impossible or contrary to the laws of nature. If successful, these arguments would help the Humean meet the burden I have put to her. After carefully examining these Humean arguments, however, I contend that they are all wanting.

In Chapter 2, I examine an argument from motivational externalism. In support of their claim that moral judgments are incapable of motivating agents directly, Humeans frequently attack a range of so-called “internalist” claims, to the effect that all (or at least some interesting range of) moral beliefs necessarily motivate agents to act accordingly. I agree with the Humean that such strong versions of motivational internalism are likely false, as standard cases involving seriously depressed agents and amoralists provide compelling counterexamples to them. However, I argue, this does not undermine motivational cognitivism. I examine cases that are significantly like the externalist cases and explain why they undermine similarly strong claims about the essentially-motivational nature of desire. This is dialectically significant, since few philosophers have a problem accepting that desires are capable of motivating agents to act. If desires can be motivational without being necessarily so, then it is no objection to the motivational efficacy of belief to show that they are at best contingently motivational too.

If the argument of Chapter 2 is successful, it suggests that the best version of motivational cognitivism will allow for contingently motivating attitudes. Chapter 3 examines some of the same issues from a more theoretical perspective. In this chapter, I take up a standard dispositional account of desire and argue that, properly understood, this account of desire supports the claim that desires motivate only contingently. In order to show this, I do two things. First, I try to motivate a dispositionalist necessary

condition on being motivated to act. Second, I develop an extended analogy between desires and commonplace dispositions—like fragility, solubility, and the disposition to accelerate at a particular rate in free-fall—surrounding the way in which they vary in degree. I contend that this analogy matters for how we should understand the basic shape of the dispositionalist account. According to any dispositionalist account of desire with that shape, desires will meet the necessary condition on motivation at best contingently. I then try to use this apparatus to provide a broad-brush account of motivational variability that does not appeal to differences in the contents of desire.

In Chapter 4, I build on the basic dispositionalist picture of Chapter 3 and use it to argue that a key dispute between Humeans and cognitivists can be understood as an empirical issue. I begin by motivating a more precise interpretation of the basic dispositionalist picture of Chapter 3 and suggest that the best way to make sense of this interpretation is to see desire as a natural kind. With this in mind, it is best to understand motivational cognitivism as an empirical hypothesis about the role this natural kind plays in human motivation. I then respond to armchair attempts to resist this hypothesis. Following Thomas Nagel (1970), among others, I contend that such arguments trade on an ambiguity in the folk concept of *desire*. I then argue that a structurally similar argument can be deployed against some competing versions of motivational cognitivism. In particular, I show that an analogous argument undermines versions of cognitivism that purport to establish a conceptually necessary claim between intentional action and cognitive representations of the good.

The picture that is developed in Chapter 4 suggests that some cognitive states are similar to desires in a significant respect—in fact there is a perfectly good sense of the

word “desire” in which it is fine to say that these cognitive states *are* desires. In Chapter 5, I turn to a widely discussed argument to the effect that such attitudes (they are often pejoratively called “besires”) would be ontologically extravagant and possibly even incoherent. I argue that there is nothing incoherent or bizarre about these besires. The argument against besires relies on the thought that beliefs and desires have essentially different “directions of fit”. First, I argue that this argument fails on its own terms. I argue that on the best extant accounts of what these two directions of fit amount to, it should not be shocking if certain attitudes ended up possessing both. In fact, I contend, it is plausible to think that moral judgments really are besires in the relevant sense. Second, I argue that the argument against besires would prove too much—it would rule out by *a priori* fiat a plausible empirical hypothesis about the nature of certain paradigmatic motivational states. More importantly, this empirical hypothesis is unrelated to the typical dividing lines between Humeans and cognitivists regarding moral motivation.

## Chapter 2

### Depression, Amoralism, and the Humean Theory

#### 1. Introduction

In Chapter 1, I argued that we need to take seriously the possibility of wholly cognitive motivation. In this chapter, I address an intuitive argument against the idea that motivation can proceed directly from an agent's moral judgments. The objection is, in effect, that this cognitivist proposal overstates the connection between moral judgments and motivation. To a first approximation, the argument is that the possibility of certain agents—severely depressed agents and amoralists—show that moral beliefs do not necessarily motivate agents to act appropriately. Moral beliefs, therefore, cannot be effective sources of motivation—something extra is needed to pick up the slack.

Insofar as moral judgments are sometimes thought to be the cognitivist's leading example of motivational cognitive states, this argument is meant to make a presumptive case in favor of the Humean Theory.<sup>1</sup> Moreover, the supporting reasons generalize. Analogous cases can be constructed to cover *any* purported motivationally-efficacious cognitive state. This “something extra”, the argument continues, must therefore be a *non-cognitive* state—some desire or passion. Motivation cannot be the product of purely cognitive forces.

The motivational cognitivist has two standard lines of response. The first is to reject the problematic cases by insisting that, as a matter of necessity, the depressed and

---

<sup>1</sup> Many cognitivists would take issue with this; these cognitivists instead think that beliefs about what one all-things-considered ought to do that are the best bet for motivational beliefs. See, for instance Sarah Buss (1999) and Ralph Wedgwood (2007, Ch. 1).

amoral either fail to make moral judgments (or fail to possess some other relevant cognitive state) or are moved (in some sense) to act. The second response is to argue that these problem cases describe instances of practical irrationality and thus pose no significant problem for motivational cognitivism—it would suffice for the truth of cognitivism if even only fully rational agents were moved by wholly cognitive forces.

In this essay, I provide an alternative way to resist the Humean argument. I start by discussing the problem cases. I suggest that these cases provide compelling evidence that moral beliefs do not necessarily motivate agents to act accordingly. In Sections 3 and 4, I respond to two versions of the argument against motivational cognitivism based on these cases. I contend that both versions depend on questionable claims about the motivational efficacy of desire. In fact, these claims are undermined by the very same considerations that support the externalist upshot of the motivational cognitivist's problem cases. The Humean argument is itself unstable. In Section 5, I turn my attention to an alternative strategy for using depression and amorality in support of the Humean Theory—one that appeals to the Humean's explanatory superiority. I argue that this strategy fails as well, and for much the same reason as the other two versions. Since desires are not necessarily connected with motivation in the requisite ways, Humean explanations are really no deeper than those offered by the motivational cognitivist. This is significant, in that it upsets a common presumption that often underlies support of the Humean Theory—the presumption that Humean explanations are easier to come by than anti-Humean explanations.

## **2. The Externalist Cases**

## 2.1 Internalism: A First Pass

It is easy to overstate the connection between our moral judgments and motivation. Take, for instance, the claim with which R.M. Hare (1964, p. 1, original italics) opens *The Language of Morals*: “If we were to ask of a person ‘What are his moral principles?’ the way in which we could be most sure of a true answer would be studying what he *did*.” If Hare is here endorsing the following version of internalism, he is clearly wrong:

***Judgment/Action Internalism:***

Necessarily, if an agent judges that an action,  $\phi$ , is morally right (best, obligatory, etc.) she will (try to)  $\phi$ ; if she judges that  $\phi$ -ing is morally wrong (inferior, impermissible, etc.) she will (try to) not-  $\phi$ .

*Judgment/Action Internalism* is incompatible with the possibility that someone could be convinced that eating meat is wrong or that she is morally required to donate money for famine relief, while her contrary desires get the better of her. Yet being moved by one’s moral judgments is clearly compatible with being moved more by other considerations.

Since our weak-willed agent may be motivated in the sense that she still sees or feels “the pull” of morality’s demands while failing to act in the appropriate ways—since she may simply be moved more by the prospect of prime rib or the conveniences that extra money provides—the possibility of competing motives is compatible with a weaker internalist thesis:

***Judgment/Motivation Internalism:***

Necessarily, if an agent judges that an action,  $\phi$ , is morally right (best, obligatory, etc.) she will be motivated to  $\phi$ ; if she judges that  $\phi$ -ing is morally wrong (inferior, impermissible, etc.) she will be motivated to not- $\phi$ .

In this section, I discuss two standard counterexamples to *Judgment/Motivation Internalism*: those involving depressed or listless individuals and amoralists. I argue that these cases really do provide trouble for the view.

## 2.2 Depression

Let us start with the case of severe depression. Several philosophers have tried to use this phenomenon to drive a wedge between moral judgments and motivation. According to these philosophers, while in the throes of a deep depression, an individual might continue to make moral and evaluative judgments while failing to be motivated. Michael Stocker (1979, 744), for example, describes things as follows:

Through spiritual or physical tiredness, through accidie, through weakness of body, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that there is less good to be obtained or produced [...]. Indeed, a frequent added defect of being in such "depressions" is that one sees all the good to be won or saved and one lacks the will, interest, desire, or strength.

Some might object to this description of the phenomenon. According to this response, if we attend to the fact that an individual can be motivated to act in a particular way without trying to do so (as happens in the case of competing motives) then it is not actually clear whether or not the depressed individual is motivated. In other words, according to this objection, the motivational failure associated with severe depression is a

counterexample to *Judgment/Action Internalism*, but not *Judgment/Motivation Internalism*.<sup>2</sup>

This response underestimates the force of some of the putative counterexamples. Alfred Mele (1996) offers one such case that merits some discussion. Mele's case involves Planet X, some of the inhabitants of which have rich moral lives, much like we do: Xians employ ethical terms, judge their own actions and the actions of others, have reactive attitudes similar to our attitudes of guilt, regret, blame, praise, and so on.

Mele asks us to suppose three further facts about Xian psychology. The first is that Humean psychology is true of Xians—an Xian will be motivated to act in a particular way only if she has a suitably-related desire. When it comes to moral motivation, we are further asked to suppose that “The great majority of adults on planet X are possessed of a long-term generic desire to do whatever is morally required [and] whenever Xians seek to satisfy what they take to be moral requirements, they are motivated by [this generic desire]” (Mele 1996, 733). The third fact is that Xians are sometimes subject to severe depression or listlessness that “In its more severe forms, [...] consists in the total absence of motivation to engage in activities of kinds that formerly were matters of deep personal concern” (Mele 1996, 734).

To complete the case against *Judgment/Motivation Internalism*, Mele (1996, 734) describes the plight of a particular Xian:

In a recent case, owing to the deaths of her husband and children in a plane crash, a woman, Eve, lost all motivation to continue aiding her ailing uncle [...]. Eventually, Eve sought treatment for her depression. Standard medical procedure on planet X in such cases requires that agents take an electronic lie detector test and then answer questions under the

---

<sup>2</sup> Eve Garrard and David McNaughton (1998) suggest a response partly along these lines.

influence of a truth serum. Both tests indicated that Eve believed herself to be morally required to assist her uncle. Another “belief test” was administered—a test for a neural realization of the pertinent belief. That test, too, indicated that Eve believed herself to be morally required to assist him.

While Mele’s discussion relies on imaginary results of neurological tests, this aspect strikes me as unnecessary. If we think about how things might seem from Eve’s point of view, we already have the makings of a compelling case against *Judgment/Motivation Internalism*; the appeal to Xian science adds nothing. From her point of view it could certainly seem to her like she has moral judgments that fail to move her, in even the minimal sense. If she thought so prior to the death of her husband and children, Eve could certainly take herself to be morally obligated to assist her ailing uncle. Nothing about this obligation has changed as a result of her recent tragedy—at least not from her point of view. Insofar as her take on her own situation is coherent, we should take it at face value. Her self-conception provides strong *prima facie* grounds for thinking that she is not appropriately moved by these moral considerations. There is clearly good reason to think that she could be as she sincerely takes herself to be. This possibility—the possibility that she makes moral judgments that fail to move her appropriately—is incompatible with *Judgment/Motivation Internalism*.

### **2.3 Amoralism**

Cases like this work so well, in part, because many of us find it easy to identify with individuals like Eve. We have at some point in our lives felt at least a modest melancholy and the decreased motivation that goes along with it. We can thereby appreciate what it might be like for someone to be in Eve’s position. Yet the perspective

of the amoralist—individuals who show a callous disregard for moral considerations—can be more difficult to appreciate. It is not straightforwardly absurd to think that there is something illicit in bald descriptions of amoralism, such as this anecdote offered by Nick Zangwill (2008, 102):<sup>3</sup>

A mercenary I once met on vacation exuded moral indifference. He was in control, reflective and articulate. Everything he said convinced me that he was perfectly aware that his vocation was genuinely morally wrong.... He fully understood the wrongness of his vocation. But he was not very concerned about that. He was more concerned with his immediate interests and concerns, that is, colloquially, looking after number one. There was no moral cognitive lack. He made that quite clear. Indeed he insisted on it. The mercenary was unusually indifferent to the demands of morality; but he shared moral beliefs with the rest of us....

There is, to be sure, no problem in thinking of such cases as counterexamples to *Judgment/Action Internalism*, for we can clearly understand how “looking after number one” might get in the way of *doing* what we know is right. Yet when confronted by such bare descriptions as counterexamples to *Judgment/Motivation Internalism*, there is a tendency to think that something fishy is going on.<sup>4</sup> There is a tendency to think that such characters, if they are genuinely unmotivated, are not really making moral judgments at all—to think that they don’t really mean it or don’t really get it. Zangwill’s

---

<sup>3</sup> For another classic discussion of amoralism, see David Brink (1986). Russ Shafer-Landau (2003, 148-155) also offers an excellent discussion of several other examples.

<sup>4</sup> In fairness, Zangwill doesn’t describe this as a case of complete amoralism—cases that he thinks are dialectically problematic. His immediate target is thus not *Judgment/Motivation Internalism*, but a different claim:

***Proportional Determination Thesis:***

The degree of a person’s moral belief that he ought to do something proportionately determines the strength of his desire to do it. (Zangwill 2008, 95)

For my dialectical purposes, the example works well enough.

mercenary, according to this objection, does not really judge his vocation to be morally wrong; he is merely parroting the moral norms of his wider community.<sup>5</sup>

Another objection to the possibility of amoralism has been offered by Michael Smith (1994, 68-71). Smith argues that even if the amoralist is sincerely trying to engage in moral discourse and to make moral judgments (rather than just parroting the moral norms of the wider community, or whatever) he fails to do this because he lacks the relevant moral concepts. Or, at the very least, he fails to master them, and they play a radically different role in his overall psychology, than they do in ours. In typical moral agents, moral thoughts are related to behavior in ways that go well beyond verbal and quasi-linguistic mental activities. Our moral concepts tend to be hooked up to our emotions, motivations, and intentional actions. Understanding such connections is necessary for complete mastery of the moral concepts and for genuinely employing them in moral judgments. If so, then the amoralist cannot provide a counterexample to *Judgment/Motivation Internalism*.

We can, however, construct cases that meet these deeper challenges head-on. I have in mind two sorts of cases. The first is modeled on the temporal structure of Eve's case—the individual is a typical moral agent through some portion of her life and then undergoes a transformation. Consider Kate. Through her early thirties, Kate seemed to be a perfectly normal moral agent. She seemed to make moral judgments, just as you or I would. She often acted as she thought morality required and sometimes she didn't; though when she didn't, she would typically still “feel the pull” of morality, and would

---

<sup>5</sup> This strategy is commonly attributed to Hare (1964, 124-126). Sigrún Svavarsdóttir (1999, 188) insightfully points out that this objection just won't do; for, the amoralist could be articulating quite *revisionary* moral standards, rather than those of her community.

often feel guilty and regret her decision. At thirty-five, Kate underwent some sort of psychic conversion. As she describes it, she simply stopped “feeling the pull” of morality. If we need some explanation for this, we might imagine that she left a religious cult or that she has been convinced of a philosophical theory according to which there is no good reason to be moral. In any event, Kate contends that she is no longer moved by her moral judgments. Her change in behavior lends credence to this claim. When she performs some morally good action, she will unashamedly avow that it was some other consideration that moved her to do it. At least this is how things seem to her.

We can imagine that after her conversion, Kate still seems to engage in moral discourse with others. She makes the same sorts of inferences involving moral terms, and takes the same sorts of considerations to bear on moral facts that she did before her transformation. Furthermore, she still readily ascribes certain moral beliefs to herself. Indeed she sometimes says things like “I have always been convinced that abortion is morally impermissible” and “I still think that we have no obligations to the poor”. In short, Kate is convinced that none of the contents of her moral beliefs have changed as a result of her motivational transformation. As far as Kate is concerned, the only difference is that these moral thoughts no longer move her as they once did.

So described, it is plausible that Kate has moral concepts and makes moral judgments employing them. This is clearly true of the first part of her life and there is no good reason to deny that she retains those concepts after her transformation. Nor should we deny that she remains competent at employing these concepts to make moral judgments. If we would take her moral claims and self-ascriptions at face value before

her motivational transformation (as we certainly should) then it is objectionably *ad hoc* to insist that we should not do so afterward as well.

Kate's case, like Eve's, highlights diachronic differences in agents' moral motivation. Cases involving synchronic differences may also help to disarm the deeper worries. One such case involves what we might call a "subject-matter amoralist"—an individual who is systematically unmoved by some proper subset of his moral judgments pertaining to some specific subject matter.

Let us suppose that Frank seems to be a sophisticated moral agent. He has sophisticated grasp of the moral language, employing it to make claims like "stealing is wrong", "if we have obligations to help the poor, then I should mail a check to Oxfam" and "allowing harm is at least as bad as causing harm." We may even suppose that he claims to subscribe to some well-developed normative theory. Frank generally acts as one would expect, given the sorts of claims that he makes: when he says that being honest is morally required or that his normative theory entails that cheating is wrong he typically tries to be honest and tries to avoid cheating.

Frank has a peculiar psychological quirk that manifests itself in a variety of ways. When a cat darts in front of his car Frank does nothing to avoid hitting it. He doesn't even try. Nor does he try to restrain his dog from chasing after neighborhood cats, or try to save them when they are stuck up in trees. It is not that Frank tries to cause wanton or malicious harm to felines; it is just that he is not at all motivated to prevent harm from coming to them, even when doing so would be quite easy. When cats are involved, his moral judgments get no traction.

Frank insists that he knows this is wrong. This verdict is entailed by his normative theory. He insists that he knows that cats are just as worthy of moral consideration as other animals (animals he would be moved to aid in a wide variety of cases). Frank will report that he thinks that he should be more willing to help cats, and that he is just as confident in this judgment as he is that he shouldn't lie, cheat, or eat veal. But he also claims to feel absolutely no compunction about letting cats be harmed; he clearly shows no signs of remorse afterward.

Except for those pertaining to cats, Frank's attitudes seem to be just like our moral judgments. These attitudes are hooked up in similar ways to his emotions and motivations, and he would take the same sorts of considerations to be evidence for and against them. These attitudes seem to be *bona fide* moral judgments, and we should thus attribute genuinely moral concepts to Frank. If we accept this much, we should accept that his cat-related attitudes are also moral judgments, employing those same moral concepts. As we have said, Frank has a sophisticated grasp of moral language—in language and thought he is able to use constructions like “If running over a dog is wrong, then so is running over a cat” and “Harming a cat is worse than harming an insect, but not worse than harming a child.” The very possibility of having thoughts like these seems to presuppose that one and the same concept can be deployed for both cat-related and cat-unrelated purposes.

Another related point also drives to this conclusion. We can imagine that Frank came to his cat-related attitudes on the basis of reflective equilibrium involving his various non-moral beliefs, his other judgments, and his normative theory. If others were to come to genuine moral judgments by this process, then Frank should be able to as well.

For others, moreover, this process would plausibly yield knowledge. Yet if Frank's cat-regarding attitudes fail to utilize the same moral concepts, then this process should not yield him knowledge. For, we must accuse him of equivocating between his genuinely moral concepts and his pseudo-moral concepts in a way that would make his inferences inappropriate. But whether or not he is moved by these beliefs, Frank clearly knows better.

The cases of Kate and Frank help meet the deeper concerns about amoralism head-on. If these cases describe genuine possibilities, as they seem to, then agents can have moral judgments that fail to move them. *Judgment/Motivation Internalism* is false.

### 3. Motivational Cognitivism and Depression

According to motivational cognitivism, motivated action can sometimes be the product of purely cognitive forces—there need not always be a distinct desire. If motivational cognitivists were committed to *Judgment/Motivation Internalism* then the cases discussed in the previous section would directly undermine the position. In this section and the next, I will explain why these cases pose no problem for the view.<sup>6</sup>

According to Michael Smith's (1994, 119-120), version of the Humean argument, motivational cognitivists are committed to the existence of "besires"—propositional

---

<sup>6</sup> There are other ways to resist the Humean conclusion. One common response is that even if moral judgments do not guarantee motivation, other cognitive states might—perhaps judgments about what an agent has reason to do (Garrard and McNaughton 1998), appearances of the good (Tenenbaum 2007), or all-things-considered "ought" judgments. The response I provide here is more similar to one offered by Russ Shafer-Landau (2003, 147-148)—that even if the motivational cognitivist thinks that moral judgments are intrinsically motivating, she need not insist that they are necessarily motivating. In effect, the argument I provide here fills out certain details missing from Shafer-Landau's discussion.

attitudes that are both belief-like and desire-like.<sup>7</sup> Cognitive states like belief have a mind-to-world “direction of fit”.<sup>8</sup> Such states are, roughly, ones that are to be changed when they fail to match the world. This is also true of the cognitive states that the motivational cognitivist takes to be motivationally efficacious—inaccurate moral beliefs, for instance, should be made to fit the moral landscape, not the other way around. Smith (1994, 116), however, contends that motivationally efficacious states have a world-to-mind “direction of fit”—they make the world conform to them, rather than vice versa. Motivationally efficacious cognitive states would thus have both directions of fit—they would be besires.

The Humean insists that there are no besires, but only concatenations of modally separable beliefs and desires—any belief-like state “can always be pulled apart, at least modally” from any desire-like state (Smith 1994, 119). If motivational cognitivism suggests that moral judgments are besires, then the belief-like and desire-like elements of moral judgments should not be able to “come apart” in this way. In other words, it seems that whenever an agent judges that she morally ought to  $\Phi$  (or, that  $\Phi$ -ing is right or best or whatever), she should thereby be motivated to  $\Phi$  (Smith 1994, 120). This is simply a restatement of *Judgment/Motivation Internalism*.

I am willing to suppose, for the sake of argument, that motivational cognitivists are committed to the idea that moral judgments (or some other cognitive states) are partly “desire-like”. Even granting this much, Smith’s argument still fails to saddle the

---

<sup>7</sup> The term “besire” was originally coined by J.E.J Altham (1986).

<sup>8</sup> The direction of fit metaphor was introduced by G.E.M. Anscombe (1957, 56). For critical discussions of the metaphor, see Mark Platts (1979) and David Sobel and David Copp (2001).

motivational cognitivist with *Judgment/Motivation Internalism*. The motivational cognitivist should presumably think that the motivational efficacy of a besire is inherited from its desire-like component. Without a necessary connection between this desire-like component and motivation, there is no reason to think that moral judgments (conceived of as besires) necessarily motivate, and hence no reason for the motivational cognitivist to accept *Judgment/Motivation Internalism*. We should not insist that besires necessarily motivate unless a similarly strong connection holds between *desires* and motivation. Here's one candidate for such a connection:

***Desire/Motivation Internalism:***

Necessarily, if an agent desires that *p*, she will be at least somewhat motivated to bring it about that *p*.

According to Jonathan Dancy (1993, 2), Humeans are committed to *Desire/Motivation Internalism*.<sup>9</sup> If so, it is easy to understand why they would press this argument against motivational cognitivism—if the desire-like components of moral judgments are really like desires, then they would necessarily motivate too. Yet *Desire/Motivation Internalism* is simply implausible. It is widely recognized that many desires do not typically motivate the agent to bring about their contents. To fix on some examples, I might desire that the snow clears out by 6:00 PM or that scientists figure out

---

<sup>9</sup> He claims that Humeans are committed to desires being “states which are *guaranteed* to motivate; they cannot exist without motivating. We can say that they are essentially or necessarily motivating states” (Dancy 1993, 2) [Dancy’s emphasis]. I am skeptical that Dancy is right about this, but this does not matter for my purposes.

human teleportation within the decade without being motivated to bring about either state of affairs.<sup>10</sup>

Even if *Desire/Motivation Internalism* is false, a weaker principle might work well enough. Standard examples of motivationless desires—including the two we have just mentioned—do not have the agent's acting in any way as their objects. There is no action *I* want to perform in virtue of wanting scientists to discover teleportation. If I want the snow to clear out by 6:00 PM, I do not want to *do* anything to try to bring this about (there is nothing I can do). According to Mele (1995), we can escape this sort of objection if we focus on action-desires—desires directed toward the agent acting in some way. According to Mele (1995, 399), “Any desire to *A* (*A* being an action variable), by its very nature, inclines the agent, in some measure, to *A* intentionally, or to try to *A*, or to try to put herself in a position to *A*.” According to Mele, these action-desires are “essentially motivation-constituting states” (“*EMCs*”, for short).

Moral judgments often have the agent acting in some way as an object in a similar way: my belief that I ought to  $\Phi$ , for instance, is in some appropriate sense about my  $\Phi$ -ing. If such judgments have a desire-like component, the Humean might argue, that component would have to be action-desire-like. On the assumption that Mele's picture is correct, if these states were desires, they too should be *EMCs*. The argument may thus still be salvageable if the following is true:<sup>11</sup>

---

<sup>10</sup> Mele (1995, 393 & 395) uses examples of someone who desires world peace or someone who wants the New York Giants to win some important game.

<sup>11</sup> It should be noted that Mele does not use his picture to defend the Humean Theory. While he thinks that moral beliefs are not *EMCs*, Mele leaves open the question of whether such judgments are contingently motivating states. He does, however, accept the claim that motivation is *constituted* by action-desires, and therefore one cannot have motivation without a desire.

***Action-Desire/Motivation Internalism:***

Necessarily, if an agent desires to perform some action,  $\phi$  (has some desire-like state to the effect that she  $\phi$ s), she will be at least somewhat motivated to (try to)  $\phi$ .

Certain counterexamples to *Desire/Motivation Internalism* seem to be equally problematic for *Action-Desire/Motivation Internalism*. Marty may want to travel back in time. However, if he is convinced that this is impossible, he may fail to be appropriately motivated.

Mele (1995, 399) recognizes this potential objection and tries to avoid it by insisting that *EMCs* must have as their objects actions that are “doxastically open to the agent: if she does not explicitly believe that she can *A*, at least she does not explicitly believe that she cannot *A*.” In support of this claim, Mele (1995, 399) insists that this condition must be met if the attitude is to count as a desire to do something: “If I am convinced that I cannot travel faster than the speed of light, or change the past, or defeat the current heavyweight champion, then although I might wish that I could do these things, I do not *desire to do* them” (Mele’s emphasis).

This is unsatisfying for the following reason. Suppose that Marty initially thinks that it is possible to travel back in time but that the conditions that must obtain are highly improbable. This is enough for traveling back in time to be “doxastically open” for Marty. If so, this is enough, on Mele’s view, for Marty to want to (and therefore to be motivated to) travel back in time. If these conditions then became even less probable by some arbitrary amount so that Marty becomes convinced that they do not obtain, Marty *automatically* goes from having a desire with one content (i.e. *to* travel back in time) to a desire with a different content (i.e. *that he could* travel back in time). This slight change

in credence *entails* a change in the content of his desires. But that this should happen as a matter of necessity is implausible.<sup>12</sup>

We can safely set these problems aside, since they are not directly relevant to the spirit of the Humean argument—cases of depression and amorality are not (or, at least need not be) cases in which the agent thinks herself unable to satisfy the relevant moral requirements. Even so, the motivational failure associated with severe depression does nothing to undermine the motivational cognitivist’s claim that moral judgments (or some other cognitive states) are desires. If cases of depression can be used to undermine *Judgment/Motivation Internalism*, they can also be used to undermine the key thesis on which the argument depends: *Action-Desire/Motivation Internalism*.

Suppose that George was just denied tenure. This event triggers a period of severe depression, during which George seems to lose his motivation to do many things, spending weeks brooding around his house feeling sorry for himself. Despite this apparent lack of motivation, there are many things that George says that he wants to do. For example, he claims that he wants to finish his manuscript or work on his garden—or even just to leave the house. In his current state, George lacks all motivation to do any of this—at least this is how he would describe things. This lack of motivation is incredibly frustrating to George and exacerbates his feelings of depression and worthlessness.

George’s case brings out a subtle point in Stocker’s and Mele’s discussion of how severe depression undermines motivation. Stocker described the depressive as having “lessened desire” and Mele explicitly discussed Xian listlessness as a problem of

---

<sup>12</sup> What this shows, I think, is that it is not enough for the action to be “open” in some abstract sense—what Mele should be insisting on here is that the agent has some idea about how to go about doing the thing she wants to do. This would bring it more in line with some formulations of the Humean Theory, such as the one offered by Smith (1994, Ch. 4).

weakened or absent desires. George's self-understanding, however, is that his desires remain intact while failing to motivate him. Moreover, these desires are action-desires—they are desires that have as their objects particular actions, such as *finishing his manuscript*, *working on his garden*, and *leaving the house*. We may also suppose that these actions are “doxastically open” to George—that he is not convinced that he would fail if he tried to do any of them.

If George's self-conception is coherent, it is reasonable to think that depression can block the motivational efficacy of desires (including action-desires). Despite the fact that they often are sources of motivated action, even action-desires do not *necessarily* motivate. It is thus no objection to the claim that moral judgments are capable of being sources of motivation—i.e. that they are desires—to say that their motivational force can be blocked by severe depression.

I do not pretend to have a knock-down argument for the claim that George's self-conception is ultimately coherent. At the same time, the burden lies with those who would want to deny that it is. A few more details may help to shift this burden even further. We may suppose that, were it not for his depression, George *would* be moved to do these things he claims to want to do. Recognition of this fact might even be part of what makes his current state so unbearable. Moreover, from his perspective, there is a clear difference between his attitude toward these activities and his attitudes toward others. Even if he is not motivated to engage in these activities, his attitude toward finishing his manuscript, working on the garden, or leaving the house is not the same as his attitude toward whether he takes an odd or an even number of breaths in the next hour (something he is utterly indifferent to) or his attitude toward divorcing his wife

(something he wants not to do). Together, these factors suggest that George has a coherent self-conception according to which he has action-desires that fail to motivate him.

It bears emphasizing that I am not leaning so heavily on George's self-conception because I think we are infallible about our own mental lives. We clearly are not. The point is simply that it is the same types of considerations that make depression a problem for *Judgment/Motivation Internalism* also make it a problem for *Action-Desire/Motivation Internalism*. Since the motivational cognitivist cannot be saddled with the former claim unless something like the latter claim is true, the argument based on depression is unstable.<sup>13</sup>

An alternative way to reach the same conclusion does not rely on an individual's self-conception. Clinical depression is often portrayed as a disorder of motivation generally.<sup>14</sup> In fact, it seems to be something of a commonplace that the decreased energy and other symptoms associated with major depressive episodes is a matter of a severed connection between an agent's goals, plans, or commitments—in short, anything that Humeans want to reduce to desires—and motivation.

---

<sup>13</sup> Mele (1995, 408) anticipates a similar objection:

Someone might contend that even if cognitivist moral ought-beliefs are not *EMC* attitudes, action-desires are on no firmer ground in this respect. It may be claimed, for example, that a jogger might desire to run another lap but be too tired even to try, so that this desire is not motivation-constituting.

In response, Mele (1995, 408, original italics), subsumes this example under the case in which the agent is convinced that he would not succeed if he were to try: "Again, an agent who takes his *A*-ing to be physically impossible might hope or wish that he could *A*, but he does not *desire* to *A*."

<sup>14</sup> See, for example, Randolph Nesse (2001).

One important correlate of severe depression that differentiates it from more typical lowered motivation is a resistance to the abandonment of goals and commitments in the face of strong evidence that, among other things, those goals are lost.<sup>15</sup> In normal individuals, such evidence leads first to a downgrading of the motivational system and then to the abandonment of the goal or commitment. Individuals prone to depression, on the other hand are more likely to continue holding the unattainable goal, despite the lowered motivation and this often leads to further downgrading of the entire motivational system (Nesse 2001). Insofar as the entire motivational system is affected, the lowered motivation is not limited to the particular goals that are seen as lost, but extends to other goals, plans, and commitments as well. If some of these goals, plans, or commitments are best analyzed in terms of action-desires—a hypothesis that cannot be ruled out by fiat—then the lowered motivation associated with depression might extend to action-desires as well.

#### 4. Motivational Cognitivism and Amoralism

It may be objected that the motivational failure discussed in the previous section is unique to major depression. Even if it is compatible with the falsity of *Judgment/Motivation Internalism*, motivational cognitivism may still be committed to a weaker thesis:

***Restricted Judgment/Motivation Internalism:***

Necessarily, if an agent who is not suffering from severe depression judges that an action,  $\phi$ , is morally right (best, obligatory, etc.) she will be

---

<sup>15</sup> Indeed, according to Charles Carver and Michael Scheier (1990, 29), “this bind—being unable to let go of something that is unattainable—lies at the heart of exogenous depression [...]” See also Nesse (2001).

motivated to  $\phi$ ; if she judges that  $\phi$ -ing is morally wrong (inferior, impermissible, etc.) she will be motivated to not- $\phi$ .<sup>16</sup>

The Humean is confident that amoralists—at least those not suffering from depression—provide counterexamples to *Restricted Judgment/Motivation Internalism*. So, the Humean may be optimistic that this claim can be used to undermine motivational cognitivism.

To the cognitivist, the Humean optimism is premature. For reasons that parallel our discussion of the previous section, the motivational cognitivist can be saddled with this thesis only if a similar claim holds true of action-desires generally:<sup>17</sup>

***Restricted Action-Desire/Motivation Internalism:***

Necessarily, if an agent who is not suffering from severe depression desires to perform some action,  $\phi$  (has some desire-like state to the effect that she  $\phi$ s), she will be at least somewhat motivated to (try to)  $\phi$ .

Yet the cognitivist thinks that there is good reason to doubt this key thesis. An initial case against *Restricted Action-Desire/Motivation Internalism* focuses on another character from the philosophical literature: John McDowell's virtuous agent. According to McDowell (1978, 90-93; 1979, 55-56; 1980, 17-18), such an agent is invariably motivated to comply with the demands of morality as she recognizes them to be because all contrary desires are "silenced" in the face of moral demands. From her perspective,

---

<sup>16</sup> This is essentially the thesis against which Svavarsdóttir's (1999, 165) main arguments are targeted. As she notes, we could frame this in terms of rational agents, rather than agents not suffering from severe depression. However, formulating the point in such terms would get us tangled up in substantive questions about rationality that we are better off avoiding.

<sup>17</sup> *Modulo* any qualification needed to accommodate cases in which an individual is convinced that she cannot  $\phi$ .

desires, partial interests, and the like, all lose their motivational force in the face of contrary moral judgments.<sup>18</sup>

I do not intend to defend the claim that the virtuous agent must be as McDowell describes. I insist only on two points. First, it is clearly possible that there could be such agents. Second, we should take seriously the possibility that such agents could still want to do things that think are contingently impermissible—that is, that they have action-desires with objects that, because of how the world turns out, should not be performed. It follows from these two observations that such agents could have action-desires lacking any actual motivational force.

To fix on an example, suppose that Marcy considers it unacceptable to spend money on frivolous items while half of the world's population lives in severe poverty. Despite this, she still wants to do certain things—things that she recognizes to be frivolous, like going to a ball game or buying a new pair of jeans. Were she a McDowellian virtuous agent, these desires would not move her in the least.

None of this strikes me as incoherent or obviously impossible. In support of this, we can leverage the very same considerations that were used to undermine our previous internalisms—in particular, that there is a coherent self-conception available to Marcy according to which all of this is true. This self-conception is supported by other attitudes she may have—for instance, her express wish *that it would be ok* for her to go to the ball game or buy the jeans. From her perspective she might have such a wish because she would then be free to do these things without any remorse. It would be hard to make

---

<sup>18</sup> This is unlike the case of the *merely continent* agent who still feels “the pull” of these desires, but systematically complies with her moral judgments nonetheless (McDowell 1979, 53-57).

sense of these attitudes—a wish that it were permissible for her to do these things—if she did not want to do them in the first place. We might even suppose that although she is not actually motivated by these putative action-desires, she is disposed to be motivated by them in different situations—in situations where the moral landscape differs sufficiently. Indeed, Marcy may even insist on this: it may seem to her like she would be moved to go to the ball game or buy the new jeans, were it not for morality demanding that she do otherwise.

There is nothing incoherent about agents like Marcy—agents that remain unmoved by particular action-desires when morality seems to require that they do otherwise. Even if we suppose that those very same action-desires would have motivational force in other situations, if their motivational force can be silenced, *Restricted Action-Desire/Motivation Internalism* is false. Motivational cognitivism cannot be tied to *Restricted Judgment/Motivation Internalism* and the argument once again fails.

Our discussion has been somewhat artificial. But as far as our argument is concerned, it does not really matter that it is the agent's moral judgments that are silencing her other desires. McDowellian virtue is merely a special case of a more general phenomenon.

When deliberating about what to do, typical agents don't consider every possible course of action; we don't even think about every obvious or likely means to our ends. Some actions are off the table from the get-go and we don't even entertain them as serious options. Despite wanting to do things they cannot afford (say, traveling around

the world or paying off one's credit cards), most people never really take bank robbery, murder-for-hire, or drug running as serious ways of making this happen.

We need not speculate about why more people do not consider robbing banks, killing people for money, or dealing drugs—it matters not whether it is because these vocations are seen as morally unacceptable or because they have significant risks associated with them. What matters for our purposes is what might happen when, after ruling out such non-serious routes to performing a particular action, an agent is left with no further options and thus has no serious way of doing what she wants to do. It is a plausible hypothesis that when left with no serious options, some agents may fail to be motivated to engage in an activity they want to engage in. Despite wanting to travel around the world, or to pay off one's credit card, when left with no serious ways of doing so, an agent may not be motivated accordingly. In such cases, the motivational force of the agent's desire may be silenced in virtue of the fact that any available route to doing what she wants to do is not a serious option.<sup>19</sup>

It should be clear why we might consider McDowellian virtue as nothing more than a special case of this more general phenomenon. In McDowell's virtuous agents, moral considerations automatically exclude a range of options as being non-serious. Marcy's desires fail to move her because there is no serious way for her to satisfy them—

---

<sup>19</sup> To avoid a potential misunderstanding, to say that the agent has no *serious* way of doing what she wants to do is not to say that the actions are “doxastically closed” to her. Such an agent need not be convinced that she would fail if she tried. One can believe that she would make a pretty good bank robber, assassin, or drug runner without taking these vocations as serious options. To take a more familiar example, there may be no serious way to satisfy one's desire to eat candy at the moment, even if there is a nearby baby from which it could safely be taken. We need not take this as a serious option even if we know that we would succeed if we tried.

she might want to buy the new jeans, but in the face of global poverty, she cannot take such frivolity seriously.

Even *amoralism* may be explainable in terms of a similar mechanism. Zangwill, you may recall, described his mercenary as being concerned only with his own interests. One way this could happen is if the agent does not take selfless behavior seriously—any action that runs afoul of his own interest is off the table from the get-go and is not considered as a serious option.

We seem to have a general mechanism by which a putative motivational state—be it an action-desire or a moral judgment—could have its motivational force silenced. If some mechanism could work like this, both *Restricted Judgment/Motivation Internalism* and *Restricted Action-Desire/Motivation Internalism* are false and for much the same reason. Once again, a crucial premise in the Humean argument is undermined by the considerations that make the other premise plausible. The two parts of the argument are in tension with one another.

## 5. Humean Explanations

I have argued that depression and amoralism are compatible with motivational cognitivism—the view does not imply that moral judgments, or any other cognitive states, necessarily motivate. In this section, I argue that similar considerations pose a dilemma that undermines the Humean's supposed explanatory advantage. On the one hand, we can accept that workaday explanations of an agent's actions citing only her moral judgments can sometimes be adequate. This would be to reject the Humean's assumption that these explanations are either missing or implicitly contain reference to

some distinct motivational state. Alternatively, we can reject the standard Humean account of moral motivation as itself being objectionably shallow and insist that if depression and amoralism cause problems for motivational cognitivism, they cause problems all around.

An adequate account of moral motivation must acknowledge two facts. It is no mere accident that people are systematically moved by moral considerations. Any satisfactory picture must explain this convergence. We must also explain variations in moral motivation. Our discussion so far highlights two broad types of variation. First, there are differences *between* agents: individuals who have similar moral beliefs might be motivated differently by them. Second, there are motivational differences *within* individuals: agents can be motivated differently by their moral judgments at different times (or over different periods of their lives), and at any given time, an agent may be moved more by some moral considerations than she is by others.

Humeans suggest that such explanations are easy, requiring only a limited set of tools: desires, means-ends beliefs (including beliefs about constitutive means), and variations in desires' strengths. The vast spectrum of human action can be produced by precise manipulations of these levers and knobs.

On the Humean picture, convergence is guaranteed by positing widespread desires. Why do agents systematically pursue food and sex? Because (the Humean might say) there is a universal (or near enough) desire for pleasure. Why do typical agents run away from large predators? Most likely, it is because they want to live. Why are teenagers and young adults systematically drawn toward dangerous and irresponsible

activities? Because nearly everyone craves excitement at that age. By positing widespread or universal desires, the Humean gets convergence for cheap.

The convergence of *moral* motivation can be explained by appealing to widespread desires of three types. First, there are desires with “thin” moral terms in their contents—the desire to be moral, to be a good person, to do what’s right and the like. Then, there are desires with contents involving “thick” moral terms. People might want to be honest, fair, friendly, trustworthy, helpful, and so on. Finally, some desires do not have contents containing moral notions, but are still properly related to moral motivation insofar as they have morally relevant objects. Well-being and harm, on most accounts, are morally relevant factors—contributing to or determining the moral worth of actions—and they are also the kind of thing that someone might care about.

Among Humeans there is significant disagreement regarding which of these desires (if any) should be considered distinctive of moral motivation.<sup>20</sup> We need not weigh in on this dispute here. All that matters for us is the general structure of the Humean explanation—that desires of one or more of these types are sufficiently widespread to account for the widespread similarities in moral motivation.

Motivational cognitivism also easily accounts for the regularity of moral motivation: moral judgments can be independent sources of motivation and can move us all by themselves. If moral judgments can directly motivate, it should be no surprise that

---

<sup>20</sup> According to Zangwill (2003, 144), for instance, moral motivation is a matter of desiring actions that have the property of “being morally better than some alternative.” Svavarsdóttir (1999) contends that moral motivation requires the object to be seen under a distinctively moral mode of presentation. Smith (1994, 75), in contrast, famously likens such desires to a moral “fetish”.

we are systematically moved by them. The motivational cognitivist, no less than the Humean, can have convergence on the cheap.

Here is where our externalist cases reemerge. According to the Humean, this explanation is too shallow because depressed agents and amoralists make moral judgments but *they* are not moved by them. This account lacks the structure needed to account for the variety of motivational profiles embodied by real or possible human agents. Moreover, the objection continues, the Humean Theory gives us the needed structure. Accounting for the variety of motivational profiles requires us only to adjust the settings of the levers and knobs. In other words, the Humean need only point out that widespread desires are not necessarily universal and can come with various strengths.<sup>21</sup>

We have, in fact, already seen this story. According to Stocker (1979, 744) and Mele (1996, 734), you may recall, the motivational failure associated with depression is due to a “lessened” or “absent” desires. Zangwill (2008) likewise explains his mercenary’s lack of motivation in terms of the weakness of his moral desires. Svavarsdóttir (1999, 217-218) extends this explanatory strategy to account for more mundane differences in moral motivation:

It is not only that we can explain the difference between the moral cynic and the morally committed by ascribing the desire to be moral just to the latter. We can also explain why the motivational impact of moral judgments seem to vary among morally committed people such that in some people it is more easily overridden by contrary motivations than in others. It is because of the strength of their desire to do what is morally valuable or required differs; this is what underlies differences in their degree of moral commitment.

---

<sup>21</sup> Here I am ignoring differences explained by reference to differences in means-ends beliefs. Such cases are not at issue in the present discussion.

Motivational cognitivists seem to be at a disadvantage. Their account, unlike the Humean's, lacks the structure to yield both convergence and variety. But this is not exactly right. One upshot of our earlier discussion was that even *Restricted Action-Desire/Motivation Internalism* is false—desires do not guarantee appropriate motivation and thus the lack of motivation does not guarantee the lack of desire. Therefore, someone might object, we don't really get convergence by appealing to widespread desires; make them as widespread as you like, if desires do not guarantee motivation, they do not do the necessary work. In much the same vein, our objector might insist, the Humean doesn't really get variety either—if one can have a desire without being motivated, then its absence will not always make a difference.

To help make the argument less abstractly, consider an extension of George's case. Prior to his tenure denial, George always tried to be a good person. Throughout his life, he has been driven by his desires to do the right thing, be honest, treat others fairly, be a good friend, and so on. Just as before, these morally relevant desires do not move him as they once did. This sometimes results in lessened motivation to do things that he knew were morally required, fair, friendly, and so on. For instance, when his friend Sally, who had performed comparable favors for George in the past, asks him to pick her up at the airport, he is not motivated to do so. He recognizes that this refusal is unfair to her, and is not an act of friendship, and so he knows that he really ought to help her out. Despite this, and despite his desires to be fair, to do the right thing, and so on, George is not at all motivated by these considerations. Given his current state of despair, these desires don't move him any more than his desire to leave the house or work on his manuscript.

We cannot merely assume that George's lack of motivation can be explained by an absence of desire. For, he might have the relevant desires. This goes for moral motivation too: George might have morally relevant desires that fail to move him. Humean explanations run aground in precisely the same way that made the motivational cognitivist's picture objectionably shallow. If this phenomenon undermines one explanation, it undermines both.

The argument need not depend on depression—silencing also does the trick. Suppose that Chuck, unlike McDowell's virtuous agent, frequently lacks motivation to do the right thing. He genuinely wants to be moral, but this desire is frequently silenced by his other wants—if it requires more than a mild inconvenience, his desire to be moral doesn't move him at all. We might even imagine that Chuck laments his failure after the fact, and feels guilty about his tendency to do the wrong thing. But when push comes to shove, he is frequently unmoved by his professed desire. In the heat of the moment, morality has no pull.

This, once again, is plausibly the same general phenomenon involved in both amorality and McDowellian virtue. This case, however, makes it explicit that morally relevant desires can be silenced in much the same way as others. Chuck might share desires with McDowell's virtuous agent while instantiating a dramatically different motivational profile. If so, positing widespread moral desires does not guarantee convergence. Moreover, Chuck and the amoralist might instantiate very similar motivational profiles, while differing significantly in desire. Differences in desire do not automatically secure motivational variety. Humean explanations are pretty shallow too.<sup>22</sup>

---

<sup>22</sup> It might be objected that I have left out variations in desires' *strength*, and that the phenomenon I have been describing as silencing is simply a strong desire swamping a weak one.

Alternatively, the motivational cognitivist might argue, it is not that Humean explanations are objectionably shallow; it is just that the motivational cognitivist's explanations are not too shallow either. Instead, the objection might go, we should reject the implicit assumption that all motivation must have its source in an essentially motivating state—that is, to reject the pretense that explanations are too shallow unless they cite or implicitly reference such a state. Ordinary sources of motivation are not necessarily motivating. In the normal case, the fact that George wants to work on his garden is a perfectly good explanation of why he is grabbing his spade. We might be missing part of the explanation—for example, we have plausibly taken the means-end belief for granted—but we need not be missing or implicitly assuming any further *motivational* element. It is true that this desire will not always move him. But this recognition does not undermine the fact that this desire is what moved him here.

Similarly, that Marcy judged that she morally ought to donate money to famine relief strikes me as a perfectly good explanation for why she wrote a check to Oxfam. Why should there be any further *motivational* element missing from or implicitly assumed in this explanation? Why isn't the moral judgment enough? It is no good to point to a different case in which it doesn't move her (or someone else). All we are

---

To see why the response won't do, we must say a bit more about the strength of desire. It often seems like when people talk about a desire's strength, they are simply talking about how much it motivates. (I suspect that this is why Stocker (1979, 744), for example, slides *immediately* from talking about one's feeling "less and less motivated" to talking about "one's lessened desire".) If this is what the strength of a desire *consists in*, then strength does no explanatory work—it is really nothing more than another name for the explanandum. If, on the other hand, a desire's strength consists in something else, then we need a deeper story about what that something else is and how it does explanatory work. Moreover, to be *dialectically* useful to the Humean, the motivational cognitivist must have no analogous story to tell. After all, if the motivational cognitivist accepts my suggestion that moral judgments don't always motivate in the same way at all times or in all people, she must be willing to posit a knob here as well—the motivational strength of moral judgments can vary just as much as the motivational strength of desire.

saying is that it moved her here. To insist otherwise would either beg the question or assume an illicit double standard. Either way, the Humean has no advantage.

## Chapter 3

### Dispositions, Motivation, and Desire

In Chapter 2, I provided some intuitive, case-based arguments to suggest that as we ordinarily think about them, desires are capable of motivating agents to act but do not *necessarily* do so. In this chapter, I provide additional support for this key idea. I develop a general argument for the conclusion that motivational states need not be necessarily motivating. The argument I will be defending is that there is a necessary condition on motivation that desires—even action desires—satisfy only contingently: roughly, that an agent *S* in circumstances *C* is motivated to perform some action  $\varphi$  only if *S* is disposed to  $\varphi$  (or try to  $\varphi$ ) in *C*, or disposed to intend in *C* to  $\varphi$  at some later point.

In Section 1, I explain and motivate the dispositional condition on motivation. My main argument that desires satisfy this constraint only contingently assumes that some version of the dispositional account of desire is true. Section 2 lays the groundwork for this argument by offering an overview of dispositionalism and a partial taxonomy of extant versions of it. In Section 3, I defend a refinement of the general dispositional account that is rooted in an analogy with commonplace dispositions. I then argue that general features of this analogy suggest that desires satisfy the key constraint on motivation only contingently. I conclude, in Section 4, by offering a programmatic account of how the dispositional account I have developed is capable of accounting for a range of variation in motivational profiles embodied by real and hypothetical agents, and can do so without appealing to essentially or necessarily motivating states.

## 1. Being Moved

### 1.1 Two Senses of Motivation

I begin by defending the constraint on motivation. As we noted in Chapter 2, there is more than one sense in which an agent can be motivated or moved to do something. In one sense, an agent is moved to do something only if she actually does it. Whenever an agent does something purposefully one can sensibly ask of her “What motivated you to do that?” When such a question is asked, it presupposes that the agent actually performed, or at least tried to perform, the act in question. Suppose that the detective asks the lead suspect, “What motivated you to kill your wife?” If the suspect responds by insisting that he did not (try to) kill her, the detective could not sincerely retort, “I never suggested that you did!”<sup>1</sup> Call this the “success” sense of motivation.

There is also a second way in which we often talk about being moved to do something. When the vegetarian has a sudden urge for a steak but resists the temptation, we are often willing to allow that there was some sense in which she was moved to order the steak. When a man is having difficulty deciding between two outfits, there is a perfectly good sense in which he is being moved or “pulled” in between his options even though he ultimately wears only one of them. It is for this reason that we must make room for both a success sense and a weaker sense of motivation.

There are other important differences between these two senses. The weaker sense tends to be somewhat phenomenologically loaded. When someone is motivated in

---

<sup>1</sup> There are, however, related terms—such as *motive*—that do not share this connotation. If the defense attorney asks, “What motives do you have for killing your wife?” and the suspect pleads his innocence, the attorney can sincerely respond, “I am not suggesting that you did—I simply need to know if you have any motives that the police will use to cast suspicion on you.” This point was inspired by a conversation with Mark van Roojen.

this sense we think of her as *feeling* tempted or pulled or moved, even if she does not give in. Yet this phenomenological quality need not be present when someone is moved in the success sense. As you are walking down the street, placing one foot in front of the other, you are motivated to get to your destination. But you need not *feel* pulled or moved to go to the coffee shop in the same way as you do when you are somewhat torn about where to go.

## 1.2 Motivation and Dispositions

Given the differences between them, it is reasonable to ask what (if anything) links these distinct senses of motivation; what do the different ways in which we can be moved to do something have in common that make them both ways of being moved? It is a natural hypothesis that whatever they have in common has something to do with their connection to overt action. But given that the weaker sense does not require that any such behavior is produced, we must look for a weaker link.

The leading candidate for such a link is to be found by examining behavioral *dispositions*.<sup>2</sup> Dispositions are properties that are conceptually tied to the production of particular types of events or episodes. *Fragility* is conceptually tied to *breaking*, *solubility* is conceptually tied to *dissolving*, and *the disposition to form a “skin”* (e.g. the property possessed by homemade pudding) is conceptually tied to *forming a “skin”*. The events that are tied in this way to the dispositions are said to be the dispositions’ *manifestations*. The reason why dispositions are a promising place to look is that, while

---

<sup>2</sup> This kind of view has been offered by Michael Smith (1987; 1994, Ch. 4), among others.

conceptually connected to their manifestations, commonplace dispositions can often be embodied by objects without these manifestations occurring: a knickknack can be fragile even if it never breaks; a substance can be soluble without actually dissolving; and a bowl of pudding can be disposed to form a skin without ever doing so.

By employing the notion of dispositions, we can help ourselves to a plausible hypothesis regarding the relationship between the two different ways of talking about motivation. As a rough first pass, when someone is moved *in either sense* to eat a steak, wear the navy blue pinstripe suit, or go for a walk, she is disposed to do these things, or at least to try to do these things. The two senses can diverge, in that when someone is moved in the success sense to do one of these things the disposition is actually manifested; when someone is moved in the weaker, non-success sense, the disposition may or may not manifest itself.

Technically, although we are focusing on motivated action, there is some reason to think that the account of motivation should be more general. Philosophers have often talked as though desires and intentions can be motivated.<sup>3</sup> In addition to being moved to act or to want something, ordinary language also allows us to talk about being moved or motivated to believe certain things.<sup>4</sup>

These “motivations” may also come in our two different senses. The question, “What motivated you to believe in modal realism?” carries the presupposition that the

---

<sup>3</sup> See, for instance, Thomas Nagel (1979, 29) and Ralph Wedgwood (2007, 33), respectively.

<sup>4</sup> Smith (1988a, 250) contends that “in ordinary language, a motivated belief is [...] not susceptible to normal rational explanation in terms of other beliefs.” But Smith is not right about this. When we talk about motivating an idea or theory or belief for someone, we are often trying to get them to “feel the pull” of the idea by showing how it aligns with other beliefs that they hold.

subject does, or did, believe the proposition in question in a way that clearly parallels the success sense related to action. On the other hand, when we have two conflicting arguments before us, we might “feel the pull” of both. Although people in such situations may either believe one of the conclusions or withhold judgment on the issue, we might still think that there was some sense in which they were moved to believe the other conclusion too. One can be tempted by modal realism while choosing to resist that temptation. This looks very much like the weaker, non-success sense of motivation. If one chooses to do so, one could modify the account I offer to accommodate these other varieties of motivation.

Setting this technicality aside, our rough idea must be sharpened in at least three further ways. First, this is meant only as a necessary condition on motivation, not a sufficient condition. Individuals can be disposed to perform a variety of behaviors while not being motivated in either of our two senses: people can be disposed to blink, twitch, snore, and have heart attacks, but are not typically motivated to do any of these things. In addition to further conditions on motivation generally, there may also be further conditions on the appropriateness of attributing to agents motivations in each of our two senses individually. It could be, for example, that there is some phenomenological condition that must be met in order to properly attribute motivation to an agent in the non-success sense—at least when this motivation does not lead to overt behavior.

Second, as far as this necessary condition is to be understood, if an agent is motivated to perform some action,  $\varphi$ , it is not enough that she is disposed to (try to)  $\varphi$  in some situation or another. Intuitively it seems like she must be disposed to (try to)  $\varphi$  in her actual circumstances or relevantly similar circumstances. To see why this is so,

suppose that Glenda would vote Republican if she thought that doing so would help protect civil rights. Suppose further that she actually believes that voting Republican would not do this and would in fact be detrimental to civil rights. Is her disposition to vote Republican in some situations enough to satisfy the dispositional requirement for her actually being moved to vote Republican (in even the minimal sense)? I think not. In order for her to actually be moved (in either sense) to vote Republican she must be disposed to vote Republican in situations relevantly similar to her own—situations where she has relevantly similar beliefs about their record on civil rights, for instance.

Third, it is also possible for an agent to be now motivated to do something in the future. The week before the election, Glenda might be motivated to vote on Election Day. Given her (true) beliefs about voting regulations, there is nothing that she was disposed to do the week before the election that would constitute voting on Election Day. For this reason, our previous clarification would suggest that Glenda was not, contrary to our original assumption, motivated to vote. Worse yet, it is also possible for agents with such future-oriented motivation to lose it in the intervening period before action is required: when Election Day arrives, Glenda may no longer feel like voting. Moreover, there are some agents in which this loss of motivation is easily predictable. An alcoholic may be strongly motivated to have only one drink. According to our story, this would require the alcoholic to be disposed to have only one drink. But when that one drink is drunk, the motivation to have only one may evaporate and the corresponding disposition may disappear.<sup>5</sup>

---

<sup>5</sup> This is similar to cases that Eve Garrard and David McNaughton (1998, 50-51) discuss to undermine attempts to give a counterfactual analysis of motivation. The literature on dispositions provides an interestingly similar case. Against attempts to provide analyses of

To avoid these problems, we may think that the disposition to now *intend* to  $\varphi$  at some future time suffices for our purposes. Even if there was nothing she was disposed to do the week before the election that would constitute *voting*, Glenda was disposed to then *intend to vote* on Election Day. If she is disposed to form such an intention in her present circumstances, then that should be enough for her to count as satisfying this necessary condition on being motivated to vote. When Election Day comes, if she is no longer disposed to vote or intend to vote, then she no longer counts as being motivated. Before taking the drink, the alcoholic may have intended to have only one drink—which suffices for satisfying the necessary condition on being motivated to do so—even if after that one drink he is no longer so disposed.

Stated formally, with these clarifications in mind, the two different senses in which someone can be motivated are related in the following way:

***Motivation-Disposition Link***

An agent  $S$  in circumstances  $C$  is motivated to perform some action  $\varphi$  only if either  $S$  is disposed to (try to)  $\varphi$  when  $C$  (or relevantly similar circumstances) or  $S$  is disposed in  $C$  to intend to  $\varphi$  at some later time.

When  $S$  is motivated to  $\varphi$  in the *success sense*, this disposition is actually manifested in  $S$ 's (trying to)  $\varphi$ ; when  $S$  is motivated to  $\varphi$  in the weaker, *non-success sense*, this disposition may or may not actually be manifested.

### 1.3 Motivation and Masking

One final clarification is necessary to show that the *Motivation-Disposition Link* is appropriately weak to allow for the non-success sense of motivation. I mentioned a few

---

dispositional notions in terms of counterfactual conditionals, C.B. Martin (1994, 2-3) provides the example of a disposition that would go away at the very instant that its triggering conditions obtain, and so the conditionals linking triggering conditions and manifestations would turn out false. For further discussion of these “finkish” dispositions, see David Lewis (1997) and Stephen Mumford (1998, 54-56; 82-92).

moments ago that the move to the level of behavioral dispositions should give us the correct strength because dispositions can be embodied by objects without actually producing their characteristic manifestations. Most dispositions produce their manifestations in relatively specific types of situations. Fragile objects do not tend to break without good reason—they typically break *when impacted by a suitable concussive force*. Soluble chemicals dissolve only *when placed in a suitable solvent*. Pudding is disposed to form a skin *when left standing for a suitable amount of time at a suitable temperature*. The conditions described by these *when*-clauses are the dispositions' *triggering conditions*. So one way in which an object can fail to produce the characteristic manifestation of one of its dispositions is if the relevant triggering conditions never arise. A soluble substance will not dissolve if it is never placed in an appropriate solvent, but it is nonetheless *disposed* to dissolve *when* it placed in such a solvent.

This is not the way in which dispositions relevant to the *Motivation-Disposition Link* fail to manifest. For, if the *Motivation-Disposition Link* is correct, the relevant triggering conditions might actually obtain; the agent in circumstances *C* who is motivated to  $\phi$  is disposed to (try to) do so in her very own situation.

To explain the way in which the relevant dispositions remain unmanifested, we must look at another analogous issue with dispositions more generally. Many commonplace dispositions are not guaranteed to produce their manifestations even in the presence of relevant stimuli. A standard example is that a good packing job will prevent a fragile knickknack from breaking when handled roughly, but this is far from the only example we could use. Pudding is disposed to form a skin when chilled, but it won't if

you cover its surface with plastic wrap. Massive objects are disposed to accelerate downward at the rate of 9.8 meters per second squared when dropped near the Earth's surface, yet wind resistance will prevent objects from accelerating at this rate. A round barrel, lying on its side is disposed to roll when a given amount of force is applied, yet this will not happen if the barrel has been chocked.<sup>6</sup> In these cases, the dispositions to break, form a skin, accelerate at a particular rate, or roll when pushed are said to be *masked* or *inhibited*.<sup>7</sup>

The *Motivation-Disposition Link* suggests that the vegetarian is disposed to order a steak and disposed to abstain from meat and that the man who is torn between wearing the navy blue suit and the chocolate sport coat is disposed to wear each. Since the vegetarian cannot both order steak and abstain from doing so, and since the man is not about to wear both outfits, one of their dispositions will be masked when the other is manifested. More generally, when an agent in circumstances *C* is moved to perform some action  $\phi$  but is not moved in the success sense to (try to)  $\phi$ , and she is not disposed to intend in *C* to  $\phi$  at some later time, she has a masked disposition to do so.

## 2. The Dispositionalist Framework

### 2.1 Some Preliminaries

Before I am in a position to defend the claim that desires satisfy the constraint on motivation embodied in the *Motivation-Desire Link* at best contingently—which will be

---

<sup>6</sup> This example is owed to Michael Fara (2005).

<sup>7</sup> For more on masking, see Alexander Bird (1998), Fara (2005, 48-61), and David Manley and Ryan Wasserman (2007, 68 & 74).

the topic of Section 3—I will first need lay some groundwork. I will be accepting the familiar assumptions that an agent's desires are analyzable in terms of an underlying set of dispositions she possesses and that desire can be better understood by employing the model provided by commonplace dispositions. However, I am thinking about these assumptions in a weaker way than they are often understood. As I am thinking about things, the key idea is just that the correct account of desire will involve something approximating the following generic thesis:

***Generic Dispositionalism:***

For all agents  $S$ , and all propositions  $p$ ,  $S$  desires that  $p$  if and only if for some class of characteristic manifestations  $M$  and relevant triggering conditions  $C$ ,  $S$  is disposed to  $M$  when  $C$ .

Understood this weakly, many extant theories of desire can be seen as endorsing my assumptions. These theories are differentiated primarily by what they claim to be the characteristic manifestations and triggering conditions of desire—those events to which desire is conceptually connected and the situations in which these events tend to be produced. To help us better understand this key idea, and the particular version of it in which I am ultimately interested, it will be useful if we briefly canvas the types of events that philosophers have claimed to be tied to desire in just this way. This will give us rough and partial taxonomy of common varieties of dispositionalism.

## **2.2 Behavioral Accounts**

Desires are frequently analyzed by reference to their role in the production of overt behavior. *Behavioral dispositionalism* is the view that, as Michael Smith (1994, 113) puts it, “we should think of desiring to  $\varphi$  as having a certain set of dispositions, the

disposition to  $\psi$  in conditions  $C$ , the disposition to  $\chi$  in conditions  $C'$ , and so on [...]", where  $\phi$ ,  $\psi$ , and  $\chi$  are understood to be courses of action or other overt behaviors, and both  $C$  and  $C'$  are understood to include (among other things) facts about what the subject happens to believe.

Behavioral dispositionalism is itself open to different interpretations, depending on how one fills in the variables in the basic statement of the view. The most common version of this view—indeed the prevailing philosophical account of desire *writ large*—focuses on what we might call *productive* behaviors. Desiring something, on this view, is a matter of being disposed to try to get it.<sup>8</sup> Robert Stalnaker (1984, 15) provides a canonical formal statement of a view along these lines: "To desire that  $P$  is to be disposed to act in certain ways that would tend to bring it about that  $P$  in a world in which one's beliefs, whatever they are, were true." Smith also clearly wants to interpret the basic framework in much the same way. He thinks that the actions that fill out the basic structure will typically be ones that given her means-ends beliefs, will help see to it that the desire is satisfied (Smith 1994, 113 & 115).

According to this standard picture, then, the characteristic manifestations of desire are behaviors that, given the agent's take on the situation, would promote the wanted thing. The manifestations conceptually tied to wanting to read the morning paper are trying-to-read-the-morning-paper behaviors—those behaviors that the agent thinks are likely to allow her to read the morning paper.

---

<sup>8</sup> Compare Elizabeth Anscombe's (1957, 68) famous claim that "The primitive sign of wanting is *trying to get*."

Over the course of the last decade or so, this prevailing view has seen an increasing amount of criticism. One standard objection is that this view has difficulty accounting for desires with necessary or impossible objects.<sup>9</sup> Someone might desire that a certain mathematical formula turns out true, that there is a necessarily-existing benevolent deity, or that *this* is a brontosaurus skull (about the thing she has just unearthed). If these things turn out true (or false), they are true (false) no matter what one does. There is nothing the agent is disposed to do to make these things happen.

The prevailing view also has a hard time explaining desires with contingent objects that cannot be promoted by the agent. While it is only a contingent fact that each of us was born, someone who wishes that he had never been born cannot be disposed to do anything that would make this so (T. Schroeder 2004, 16). Or, while on vacation, an individual may hope that she had turned off the water before she left home. There is nothing she is now disposed to do to make it the case that she turned off the water then (Schueler 1991, 280).<sup>10</sup> One's behavior has no impact whatsoever on whether or not these desires are satisfied.

Such cases may or may not make it impossible to give a fully satisfactory formalization of the prevailing view.<sup>11</sup> But they do make it quite clear that for many

---

<sup>9</sup> These problems are discussed by Timothy Schroeder (2004, 16-17), among others. Compare Anthony Kenney (1968, 121-122), who contends that agents cannot desire impossibilities—even impossibilities that agents do not recognize as such.

<sup>10</sup> See Also Alfred Mele (1995, 394).

<sup>11</sup> In a note, Smith (1994, 208-209 n. 7) contends that his version of dispositionalism needs to be refined in light of these problematic cases. Here, he contends that (at least these) desires must be analyzed, at least in part, by reference to dispositions “to make certain sorts of bets when faced with lotteries where the outcome is *inter alia* that *p*.” Smith (1998, 450-451) reiterates the betting behavior picture when responding to Galen Strawson's (1994, Ch. 9) discussion of these sorts of desires. The betting-behavior strategy faces an important objection.

desires, the behaviors that seem most relevant are not those that involve trying to get the desired object. By reflecting on these and other cases we can see that a hodgepodge of non-productive behaviors can be paradigmatic indications of desire. There are things one could do to make it more likely that the two horse will win in some particular race—one could drug it, threaten or bribe the other jockeys, or hire a hit-man to do his or her dirty work. While many of us might recognize such options, we are more likely to just stand and yell things like “Get up two!” at the top of our lungs. Speaking for myself, I don’t think this actually helps the horse win; nor do I think that cheering (or booing) after the fact helps either. Yet I (and most other people) do such things on a regular basis. These behaviors are, for the most part, pretty clear manifestations of our desires.

For this reason, we should at least expand the relevant range of behaviors to include those that occur in reaction to the information that the object will (or will not) obtain. Other salient behaviors might include attempting to verify whether or not the object obtains, acting as though the object obtains, and verbal behaviors such as avowing that one has the desire in question.<sup>12</sup>

More liberal versions of behavioral dispositionalism give these and other non-productive behaviors a bigger role to play. According to the most ecumenical version, the characteristic manifestations of a desire that *p* include “*all the possible behavioural*

---

A desire to make bets of a certain kind, on the dispositionalist picture, would itself have to be constituted by dispositions of some sort or another. However, it seems possible to have a desire to make a bet of a certain kind without having the desire that *p*, or *vice versa*, given that these two desires have different propositional contents. Yet if the desire that *p* is constituted by the dispositions to make the bet, as the betting-behavior strategy contends, then it is quite difficult how we could have one of these desires without the other. For further criticism of the betting-behavior strategy, see David Sobel and David Copp (2001, 48).

<sup>12</sup> See, for instance, Robert Audi (1973, 39) and Alfred Mele (1995, 393-395).

*evidence* that would bear upon whether someone desires that [*p*]” not just those behaviors that are conducive to *p* (Smith 1998, 450, italics added).

### 2.3 Non-Behavioral Accounts

Behavioral dispositionalism is not the only way to develop the generic dispositionalist assumption. Many accounts of desire give non-behavioral manifestations pride of place. I will focus on a cluster of these views that consider specific conscious mental states to be desire’s characteristic manifestations.

When we theorize about desires, there is a stereotype that seems particularly important. We often think about the desires for food, drink, and sexual gratification that are associated with hunger, thirst, and other brutish drives. Going beyond these paradigmatic cases, we might also think about other common cravings and urges: the craving for chocolate or salty snacks, or the urge to smoke a cigarette, for instance. When we examine these stereotypical desires, it is remarkable how they all feel a certain way—how they share a similar occurrent phenomenology.

It is difficult, if not impossible, to describe this common phenomenology in any precise way. Perhaps the best we can do is say that these attitudes all feel *urgent*. Nor is the phenomenology of these states exactly the same as one another. At the very least, it can vary in intensity. While going through withdrawal, an addict’s craving for heroin probably feels much more urgent than my current itch for a salty snack.

If we take these stereotypical desires to be a good model for desires generally, one might be convinced that this felt urgency is part of the concept of desire. One way this could play out is if desires are identified with this type of phenomenology. Intuitively,

however, this is not quite right. For, we can have standing desires that do not always feel this urgent to us. I might want the House Republicans to give up their quest to repeal the Affordable Care Act even when this issue is not at the forefront of my mind.

An alternative way to acknowledge the importance of phenomenology is to incorporate it into a dispositionalist framework as the characteristic manifestation of desire. According to such a view, desires are to be analyzed in terms of an agent's dispositions to be in certain occurrent mental states with this felt urgency in certain situations. This allows one to say a subject who desires that  $p$  need not actually have an experience with this sort of flavor with regard to  $p$ —the point is only that there must be a disposition to produce an experience of that sort when suitable conditions arise. In the paradigmatic examples of desire, these dispositions have succeeded in producing the event of the appropriate sort. In other cases, the disposition may never be manifested in this way, and the phenomenology of felt urgency never materializes.

Another class of conscious mental states that are often thought to be conceptually tied to desire includes affective and hedonic states. While there has been a longstanding philosophical debate about the precise nature of the connection between desire and (dis)pleasure or (un)happiness, there is broad agreement that there is some intimate connection. According to one hypothesis, the link is just this: these states are characteristic manifestations of desire. On this view, desiring that  $p$  involves, dispositions to be “pleased or happy or contented should [ $p$ ] come about (or at least to [cease] to be unhappy or discontented should it come about)” and dispositions to be “unhappy or discontented or disappointed should [ $p$ ] not come about” (Strawson 1994, 280). Affective and hedonic states can be produced in other situations as well. For

instance, an individual may be disposed to be (dis)pleased when she daydreams about *p* or thinks about what it would be like if (not-) *p* were the case (Audi 1973, 39).

It is also possible that the felt urgency that I have attributed to many stereotypical desires is also related to the connection between desire and (dis)pleasure. It is often the case that the phenomenology of these desires is itself unpleasant. This, I think, is behind Stephen Schiffer's (1976, 198) idea that desires of this paradigmatic sort (he includes, as examples, thirst, cravings for chocolate, and the desire to scratch one's nose) can be uncomfortable, and the relief of this discomfort tends to be a further source of enjoyment.<sup>13</sup>

Yet another version of non-behavioral dispositionalism focuses on patterns of attention and evaluation as the characteristic manifestations of desire. A desire for something involves dispositions to think about that thing in certain ways and to make certain normative or evaluative judgments about it—especially dispositions to see it as good, valuable, desirable, or otherwise worth pursuing. T.M. Scanlon (1998, 39) adopts a version of this view, claiming that desiring “involves having a tendency to see something as a reason” and that

A person has a desire in the directed-attention sense that *P* if the thought that *P* keeps occurring to him or her in a favorable light, that is to say, if the person's attention is directed insistently toward considerations that present themselves as counting in favor of [*it*].

---

<sup>13</sup> In reference to desires of this sort (he calls them “reason-providing desires”, or “r-p desires”, for short) Schiffer (1976, 198) claims that “they are desires which, almost always, are both pleasurable to satisfy and discomfoting to endure, always one or the other; in fact the anticipated pleasure and relief of discomfort are nearly always inextricably related, in that what one anticipates is just the pleasurable relief of discomfort...”.

A related picture might center on dispositions to make certain normative or evaluative judgments in certain idealized situations, such as when one has complete or more complete information. A subject may be thought to want that  $p$  if she is disposed to judge that it was or would be a good thing that  $p$ , or that it should be or ought to have been that  $p$ , when she has all of the relevant information regarding what that was, would be, or would have been like.

## 2.4 Hybrid Accounts

While philosophers have often wished to argue that one or another of these phenomena is *the* characteristic manifestations of desire, some have been willing to be more ecumenical. Hybrid accounts allow for a variety of different things to count as the characteristic manifestations of desire—there needn't be a single type manifestation that is given pride of place. According to this type of picture, desires can be understood on the model of multi-track dispositions that have manifestations of multiple types.<sup>14</sup> Some of these manifestations may be of one type of conscious state, some of them may be of another, and some of them may be overt behaviors. Perhaps desires can be manifested in other ways as well.<sup>15</sup>

One especially clear example of a hybrid view has been put forward by Robert Audi. This account includes seven “laws” involving both behavioral and non-behavioral components that are supposed to be constitutive of our concept of wanting. On the

---

<sup>14</sup> For more on multi-track or variably manifesting dispositions, see Elizabeth Prior (1985, 96-98) and Stephen Mumford (1998, 66 & 182).

<sup>15</sup> Desire's role in reward-based learning might also be important. For a thorough discussion of this “face” of desire, see Timothy Schroeder (2004) who credits Fred Dretske (1988, Ch. 5).

behavioral side, Audi (1973, 39) cites trying-to-get behavior, avowals, as well as the tendency “in free conversation ... to talk about at least occasionally about  $p$  and subjects he believes to be connected with  $p$ .” On the non-behavioral side, he includes tendencies to find certain things (un)pleasant and disappointing, and the tendency to “think (reflect, muse, or the like) or daydream about  $p$ ” (Audi 1973, 39).

More recently, Mark Schroeder (2007, Ch 8) has offered a more restricted hybrid account. On his account, desire is connected to subject’s dispositions to pay attention to certain patterns in the environment. The intuitive idea is roughly that desiring something involves a tendency for certain features of the environment to be salient or to “stick out”—especially facts about potential means to that thing. In addition to these patterns of salience, Schroeder (2007, 156-157) adds a behavioral component to the formal statement of the view:

For  $X$  to have a desire whose object is  $P$  is for  $X$  to be in a psychological state grounding the following disposition: when for some action  $a$  and proposition  $r$  believed by  $X$ , given  $X$ ’s beliefs  $r$  obviously helps to explain why  $X$ ’s doing  $a$  promotes  $P$ ,  $X$  finds  $r$  salient, and this tends to prompt  $X$  to do  $a$ , and  $X$ ’s attention is directed toward considerations like  $r$ .<sup>16</sup>

Indeed, while he is held up as the leading proponent of behavioral dispositionalism, in some voices even Smith (1994, 114) sounds sympathetic to a hybrid picture that allows for phenomenological states to be included among the manifestations of at least some desires:

For, according to [a dispositional conception of desires], desires have phenomenological content just to the extent that the having of certain

---

<sup>16</sup> For another recent hybrid account, see Neil Sinhababu (2009, 468-472).

feelings is one of the things they are dispositions to produce under certain conditions. Some desires may be dispositions to have certain feelings under all conditions: these desires have phenomenological content essentially. Other desires, though they are dispositions to behave in certain ways, may not be dispositions to have certain feelings at all: these desires lack phenomenological content altogether.

## 2.5 The Structure of Dispositionalism

Before continuing with the main argument, it is worth highlighting the structure of the dispositionalist picture. According to *Generic Dispositionalism*, the relation between desire and dispositions is an asymmetric one: an individual has a particular desire because of or in virtue of embodying the relevant dispositions. Dispositions are the more basic piece of the theory.

Against behaviorism and behavioral dispositionalism (a view he calls neo-behaviorism), Galen Strawson (1998, 472) complains that “it seems to me that facts about how a subject behaves (or is disposed to behave) are fixed by facts about what she desires, not the other way about.” It is certainly true that one could endorse a law-like connection between desires and the dispositions in question while denying the dispositionalist’s direction of explanation. It could be that subjects who want something are (at least in general) disposed to produce certain characteristic manifestations even if this is no part of what it is to want something. An oxygen atom is disposed to react in certain ways with carbon atoms, but being so disposed is not *what it is* to be an oxygen atom. Facts about how an atom is disposed to interact with carbon are fixed by facts

about what kind of atom it is. So, one might be skeptical about the dispositionalist's direction of explanation.<sup>17</sup>

It is also important to recognize that despite this asymmetry, dispositionalism is compatible with the standard practice in philosophy and folk psychology to explain certain events—those that have the appropriate conceptual or nomological connection to desire—as the *product* of desire. It is extremely common for philosophers to explain an agent's productive behaviors by referencing her desire for something or another. Why did Mitch pick up eggs on his way home? Because he wanted to make pancakes for breakfast and he knew that he would need some eggs to do so. Dispositionalism offers a deeper story regarding this explanatory practice: the event to be explained is the manifestation of a disposition that is partly constitutive of that particular desire. While it is true that Mitch wants pancakes because he is disposed to do certain things, it is also true that he does those things because he wants pancakes.<sup>18</sup>

If non-behavioral phenomena are also to be explained by appealing to desire, then a similar story might be available for other varieties of dispositionalism as well. Lucinda is happy *because* she wanted the Boston Red Sox to win the division. Intuitively, the disposition to be pleased by the victory could be partly constitutive of a desire that the Red Sox win the division. Similar thoughts could apply to explanations of why someone keeps thinking about *p*. And so on. In general, an explanation of the form “*S M*-ed in *C* because she wanted that *p*” is true on the dispositionalist picture if and only if *S M*-ed as

---

<sup>17</sup> In the end, I share Strawson's unease about this. However, for the purposes of this paper, I am simply setting this concern aside.

<sup>18</sup> Some philosophers object that the appeal to dispositions does no real explanatory work. See, for example, Ulrike Heuer (2004, 56-57).

the result of the triggering of an underlying disposition to *M* when *C* and this disposition is partly constitutive of her desire that *p*.

### 3. Refining the Dispositionalist Account

#### 3.1 The Standard Move

In this section, I argue that desires are at best contingently motivating. General features of the analogy between desires and ordinary dispositions suggest that any satisfactory version of dispositionalism—behavioral or not—will be committed to the claim that desires are only contingently related to the dispositions necessary for satisfying the *Motivation-Disposition Link*.

Before continuing with the main line of argument, I wish to point out a way in which desires can fail to motivate that is not at issue. Any plausible theory of desire will acknowledge that desires are not guaranteed to produce the relevant manifestations if the relevant triggering conditions fail to materialize. When the agent has no thoughts about how to get what she wants, she may not be in a position to do anything to help her attain it and will therefore not be motivated to satisfy it. If Marty has no idea how to travel back in time, or thinks that this is impossible, he may not be disposed to do anything in his actual situation to do this, even though traveling back through time is one of the things that Marty wants to do. Other putative characteristic manifestations of desire can also fail to materialize in the absence of appropriate triggering conditions. If Lucinda does not find out that the Red Sox won, she is not likely to be happy about this. This type of failure is unobjectionable. Dispositionalism only claims that an agent is disposed to undergo the relevant manifestation *when the appropriate triggering conditions obtain*.

As we have already made sufficiently clear, many common dispositional notions are similar in this respect.

As I said, I am not interested in such cases. The cases that concern me involve competing desires. Typically, when an agent cannot get everything she wants, only some of her desires will manifest. Take the case of productive behavior. As I noted earlier, a steadfast vegetarian might crave a steak without giving in to the temptation—her desire to eat meat need not manifest in any attempt or intention to do so. This can happen to other putative manifestations of desire too, like patterns of attention or pleasure and happiness. Normally, if someone wants to catch a movie that he knows starts at 9:30, he is disposed to have his attention directed toward the time as show time approaches. However, if he wants to continue a particularly interesting conversation with his date, his attention might be directed so much toward his conversational partner that he doesn't notice when 9:30 comes and goes. Similarly, while in general, someone who wants great wealth might be disposed to be happy at receiving a windfall, if the money was received in some unfortunate way—say, he inherited the money because of the unexpected death of a dear loved one—he may not be pleased or satisfied in the least.

To the extent that people notice it, the standard line of response to this issue is to gesture toward the phenomenon of masking. This is the move Audi (1973, 45) makes when he insists that his seven “laws” of desire should be understood as to allow for the presence of inhibiting conditions. In support of this, he draws an analogy to the case of scientific laws, noting that magnetic interference (in addition to air resistance) can interfere with the rate at which objects accelerate. In much the same way, according to Audi (1973, 44), a subject may want a library to do well while being displeased when its

budget has been increased—if he “wanted even more for the money to go to the museum,” this contrary desire could be “one of the relevant inhibiting factors.”

The intended story seems to be something like this. When someone has conflicting desires, she will have at least two different dispositions set to be triggered in those circumstances. If the manifestations of these dispositions would be jointly incompatible—say, for example, they are two incompatible actions, or two incompatible phenomenological states—then only some of them can actually be produced, thereby preempting or masking the others. The vegetarian who is craving meat may both be disposed to order the meaty dish *and* disposed to abstain. The man on the date may be disposed to have his attention directed toward the time *and* disposed to have it devoted to his conversational partner. Audi’s concerned citizen may both be disposed to be pleased when the library’s budget is increased *and* disposed to be displeased when available funds are not used for the museum. Each of these dispositions is partly constitutive of a different desire. At most one of the desires in each person can win. According to the masking story, the winning disposition (and hence the winning desire) prevents the losing disposition (and desire) from producing the relevant manifestation.<sup>19</sup>

### 3.2 Wanting More (or Less)

While the appeal to masking is appropriate in many cases, this strategy cannot be fully general. To show this and its implications, I wish to draw out a further analogy between

---

<sup>19</sup> While the masking story provides us *something* of a deeper structure for what goes on when desires conflict, it isn’t all that much deeper. For, we are *immediately* faced with the question of why one of these dispositions should be stronger than another—why, in some given case, the disposition to  $\psi$  preempts the disposition to  $\phi$ , rather than *vice versa*. Without a story about what grounds the relative “strengths” of these underlying dispositions, we will have run aground remarkably fast. But I will not insist that such a story cannot be told.

desires and commonplace dispositions involving the way in which they come in varying degrees.<sup>20</sup>

Suppose we learn that some knickknack was fragile—it was knocked off a shelf and is now shattered to bits on the floor. If the object broke when dropped from the distance of five feet, it is reasonable to infer that it was also disposed to break when dropped from twice or three times that distance—or when run over by a steam roller. Other inferences, in contrast, are plainly unacceptable. While the presence of some dispositions will dramatically raise the probability of the presence of others, these relations are often not symmetrical. These dispositions often have some degree of independence from one another.<sup>21</sup> We could not justifiably infer that it was disposed to break when dropped two inches or when jostled by a slight breeze. It is simply not true that a fragile object must be at least *somewhat* disposed to break in such circumstances. While some fragile trinkets may be disposed to break at the slightest touch, others are not disposed to do so *at all*.

Whether or not an object is fragile is a matter of whether it is disposed to break when impacted by certain ranges of concussive forces. But objects can also be *more* or *less* fragile than others. They can also be more (or less) elastic, soluble, flexible,

---

<sup>20</sup> My thoughts here owe a great deal to an excellent discussion of the gradability of dispositions by Manley and Wasserman (2007). I am, however, somewhat hesitant to accept Manley and Wasserman's analysis of dispositional language in terms of conditionals—in part because I am unconvinced of their ability to fully explain masking and finkish dispositions. As far as our interests are concerned, it could just as well be that dispositional language is ineliminable.

<sup>21</sup> To take another example, in the movie *Star Wars: Episode IV* the Death Star is disposed to explode only given an *extremely* precise missile strike down the vent that leads to the reactor core; it is not at all disposed to explode when struck by a missile a foot to the left of or to the right of that vent. These dispositions are independent of one another.

predictable, or prone to developing a skin. *Ceteris paribus*, the broader the range of stimulus conditions under which the object is disposed to produce the characteristic manifestations (e.g. breaking, stretching and then returning to its original shape, dissolving, forming a skin) the more it has the relevant disposition—the more fragile (elastic, soluble, etc.) it is.<sup>22</sup> Other things being equal, something that is disposed to break when slightly jostled is more fragile than something that isn't.

Much like these commonplace dispositional notions, our folk concept of desire is gradable and allows for comparisons between agents and within subjects. We frequently make comparisons between agents about who wants something more than someone else. We think that Rocky wanted to win more than Apollo, that the child wants to climb the tree more than her parents, that Dennis Kucinich wants there to be universal medical coverage more than John Boehner. Such examples are ubiquitous. We also make comparisons between a single subject's desires. Scrooge wants to be wealthy more than he wants to be loved. I want to visit Lisbon more than I want to maintain a modest credit card balance. Many students want to sit through a lecture about as much as they want to sit through a three-hour-long loop of "Who's on First."

We can account for this key feature of our folk concept of desire in a way that closely parallels the story we just told about commonplace dispositions. Consider the version of dispositionalism according to which, roughly, whether or not an agent desires that  $p$  is a matter of whether she is disposed do various actions were she to think them conducive to  $p$ . It is plausible to think that the dispositions to perform these various actions are somewhat independent of one another. If Hal wants a doughnut, he may be

---

<sup>22</sup> See Manley and Wasserman (2007, 72-73).

disposed to run to the break room or to spend a few dollars were he to think that doing either of these things would help him get a doughnut. He can be disposed to do either (or both) of these things without being disposed to, say, walk across hot coals when he thinks there is a free doughnut on the other side.

If one is sympathetic to behavioral dispositionalism, one should find it plausible to think that *ceteris paribus*, the more actions that an agent is disposed to perform in service of obtaining *p*—the wider the range of circumstances in which he would engage in *p*-promoting behavior—the more he desires that *p*. Other things being equal, someone who is disposed to walk across hot coals to get a doughnut wants one more than someone who is not so disposed. We find plausible analogies for other versions of dispositionalism as well. Other things being equal, the wider variety of situations in which she is disposed to be pleased that *p* and displeased that *not-p*, or disposed to have her attention directed toward *p*-related facts, the more she desires that *p*.

There is a further, related analogy between commonplace dispositions and desires that is also worth noting. Commonplace dispositions can often manifest in more or less “dramatic” ways. For instance, some fragile things merely break in two when struck with a given force while others would shatter into thousands of tiny shards. *Ceteris paribus*, the more dramatically an object is disposed to break, given a particular stimulus condition, the more fragile it is (Manley & Wasserman 2007, 73). Similarly, the more quickly and completely an object is disposed to dissolve in a given amount of a particular solvent (at a particular temperature, pressure, etc.), the more soluble it is. And so on.

Many of the phenomena that dispositionalists cite as manifestations of desire can similarly occur in more or less “dramatic” ways: one occurrent phenomenological state

regarding  $p$  may feel more or less urgent than another, one can be more or less (un)happy or (dis)satisfied about the fact that (not)  $p$ , one's attention can be pulled more or less completely toward signs that  $p$ , and so on. Once again, *ceteris paribus*, the more dramatically one is disposed to produce the relevant manifestation given a particular triggering condition, the more one wants that  $p$ .

### 3.3 Wanting *Simpliciter*

I have been drawing out the way in which both desires and workaday dispositions are gradable. One interesting aspect of the analogy is that it is not tied to any particular theory about which events are desires' characteristic manifestations—this is a general feature of desire that any account should satisfy. This picture also suggests that any dispositionalist account of wanting *simpliciter* should take a particular shape:

***Moderate Dispositionalism:***

For all agents  $S$ , and all propositions  $p$ ,  $S$  desires (*simpliciter*) that  $p$  if and only if for a *sufficiently extensive* (but not necessarily complete) class of characteristic manifestations  $M$  and relevant triggering conditions  $C$ ,  $S$  is disposed to  $M$  when  $C$ .

On this view, an individual may count as desiring that  $p$  while being disposed to produce less than the full range of characteristic manifestations. To render this account precise, we would need a story to tell about what it takes for a set of dispositions to count as *sufficient*. One suggestion is that for each proposition  $p$  that could be the content of desire, there is some privileged class  $\mathcal{P}$  of manifestations and triggering circumstances of the appropriate types. For each manifestation/triggering condition ( $M/C$ ) pair in  $\mathcal{P}$ , being disposed to  $M$  when  $C$  is a necessary condition for desiring that  $p$ ; being disposed to  $M$

when  $C$  for *all* of the  $M/C$  pairs in  $\mathcal{P}$  is sufficient for desiring that  $p$ . Thus, according to this suggestion, there is some precise threshold that distinguishes desiring that  $p$  from not desiring that  $p$ .

Upon further reflection on the general analogy that supported *Moderate Dispositionalism*, it is clear that the threshold view is problematic. An analogue of the threshold view for fragility would maintain that there is some privileged class of concussive forces that would break any fragile object and no non-fragile objects would break when impacted by all the forces in that class. This is implausible because even though some concussive forces could break any fragile object—for instance, the force produced three feet away from a nuclear explosion—it is implausible that there is a privileged set that can supply both necessary and sufficient conditions for fragility. After all, many non-fragile things would also shatter in a nuclear explosion. Any choice we make regarding the lower boundary is bound to look arbitrary.

Similar thoughts apply to desire. Suppose the following is an absolute principle of wanting: unless  $S$  is disposed to accept  $x$  when it is offered with no strings attached  $S$  does not really want  $x$ . Even if this were a fixed point, the chance of finding a privileged set of dispositions that provides both necessary and sufficient conditions looks bleak. People might satisfy this condition with respect to  $x$  while not wanting  $x$ : someone might be willing to accept a doughnut with no strings attached even if they didn't want one—for instance, if they thought it would be impolite to refuse. Moreover, there is no non-arbitrary lower bound. Must a subject be disposed pay \$1.50 for a doughnut in order to count as wanting one? Must she be willing to pay a dime? A hair from the top of her head? Any cost placed on a doughnut could be a potential deal breaker, at least in

principle. If *Moderate Dispositionalism* is correct, there is likely no precise boundary between wanting something and not wanting it.<sup>23</sup>

An alternative to *Moderate Dispositionalism* requires that the subject embody all of the relevant dispositions (of the appropriate type) in order to count as really wanting something. The view might be stated more precisely as follows:

***Radical Dispositionalism:***

For all agents *S*, and all propositions *p*, *S* desires (*simpliciter*) that *p* if and only if for *every* characteristic manifestation *M* and relevant triggering conditions *C*, *S* is disposed to *M* when *C*.

*Radical Dispositionalism* is just a version of the threshold view that takes the maximal set of relevant dispositions as privileged. This view has at least two major disadvantages. First, *Radical Dispositionalism* is incompatible with the intuitive account of the gradability of desire that grew out of the analogy to commonsense dispositions. If two people both want *p*, according to the radical view, both must embody the full range of relevant dispositions—in which case we need some metric by which we can measure who wants it more.

A second problem is that *Radical Dispositionalism* cannot vindicate the ordinary practice by which we come to make and accept desire attributions. Take, for instance, the corresponding thesis for the standard version of behavioral dispositionalism:

***Radical Behavioral Dispositionalism:***

For all agents *S*, and all propositions *p*, *S* desires (*simpliciter*) that *p* if and only if for *any* action  $\phi$ , *S* is disposed to  $\phi$  when she believes that  $\phi$ -ing will bring it about or make it more likely that *p*.

---

<sup>23</sup> If there is any boundary, it is likely to be a context sensitive one. For issues involving vagueness and context-sensitivity of dispositions, generally, see Prior (1985, 8-9), Fara (2005), and Manley and Wasserman (2007, 73).

*Radical Behavioral Dispositionalism* cannot vindicate the reliability of our ordinary talk about desire. We attribute desires to others on the basis extremely limited behavioral evidence. Hearing Hal assert “I would like a doughnut” while handing money to the cashier is evidence enough to infer that he wants a doughnut. While this inference is not an infallible one, it seems to justify the belief well enough. If this inference is reliable, then so should the weaker inference (according to *Radical Behavioral Dispositionalism*) from this evidence to the conclusion that Hal is disposed to walk across hot coals or to throw someone’s mother off a train when these are means to getting a doughnut. According to the theory under consideration, a miss is as good as a mile: unless he is disposed to do these other things, he does not count as wanting a doughnut. Yet intuitively, this isn’t the sort of thing that we should be able to infer from our meager set of evidence—just as we could not legitimately infer that the knickknack would shatter when slightly jostled from the fact that it did so when struck with a sledgehammer. If we cannot reliably move from central to far-out dispositions, then we cannot reliably infer facts about what agents want either.

A similar argument undermines other versions of *Radical Dispositionalism* as well. If we cannot reliably determine that the agent is disposed to be pleased when she receives a sum of money from the loss of her beloved spouse or child, we cannot safely infer that she wants money. This is surely a cost.

### **3.4 Why Desires Motivate Only Contingently**

I contended earlier (Section 3.2) that the standard explanation for unmanifested desires—the explanation that appeals to masking or inhibiting conditions—cannot be fully general. We are now in a position to see why. Generalized, the standard move amounts to the following claim:

***Masking:***

Whenever an agent *S* desires that *p*, but does not exhibit a characteristic manifestation *M*, despite the fact that she is in a triggering condition *C* of the relevant type, this is because *S*'s disposition to *M* when *C* is masked by the triggering of a stronger, incompatible disposition to *M\** when *C*.

The problem with *Masking* is that it is committed to *Radical Dispositionalism*. For, *Masking* claims that the dispositions to which *Radical Dispositionalism* appeals are the ones that are masked when a desire remains unmanifested.

The picture I have been developing in this chapter suggests another way in which a subject's desires can fail to manifest when triggering conditions of the relevant sort arise. Take our individual who has just found out she has received a great deal of money because of the passing of a beloved spouse or child. If our subject takes no pleasure in this event, and if she does not consider it at all enjoyable or acceptable, we need not insist that she has a masked disposition to be pleased by this state of affairs—it may simply be that she was not even disposed to find it pleasing in the first place. Although she may have desired money, she didn't want it *that* much. According to *Moderate Dispositionalism*, she may have had a sufficiently extensive set of affective or hedonic dispositions without having this particular one. The fact that she does not want the

money that much along with the related fact that she lacks this particular disposition explains why her situation fails to make her happy.<sup>24</sup>

Working our way back to the case of productive behavior, suppose William knows that a particular course of action would help him pay off his credit card debt: he has just been offered the money to throw someone off a train. If William does not even seriously consider the offer, this may be something that he is simply not disposed to do—at least not to pay off his credit card. Although William wants to pay off his credit card debt, he doesn't want to pay it off *that* much and this explains why he does not take the available means. According to *Moderate Dispositionalism*, a subject may have a sufficiently extensive set of dispositions to engage in productive behavior without being disposed to engage in *that* particular productive behavior.

As simple as it is, this insight has an important implication if the *Motivation-Disposition Link* is true. It shows that within even behavioral dispositionalism there is conceptual space for the possibility that an individual desires that *p* while not being disposed to take any of her *actual* options to promote this desire. Thus, if the *Motivation-Disposition Link* is true, there is room for desires that do not actually motivate the agent to try to achieve the desired good in even the weak non-success sense. It could turn out that there are individuals who, while having a sufficiently extensive set of *p*-promoting dispositions, ready to be triggered when she is faced with means she would seriously

---

<sup>24</sup> To avoid misunderstanding, the dispositionalist need not think that masking can never be used to explain unmanifested desires. The point is simply that if the analogy I have been developing is on to something, the dispositionalist should not think that such explanations are always available.

consider, has no such dispositions set to be triggered in her actual situation. This is so even if there are means of achieving  $p$  that she is aware of.

#### 4. Silencing, Continnence, and Motivational Variation

I would like to conclude this chapter by returning to some ideas that we brushed up against toward the end of Chapter 2. I suggested there that we do not need to appeal to essentially motivating states in order to explain motivational variation. Instead, all we need is some theory to explain how states could motivate contingently. In this chapter, I have been arguing that desires only contingently satisfy a condition that is necessary for an agent to be motivated. In other words, I have been arguing that desires motivate only contingently. In this final section, I show that the picture I have been developing can make sense of the range of different motivational profiles that can be embodied by individuals who all desire that  $p$ .

We can begin with the virtuous agent. As John McDowell described her, the virtuous agent is an individual who is invariably moved to act as directed by her moral judgments because she is never even tempted to do something that she takes to be morally impermissible.<sup>25</sup> The dispositionalist account we have been developing has resources to make sense of such an agent: the McDowellian virtuous agent is the kind of agent who is not disposed to do anything she believes to be morally wrong. If so, the *Motivation-Disposition Link* predicts that such an agent is not moved (in even the minimal, non-success sense) to do anything she thinks is wrong.

---

<sup>25</sup> McDowell (1978, 90-93; 1979, 55-56; 1980, 17-18).

The virtuous agent promotes her desires only in ways that are, by her lights, morally acceptable. If there are only one or two permissible ways to promote some wanted thing, the virtuous agent will be motivated to pursue one of those options if she is motivated to satisfy that desire at all. If there are no acceptable ways for achieving the wanted thing, she will not be moved to satisfy that desire at all. To make this point clear, as I argued in Chapter 2, being unmoved in this way is nevertheless compatible with wanting to do the impermissible thing, for this desire can be silenced and lack motivational force. When any way of promoting *p* requires doing something she finds morally objectionable, the virtuous agent's desire that *p* will not actually motivate her to act in even our minimal, non-success sense.

Suppose that the Manchester Estate has just gone on the market and the agent who closes the deal is set to make a sizeable commission. Gil has only one buyer on the line, who would not buy the property if she learned some particular fact about it. We can pretend that it is haunted by the ghost of Benjamin Franklin, who likes to host occasional dinner parties for his otherworldly friends. Gil knows that given his potential buyer's disinterest in having such a roommate, the only way that he is going to close the deal is by lying about or failing to disclose this fact about the house. Yet he also thinks that this would not be appropriate—he would be doing his client a serious disservice and would not be living up to the high ethical standards embodied by the real estate profession.

If Gil were a fully virtuous agent, lying about Franklin's ghost would not be the kind of thing he is willing to consider. Although he wants to close the deal and receive the commission, among the sufficiently extensive set of deal-closing dispositions constituting this desire there is no disposition to do anything he takes to be morally

objectionable. According to the *Motivation-Disposition Link*, he is never motivated to do anything immoral. Since (we are to suppose) there are no permissible ways for Gil to achieve the desired end in his given circumstances, there is nothing he is disposed to do in his present circumstances to achieve that end. Thus, according to the *Motivation-Disposition Link*, he is not motivated by this particular desire.

In Chapter 2, I suggested that the way in which the virtuous agent's desires are silenced by moral concerns is really just a special case of a more general phenomenon. The picture I have developed here can support this claim, and can help us to better understand two important aspects of the more general phenomenon.

First, as far as the more general phenomenon of silencing is concerned, there need not be anything special about morality. Silencing occurs when a range of options are automatically excluded on the basis of some feature or another. For virtuous agents, options are automatically excluded in virtue of being morally impermissible. But there are clearly other factors that might rule out options for other individuals. For egoistic agents, being against one's narrow self-interest could make an option a non-starter. For the slave of fashion, being unhip automatically excludes some alternatives from consideration. For the slothful individual, doing something difficult or inconvenient is not taken as a serious option. According to a more general characterization, in each of these cases, the agent is not disposed to perform any action that has a certain feature,  $x$ . According to the *Motivation-Disposition Link*, the agent is not motivated in even the minimal sense to do things like that. If there are only one or two non- $x$  ways to promote some wanted thing, the agent will be motivated to pursue one of those options if she is

motivated to satisfy that desire at all. If there are no non- $x$  ways to achieve the wanted thing, she will not be moved to satisfy that desire at all.

Second, the picture I have developed helps us to better understand the contrast McDowell draws between the virtuous agent and the merely continent or strong-willed agent—a contrast that is also mirrored in the other cases. Like the virtuous agent, the continent agent, as McDowell (1979, 53-57) portrays her, will uniformly do what is morally required by her lights. However, unlike the virtuous agent, the continent agent is tempted by her competing desires. The continent agent is moved by these desires in our minimal sense; however, because of her strength of will, she is never moved by them in the success sense. Suppose that Willow, unlike Gil, is tempted to paper over the ghost infestation at the Manchester Estate, but ultimately gathers her strength to resist the temptation. According to the *Motivation-Disposition Link*, if she is moved to conceal the facts about Franklin's spirit to her client in even this minimal sense, she must be disposed in her actual situation to do so, or at least to intend to do so at some later time. In Section 1.3 I suggested that cases in which an agent is torn between options should be dealt with by appealing to masking. We can thus see the difference between strength of will and motivational silencing is the difference between having a masked disposition to perform a particular type of act and having no disposition to do it at all.

Of course actual agents are neither completely virtuous nor completely continent. Yet our actual motivational profiles contain some similarities to both kinds of agents. I find it plausible that normal agents rule out a wide-variety of ways of achieving their desires, without any serious thought. We may sometimes be tempted to commit minor moral infractions for the sake of personal benefit (and may or may not give in to that

temptation), but few of us are ever tempted to do things that we consider seriously immoral. Or, if even that is too optimistic, there are at least some seriously immoral behaviors that we are not even tempted to do. If this is right, we might think that our desires are incompletely silenced by moral considerations. We may be disposed to perform some acts that we consider only a little bad or a little wrong, or we may even be disposed to perform some significantly bad actions for a high enough benefit. But in most situations, a significant range of options will be ruled out from the very beginning, in part, on the basis of moral considerations. This is, once again, mirrored in the more general picture too. It is not only moral considerations that can play the kind of role in question—matters of convenience and fashion, for instance, commonly preclude a range of behaviors even if they do not rule out all inconvenient or unhip behaviors.

My discussion here has admittedly been sketchy and has left important questions unanswered. For instance, I have said nothing about how we should explain differences in the underlying distribution of behavioral dispositions that my picture assumes exist. The broad-brush story I have provided nevertheless gives us some hope that it is possible to give an intuitive account of motivational variation without appealing to essentially or necessarily motivating states. In Chapter 4, I try to explain how the story I have presented here can be accommodated by a version of motivational cognitivism.

## Chapter 4

### Cognitivism as an Empirical Hypothesis

The debate between Humeans and motivational cognitivists is frequently understood as a conceptual matter, to be settled by *a priori* argument. In this chapter, I develop a picture according to which the possibility of cognitive motivation is an *empirical* hypothesis. The motivation for this view consists of two elements. I begin in Section 1 by providing an argument in favor of hybrid accounts of desire over those that single out one of the phenomena related to desire as *the* characteristic manifestation. This account, I contend, is best understood as positing a natural kind that helps to unify the various dispositions. In Section 2, I argue that on this account of desire, the Humean Theory is understood as making an empirically evaluable claim about the law-like role that this natural kind plays in producing action—claims that the cognitivist is skeptical about. Sections 3-5 add detail to the cognitivist proposal, by placing it in the context of two more familiar issues in the dialectic between cognitivists and Humeans—Nagel’s insight that there is a key ambiguity in the ordinary concept of desire that makes the Humean Theory appear more compelling than it really is, and the claim forwarded by many cognitivists to the effect that all action is pursued under the guise of the good.

#### 1. A Preliminary Account of Desire

##### 1.1 Characteristic Manifestations of Desire Redux

In Chapter 3, I argued that the most plausible dispositionalist account of desire would take a particular shape—it would be a version of *Moderate Dispositionalism*:

***Moderate Dispositionalism:***

For all agents *S*, and all propositions *p*, *S* desires (*simpliciter*) that *p* if and only if for a *sufficiently extensive* (but not necessarily complete) class of characteristic manifestations *M* and relevant triggering conditions *C*, *S* is disposed to *M* when *C*.

My discussion thus far was largely neutral on the question of how best to fill out the details of this account. More precisely, while I noted a range of options, my discussion in Chapter 3 did not come down on the question of what really are desire's characteristic manifestations. I take up this issue presently.

As we mentioned in Chapter 3, a standard position appeals to behaviors—productive behaviors, in particular—as desire's characteristic manifestations. We already noted a number of problems for this position. To briefly rehearse them, this view has trouble accounting for desires with necessarily true or necessarily false objects, desires for things beyond the agent's control, and certain widespread desires that are not normally promoted by the agent (even though the agent could in theory try to promote them). For this reason, there was pressure to modify this view, to include non-productive behaviors among desire's characteristic manifestations.<sup>1</sup>

This more ecumenical version of behavioral dispositionalism is also subject to objections. The presence of behavioral dispositions does not appear to be sufficient for desiring. In *The Manchurian Candidate*, Raymond Shaw has been brainwashed to become a political assassin. Given his programming, Shaw is now disposed to kill a

---

<sup>1</sup> Michael Smith (1998, 450) does just this—he thinks that dispositionalists should include “all the possible behavioural evidence that would bear upon whether someone desires that [*p*]” among a desire that *p*'s characteristic manifestations.

specified individual when certain conditions materialize. Despite the fact that he is so disposed, it is a stretch to say that Shaw presently *wants* to assassinate the politician.<sup>2</sup>

In a more general assault on behavioral dispositionalism, Galen Strawson (1994, Ch. 9) has argued that the typical behavioral dispositions are entirely unnecessary for desire. Strawson's argument involves a hypothetical scenario in which a species of creatures intuitively want the weather to turn out in a certain way despite the fact that they lack *any* of the relevant behavioral dispositions. The Weather Watchers, we are told to imagine, pay close attention to changes in the weather. They are quite happy on days when it is seventy-five degrees and sunny. Their mood also improves when it seems to them like the weather will soon be the way they enjoy so much. On the other hand, when it is cool and gloomy (or when it seems like it will be shortly) the Weather Watchers tend to be thoroughly displeased.

As these creatures are described, many of us find it plausible to say that the Weather Watchers want it to be sunny and seventy-five degrees. Yet as Strawson (1994, 254) is keen to notice, what we have just said about the Weather Watchers is compatible with them being completely incapable of acting: we can imagine that they are like mossy boulders or trees firmly rooted to the ground and even that they completely lack the concept of action.<sup>3</sup> The force of this insight is that the Weather Watchers' dispositions to be (dis)satisfied or (dis)pleased by certain perceived outcomes, along with their dispositions to pay attention to signs of those outcomes, seems like sufficient reason to say that they want the weather to turn out one way rather than another. The Weather

---

<sup>2</sup> Mark van Roojen (2002, 213) makes a similar point about hypnosis.

<sup>3</sup> It must be noted that if the subject does not even possess the concept of action, she would likely not have action desires in Alfred Mele's (1995) sense.

Watchers' lack of behavioral dispositions does not cancel this. From this it seems to follow that dispositions to behave in the relevant ways are not necessary for our ordinary concept of desire to apply.

If there is no essential connection between desire and behavioral dispositions, it is tempting to think that there must be an essential connection between desire and one of the other phenomena that are sometimes posited as desire's characteristic manifestations. The problem cases for behavioral dispositionalism might suggest, for instance, that the essential connection is between desire and happiness, pleasure, and satisfaction (or unhappiness, displeasure, and dissatisfaction). For, it is the fact that the Weather Watchers are pleased by warm sunny days (and the prospect of them) that is why we attribute desires to them. Were there no correlation between changes in the weather and changes in the Weather Watchers' affective and hedonic states, we would not think that they want the weather to be one way or another. Notice that this connection to affective and hedonic states is also plausibly absent in Shaw—the brainwashed assassin of *The Manchurian Candidate*. Insofar as he is normal, the violent death of a politician is not likely to please Shaw. Indeed, he would likely be quite displeased were he to discover that the politician was to meet an untimely demise at his own hands. This connection to affective and hedonic states may partly explain why Shaw's behavioral dispositions are insufficient grounds for attributing the desire to him.

This is too quick. If someone can be brainwashed to commit a certain act, perhaps he could also be programmed to enjoy doing it or to have positive affect when the thought of committing the murder randomly crosses his mind. Even so, we may not want to attribute a desire to him—the hypnotist can make him do something and make

him enjoy doing it without thereby making him want to do it.<sup>4</sup> In reality, there are many things that are known to be enjoyable or pleasurable that some people never the less do not want—rumors have it that shooting heroin is quite nice, but a large number of people have no desire whatsoever to shoot up. Affective and hedonic dispositions are also insufficient for desire.

Nor are these dispositions necessary. It is in line with our ordinary concept of desire that agents can want things that they will not ultimately find pleasant. An agent can even want something that she knows (or at least strongly believes) she will not enjoy. Not too long ago, I found myself having a strong craving for beets even though I had always found them disgusting in childhood. At the time I had the craving, I could no longer remember what beets actually tasted like. I never the less had very good reason to think that I would not enjoy them. Eventually, I gave in to the urge. It was abundantly clear that I had not outgrown my childhood tastes. I was thoroughly dissatisfied—as I suspected I might be.

My experience is not so uncommon. In the most extreme cases people can be moved to eat even less appetizing items: individuals with pica crave (and eat) things like earth, raw corn starch, chalk, hair, charcoal, and even feces.<sup>5</sup> Even though individuals with pica will often enjoy the strange objects of their craving, it is clear that according to our ordinary concept, the individuals in question desire the things that they are craving, regardless of whether or not they are find them pleasing in this way.

---

<sup>4</sup> However, someone could be hypnotized to *desire* something or hypnotized to *want to do* something. It is an interesting question what the difference would be between being hypnotized to have the desire and being hypnotized to act and to enjoy it.

<sup>5</sup> For a fascinating survey of many scientific and cultural issues related to pica see Sera L. Young (2011).

When considering cases like this, it seems like the attendant phenomenology—the thing I have previously called “felt urgency”—and the dispositions to act on it are jointly sufficient for our ordinary concept of desire to apply. We could offer a variation on the Weather Watchers case that isolates this felt urgency. Suppose that it has not snowed for years. Despite the fact that he had always hated snow in the past, one particular Weather Watcher finds himself craving a blizzard in our strong phenomenological way. As it turns out, blizzards are no more enjoyable than they used to be, and this Weather Watcher is not disposed to enjoy himself at all were a blizzard to blow ashore. But this does not seem important—as long as he is craving a blizzard in this strong phenomenological way, this Weather Watcher wants there to be a blizzard. The connection to affective dispositions, hedonic dispositions, and behavioral dispositions is not required.

In this felt urgency, it appears that we have found one thing that may be sufficient for desire. Even if this is right, it should be clear that this strong phenomenological craving is not necessary for wanting something. One might quite mindlessly munch on a snack while watching a movie or buy a new pair of jeans on a whim without either of these things seeming at all urgent. In such cases, it is perfectly acceptable to say that the agent wanted to eat the snack or to buy the jeans anyway. Even if we have found a manifestation that is sufficient for desire, we have not found any manifestation that is necessary.

## **1.2 The Hybrid Account of Desire**

There are a number of things that we might say about this philosophical dialectic. We could say that the ordinary concept of desire seems irredeemably confused. In everyday

deployment of the term “desire” we seem to muddle together and slide between three very different concepts—one corresponding to each of the relevant type of dispositions. While the ordinary concept of desire is no doubt somewhat confused, it is surely too pessimistic to insist that we cannot make better sense of it than to posit three disconnected senses of “desire” and leave things at that.

We might instead try to resist some stage of the dialectic by rejecting some of the cases as *misuses* of the ordinary concept of desire. We might even provide an error theory for why we would misapply the concept in those cases. For instance, one might argue that if the Weather Watchers are really not at all disposed to act in any way then they do not *really* desire anything—perhaps when we attribute desires to them, we are not taking this stipulated aspect of the case seriously, and our intuitions are clouded by the shadows of our illicit behavioral assumptions.<sup>6</sup> Perhaps this is right; but I see little hope of making this case. After all, defenders of the alternative conceptions of desire can say much the same thing in defense of their preferred accounts. In order to settle this dispute, one would need to say much more to canvas all of the possible theories and corresponding error theories and determine which makes best sense of all the available data.<sup>7</sup> This is a daunting task that I cannot pretend to undertake here.

I can however, say this much. While this philosophical dialectic tells us something important about the ordinary concept of desire, it is crucial to recognize just

---

<sup>6</sup> This point was inspired by several conversations about this case with David Chavez.

<sup>7</sup> Although one particularly difficult piece of data for any such account is the way in which the account of the gradability of desire I developed in Chapter 3 works uniformly well for all of these characteristic manifestations of desire—for each class of characteristic manifestations, other things being equal, the wider the range of situations in which an appropriate manifestation of that type would be produced, the more the subject desires that *p*.

how artificial it is. Most people are capable of acting and do so on a regular basis—unlike the Weather Watchers, *we* are not rooted to the ground, unable to influence the world as we will. Nor are we brainwashed to do (or enjoy doing) things that we do not want to do. If we think of behavior, pleasure, happiness, felt urgency, and the rest only as isolatable phenomena, we may lose track of the fact that they regularly go hand in hand. It is this systematic connection that lends credence to the thought that these are all manifestations of a single thing—of *desire*. It is natural, then, to consider the possibility that, as far as our ordinary concept is concerned, desire is something of a syndrome, best captured by a hybrid account that ties these various manifestations together in a non-accidental way.

I suspect that this idea is really already implicit in the intuitions that support dispositionalism in the first place. Dispositionalism is supposed to provide an explanation for, *inter alia*, an agent's actions by reference to what she wants. At the same time, according to the basic structure of the dispositional account, what she wants—that is, the contents of her desires—is a function of what dispositions she has. Smith's (1998, 449) version of dispositionalism, you might recall, was committed to the idea that the contents of a subject's desires are "determined by the ways in which she behaves in possible worlds in which her beliefs are true." According to the basic structure of this picture, the entire constellation of the agent's dispositions are ontologically prior to what she wants. It is from this constellation that the contents of the agent's desires are derived. But how does this derivation work?

Let us call the function mapping bundles of dispositions onto desires a "content function". We can think of content functions as attempts to make sense of the manifold

of an agent's various dispositions by organizing them within an overarching narrative. This narrative will "bundle" certain classes of dispositions together for the purposes of attributing purposes and desires to the agent. For, when an agent's behavior is explained within the dispositional framework, these explanations must be more than a series of one-off descriptions of how that person will act in certain precisely defined situations. In order to understand someone as having a particular purpose or goal in mind that explains their behavior, we must see that purpose or goal as constraining her actions in other relevantly similar situations as well. It is this feature that is supposed to help us distinguish the purposes of the action from other, accidental features of the action (such as known side-effects).

For a given set of behavioral dispositions, there will often be several competing content functions, all of which are compatible with the data—there will typically be a number of different ways to "bundle" them for the purposes of attributing various purposes to the agent. One way to decide between competing narratives is to expand the range of dispositions that must be accounted for. In particular, we might require that a content function accounts, as far as possible, for dispositions to produce the other characteristic manifestations of desire. Other things being equal, a narrative that provides a unifying story about an agent's dispositions to go to the country club, to find pleasure in taking a few swings, to be displeased in coming to learn that his favorite nine holes are closed, to have thoughts about hitting a round of golf feel urgent to him, and so on, will be more compelling than a function that leaves these different elements unconnected.

It is a deep part of how we attribute desires and purposes to others that we systematically calibrate our interpretations of what people are up to when they act by

lining these interpretations up with, among other things, facts about the kinds of things that we know they like or enjoy. Suppose that we know Jerry well enough to know that he is not the kind of person who tends to enjoy or value learning for its own sake and that he finds learning new languages especially frustrating and not enjoyable. When we find out that Jerry is trying to learn French, we automatically look for a deeper explanation to account for this. In other words, we do not interpret him as trying to learn a new language for its own sake, but automatically look for some ulterior motive or purpose for his behavior. Is he trying to get a job in Quebec? Is he planning a trip to France? Is he trying to impress his new multi-lingual neighbor? Our knowledge of the non-behavioral facts clearly influences the way we interpret the behavioral facts.

Can we say anything more here? It is not mere happenstance that our practices of desire attribution commonly tie together these various dispositions. After all, they tend to come together in predictable ways. While the various manifestations of desire can be isolated from one another in certain cases, the fact that an agent is disposed to enjoy something increases the probability that she is disposed to pursue it, and vice versa. It is also more likely that she is disposed to be dissatisfied with not getting it, to think about it and to have her attention directed toward it, to have it feel urgent to her that she gets it, and so on. The presence of one of these dispositions makes it more probable that other related dispositions will also be possessed.

The fact that these dispositions are probabilistically connected in these ways provides *prima facie* evidence for the hypothesis that desires constitute a natural kind. More specifically, this suggests that desires likely constitute a *homeostatic property cluster* kind (or “HPC”), in the sense explained by Richard Boyd (1988; 1991; 1999).

According to Boyd's (1988, 197) account, an HPC is a natural kind that is associated with a set of properties, the co-occurrence of which is underwritten by some further unifying explanation: "Either the presence of some of the properties [...] tends (under appropriate conditions) to favor the presence of the others, or there are underlying mechanisms or processes which tend to maintain [their joint] presence [...], or both." It is a plausible hypothesis that unifying explanations of both sorts can be found to explain the co-occurrence of many of the dispositions associated with desire.

First, it is plausible that many of the relevant dispositions reinforce certain others. Consider the disposition to have it feel urgent to an agent that she gets  $x$ , and the relationships that this disposition might bear to others. Having a phenomenal experience of this sort associated with  $x$  is likely to make it easier for the agent's attention to be directed toward  $x$ . This felt urgency can also be uncomfortable, thus adding to the agent's frustration and displeasure in not having  $x$ ; we also find it pleasurable to have this phenomenology lessened or removed.<sup>8</sup> Other possible cases of mutual support among the various dispositions are also plausible, although this may be more speculative. For instance, the fact that an agent finds pleasure in something might reinforce her disposition to pursue it or to frequently think about it. Even the agent's tendency to have her attention directed toward an object might itself reinforce her disposition to pursue it—if only by increasing the likelihood that she will notice means to attaining it that would have otherwise gone unnoticed.<sup>9</sup> For these reasons, there may be sufficient connections

---

<sup>8</sup> I made this point earlier (see Chapter 3, Section 2.3 above—especially note 13) in connection to Stephen Schiffer's (1976) discussion of the link between reason-providing desires and (dis)pleasure.

<sup>9</sup> Recall Mark Schroeder's (2007, 156-157) hybrid account of desire that included both behavioral and attention-related dispositions in just this way.

among the various dispositions to create the kind of “homeostasis” that Boyd thinks gives rise to an HPC.

Second, it is also a plausible hypothesis that many of the various phenomena associated with desire are explained by common processes and mechanisms—at least in typical human agents. Recent work by Timothy Schroeder (2004) can be understood as an attempt to support such a hypothesis.<sup>10</sup> According to Schroeder, human desire (and mammalian desire in general, it appears) is housed in the brain’s reward system, which consists of a specific set of neural structures that are deeply involved with reward-based learning. In addition to this connection to learning, these structures (and the learning signal they transmit) also have many other effects commonly associated with reward and desire, including the way in which they are connected to pleasure, motivation, and certain cognitive tendencies (T. Schroeder 2004, 35-37, 53-54). Schroeder (2004, 54), understands this neuroscientific picture as “explaining the mechanism by which these effects are brought about.”

There are reasons to be skeptical of the details of Schroeder’s account. In particular, it strikes me that this picture places too much of an emphasis on role of reward-based learning. One problem with this is that people can be “open to” receiving something (in Fred Dretske’s (1988, Ch. 5) terms), and thus treat it as a reward, even if they do not initially want it. For instance, we can be rewarded by receiving something that is entirely novel to us—if behavior can be reinforced through completely novel rewards, then the subject’s antecedent desires seem to play no role.

---

<sup>10</sup> Also see Timothy Schroeder, Adina Roskies, and Shaun Nichols (2010).

More generally, the “rewards” that are sufficient to trigger the learning signal transmitted by the reward system, are not always plausibly tied to our desires. Not surprisingly, being told whether one’s answer is “Right” or “Wrong” can help people learn. This type of correction is frequently considered a type of “reward” or “punishment” of the sort that those studying reward-based learning have long-been interested in.<sup>11</sup> Some of the experiments that Schroeder discusses to support his view employ precisely this type of “reward”. A response of “mmm-hmm” by an experimenter and immediate feedback about whether one’s prediction was correct are “rewards” that Schroeder (2004, 48 & 52, respectively) notes can help subjects learn arbitrary rules, and this learning seems to depend crucially on a healthy reward system, in his sense.<sup>12</sup> It is difficult to see this type of feedback as the kind of “reward” that commonsense links to desire. The worry is that the reward system may be implicated *too broadly* in human learning to be plausibly identified with the mechanism that underwrites human desire specifically.

For my purposes, I only wish to point out that, in its most general terms Schroeder’s picture is the kind of story that promises to provide a unifying explanation in terms of an underlying neurophysiological and neurochemical mechanism. If some account similar to this—and/or one that posits a homeostatic mechanism like the one discussed a few paragraphs ago—is correct, which is an entirely *a posteriori* matter, then desire is an HPC.

---

<sup>11</sup> See, for instance, Edward L. Thorndike and Robert T. Rock Jr. (1934).

<sup>12</sup> The experimental results are from Joel Greenspoon (1955) and Barbara J. Knowlton, Jennifer A. Mangels, and Larry R. Squire (1996), respectively.

## 2. Two Anti-Humean Empirical Hypotheses

### 2.1 The Negative Hypothesis

If the account of Section 1.2 is correct, the Humean Theory amounts, roughly, to the claim that motivation is tied in some necessary way—at least in human agents—to the specific natural kind associated with desire. Whether or not the Humean Theory was originally intended as an *a priori* truth, the account of desire I have just articulated suggests that this account makes empirically testable predictions about actual kinds and their role in producing motivated action. For our purposes, we can treat the Humean as claiming that, as a law of human psychology, the human motivational system is only activated by the desire HPC. The cognitivist is less sanguine than the Humean that these empirical predictions will pan out.

The reason for the cognitivist's pessimism regarding this connection between human motivation and desire is that there is a recognizable agent who acts in ways that are not obviously connected to affect, pleasure, happiness, and the rest. Consider the following familiar case.<sup>13</sup> Martha is supposed to go to a meeting this afternoon for the University's Parking Advisory Committee—a committee she had agreed to serve on for the year. The meetings are always extremely boring and the other members of the committee are unpleasant. Moreover, Martha lives only a short walk from campus, and so the committee work has no impact on her daily life. She is tempted to skip it and work on her book manuscript instead. All in all, she would much rather not go to the meeting.

---

<sup>13</sup> My immediate inspiration for this example is from G.F. Schueler (1995, 29), who uses going to a meeting at his son's school as a mundane example of something that one might do, despite not wanting to. In the *Groundwork of the Metaphysics of Morals*, Immanuel Kant (1785, 4:398-399) offers a more famous case that is supposed to get at the same issues: the cold-hearted or grief-stricken philanthropist who provides assistance to others, not because of self-love or even a natural sympathy for others, but only because it is his duty.

In fact, as far as she is concerned, she had no desire whatsoever to go. But since she agreed to serve, Martha decides that she really should go anyway, whether she wants to or not. So she leaves her office and heads off toward the Administration building, where the meeting is being held.

The motivational cognitivist thinks it likely that the begrudgingly moral agent acts without having a desire of the relevant sort to act in that way. She goes to the meeting simply because she believes that is what she is supposed to do—because she has a responsibility to go to the meeting and to do the work that she has agreed to take on. In Martha's case, we are lacking the typical connections to pleasure, happiness, etc. that are present in the typical case of desire-induced behavior. This lends some credibility to the cognitivist's suspicion that there is no desire present in her case.

To avoid a potential misunderstanding, the cognitivist is not suggesting that the absence of the connection between an agent's actions and pleasure, attention, and the rest automatically entails that the agent is motivated by something other than desire. For, according to Boyd's (1988, 197; 1991, 141-142) account (and *Moderate Dispositionalism*, for that matter), the desire HPC may still be involved even though not all of the associated dispositions are possessed—the homeostasis between these various dispositions needn't be perfect in order to constitute an HPC.<sup>14</sup> The point is, rather, that given *how few* of the associated dispositions are possessed by the begrudgingly moral

---

<sup>14</sup> This is why HPCs are different from natural kinds defined by sets of necessary and sufficient conditions (Boyd 1991, 127-128).

agent, it is reasonable to wonder whether the behavioral dispositions are in this case the result of some mechanism other than the one associated with the desire HPC.<sup>15</sup>

The Humean has alternative hypotheses about cases like this. Even if Martha doesn't want to go to the meeting for its own sake or for the fun of it, she may nevertheless want to go for other reasons. In fact, she may very well desire *not* to go. Desiring not to go and having no *intrinsic* desire to go are both compatible with having some other *extrinsic* desire to go. Maybe she wants to keep her job. Or perhaps her motives are more benevolent—she may want the university to function effectively and to make improvements that will help the next generation of students and faculty. Or maybe she just has some fetish for doing the morally responsible thing. In any case, it is hasty to say that she has no desire whatsoever that explains why she goes to the meeting.

There are two things that the cognitivist can say in response. First, the cognitivist I am imagining does not insist that there couldn't *possibly* be another explanation—such as those offered by the Humean. The cognitivist has simply offered one possible explanation that she thinks should be taken seriously. According to my cognitivist, we cannot set this possibility aside without good reason. Second, even these further motives that the Humean attributes to Martha are not necessarily connected to affective and

---

<sup>15</sup> One further point is worth making here. I suggested that it is plausible that there is a Boyd-like homeostatic mechanism by which various dispositions associated with desire might reinforce one another (see Section 1.2 above). As part of this discussion, I offered the speculative suggestion that some of these dispositions (e.g. dispositions to find *x* pleasurable) might help reinforce the behavioral dispositions to promote *x*. However, it is difficult to find plausible ways by which the behavioral dispositions could provide causal support for the non-behavioral dispositions. If so, then the behavioral dispositions may have a relatively unique position in the desire HPC. More importantly, this makes it more plausible that these behavioral dispositions can exist in the absence of the desire HPC, which is what the cognitivist wants to say. If the behavioral dispositions generally helped to reinforce the non-behavioral dispositions, then we would suspect that even the behavioral dispositions of the begrudgingly moral agent would also help to reinforce those other dispositions. If that were the case, her actions would likely be the result of the desire HPC.

hedonic states, directed attention, or the strong phenomenology of a craving. It could be that the only thing that feels urgent to her is her desire to bail on the meeting. It may be that she is the type of person who grumbles through her job day after day. It may turn out that she would be much happier with a life that isn't devoted to constant production and committee work. She might be much happier as a professional dog trainer. Indeed, she may even daydream about this from time to time! Nor need she be disposed to be pleased by being morally responsible. After all, it can be a drag to always do the right thing. Thus even these other potential motives may fail to be appropriately connected to a sufficient number of desire's characteristic manifestations. We have no more reason to think that these further explanations would ground out in the desire HPC than we have to think that the motive to go to the meeting is.

## **2.2 The Positive Hypothesis**

The cognitivist hypothesis we just discussed was a negative one—that some purposeful action is the result of processes that do not involve whatever natural kind (if any) happens to underlie the phenomena typically associated with desire. It is also possible to forward a positive cognitivist hypothesis. This hypothesis involves the further assumption that there are natural cognitive kinds—including judgment, belief, experience, and perception, among others. Characterized positively, the cognitivist hypothesis is that purposeful action sometimes has its origin wholly in one or more of these cognitive kinds.

The cognitivist may be optimistic about this positive hypothesis because of cases where agents, like Martha, are moved by moral considerations. In Martha's case it is plausible that her normative belief about what she ought to do, coupled with the

instrumental belief about what it takes to do that, accounts for why she heads off toward the administration building.<sup>16</sup> The structure of Martha's reasoning gives credence to the claim that there is a mental process that begins in a motivating belief (in this case Martha's normative belief), is guided by her instrumental judgments (her beliefs regarding how to promote the ends towards which her normative judgment direct her), and concludes in the behavior to be explained. That is some reason to think that the cognitivist's optimism is warranted.

The cognitivist's positive hypothesis is more useful than the negative hypothesis. First, one might consider it an open empirical possibility that there is no deep difference between the cognitive and the passionate sides of the mind. Desires could turn out to be a special case of some broader cognitive kind.<sup>17</sup> If all purposeful action is the product of processes that involve the natural kind that we identify as desire, but this natural kind is itself a special case of some cognitive kind—say belief or perception—then the positive cognitive hypothesis is vindicated while the negative hypothesis is not.<sup>18</sup>

A second reason why the positive characterization is more useful is that it provides a framework in which it makes sense to ask questions about motivational overdetermination. Our present discussion has relied heavily on the case of Martha—the begrudgingly responsible agent. There is good reason for the cognitivist to be uncomfortable relying exclusively on cases like this. For, this focus gives the appearance

---

<sup>16</sup> The cases discussed in Chapter 1 are also directly relevant.

<sup>17</sup> I suppose that someone might defend the claim that all beliefs, judgments, and the rest are all special cases of some desiderative kind. This type of claim has been defended about certain kinds of judgment—evaluative beliefs, in particular—but non-cognitivism about all of the mental does not strike me as particularly plausible. I will thus set this possibility aside.

<sup>18</sup> I return to this issue in Section 5.

that non-desire-based explanations are somehow aberrant. A cognitivist might think that cognitive motivation is a more widespread phenomenon than this focus indicates. The cognitivist may believe that our moral scruples can be implicated as sources of motivated action even when others are also implicated. It is not just when one doesn't want to do something that one's sense of responsibility (or whatever) kicks in to gear. The reason for focusing on the case of the begrudgingly responsible agent is only that it is a convenient way to isolate one potential source of motivated behaviors among others.

Stated positively, the cognitivist hypothesis allows for the mechanisms that lie at the source of Martha's behavior to also be in play when there is an additional desire-based explanation of her action. For instance, many people are not so reluctant about doing the right thing; many people are often *happy to* do it, which, I take it, is just another way of saying that we *want* to do it. It is nevertheless possible that some of this happy group would forge ahead and do the right thing even if they were not so happy about it. There are a number of possibilities compatible with this. It could be that the moral belief is the state that actually plays the causal role in initiating the action while the desire does not.<sup>19</sup> Alternatively, it could be that the desire initiates the action while the belief stays in the background, ready to be activated if the desire stops playing that role. Or, it could be that there is genuine overdetermination, where both the belief and the desire are independent initiators, each sufficient (in this case) to produce the action. If sources of motivated action could be isolated empirically, these are all interesting empirical possibilities, compatible with the positive cognitivist hypothesis.

---

<sup>19</sup> The desire might still play a causal role with respect to other things—for instance, in producing the relevant affect state.

### 2.3 Are the Cognitivist Hypotheses False?

The foregoing discussion was meant to highlight certain cognitivist hypotheses, and explain why the cognitivist finds them plausible. I have not said anything, however, to establish their truth. Doing so would require us to undertake serious empirical work that I am in no position to do here. However, it is worth taking a moment to assess the position of Timothy Schroeder, Adina Roskies, and Shaun Nichols (2010, 93-98) (hereafter, ‘SR&N’, for short), who contend that the current state of neuroscience suggests that cognitivism is not an empirically adequate picture.<sup>20</sup>

To make their case, SR&N provide three key pieces of evidence: (1) Parkinson disease interferes with the way in which an agent’s “intrinsic desires ... causally influence motivation” (SR&N 2010, 93). Insofar as Parkinsonian subjects lose *all* ability to influence motivation, this “shows that intrinsic desires are necessary to the production of motivation in normal human beings” (SR&N 2010, 93). (2) The only known model for behaviors that have their immediate source in the higher cortical regions of the brain is Tourette syndrome—a pathological state. As SR&N (2010, 94), point out, “the direct causation of motivation by higher cognition via this pathway, quite independently of desire, is the sort of thing that results in a Tourettic tic, but [this] seems a very unpromising parallel to be drawn for a cognitivist picture of motivation.” (3) Sociopaths and other individuals with damage to certain regions of the brain appear not to be motivated by their moral judgments or other types of normative beliefs (e.g. beliefs about

---

<sup>20</sup> Technically, their explicit conclusion is that cognitivism is not an empirically adequate picture of moral motivation, but the arguments I will address seem (for the most part) to be aimed at the claim articulated in the text.

what types of strategies are “best” for playing certain games or about how one should go about navigating certain social situations) (SR&N 2010, 94-98).

While this evidence is worth giving closer attention than I am able to give it here, it is worth pointing out two fairly large worries about its adequacy. First, (1)-(3) all rely heavily on Schroeder’s (2004) work that takes desire to be realized in humans by the brain’s reward system.<sup>21</sup> In support of the claim that Parkinson disease impacts *desire’s* ability to motivate, SR&N (2010, 93) tell us that this disease works by killing the cells responsible for transmitting signals from the brain’s reward system to the action centers. In support of the claim that Tourette tic are produced by a mechanism that proceeds *independently of desire*, SR&N (2010, 94) tell us that it works independently of the brain’s reward signal. The particular brain anomalies that exist in agents who are apparently not motivated by their moral and normative judgments are all located in the structures that are inputs into the brain’s reward system SR&N (2010, 94-98).

In Section 1.2 above, I discussed some of my misgivings with this specific account of desire. In particular, I expressed my concern that this mechanism is implicated too broadly in human cognition to be identified as the neurological realizer of desire, specifically. The evidence from Parkinson disease seems especially problematic in this light. One of the bits of evidence on which Schroeder (2004, 52) relies to establish the importance of the specific structures he identifies as the brain’s reward system also comes from this disease—Parkinsonian patients are unable to engage in at least some forms of reinforcement learning.<sup>22</sup> In the experiment Schroeder describes, subjects were

---

<sup>21</sup> See Section 1.2 above.

<sup>22</sup> This experiment was performed by Barbara J. Knowlton, Jennifer A. Mangels, and Larry R. Squire (1996).

asked to predict the “weather” in a computer simulation (the “weather” consisted of a cartoon picture of either a sun or a raincloud, along with either a high-pitched or low-pitched sound), on the basis of some novel pieces of evidence. Symbols arranged in certain patterns were made, by some arbitrary rule, to be reliable, but fallible indicators of the “weather”. After each attempt to predict the “weather”, subjects received feedback in the form of information about how things “really” turned out (and hence about the correctness of their prediction). This feedback about the correctness of their predictions was the only reinforcement (i.e. the only “reward”) involved. Unlike other research subjects, those with Parkinson disease were unable to learn how to reliably predict the “weather” on the basis of the symbols. Given this result, it appears that Parkinson disease has a fairly significant impact on at least some stereotypically *cognitive* functions.<sup>23</sup> It is thus difficult to see what warrants SR&N’s conclusion that Parkinson disease works specifically by affecting an agent’s desires.

Second, SR&N’s use of (3) against the cognitivist is based on the mischaracterization of cognitivism that I urged against in Chapter 2. SR&N (2010, 98) describe the upshot of this evidence, in the following way:

[The brain anomalies in question] sever the link between the cognitive judgments and motivation, leaving intact the judgment and its content, but not causing the motivation that might normally result. Such an effect would establish that moral judgment was not intrinsically or necessarily motivational, but that instead the link between judgment and motivation was contingent and defeasible.<sup>24</sup>

---

<sup>23</sup> One thing to note here is that the type of learning that the Parkinsonian subjects were unable to accomplish was not particularly action-oriented—it was not like learning how to perform a particular task in order to get some external reward.

<sup>24</sup> They credit this point to Adina Roskies (2006).

SR&N, thus understand (3) as establishing the claim that moral judgments and motivation can come apart. I do not deny this. I think that this poses no problem for cognitivism, properly understood. My argument in Chapter 2 was that the Humean's favored motivational states—desires—also have at best a contingent link to motivation. If desires are capable of motivating without being necessarily motivating, then the fact that moral beliefs are not necessarily motivating cannot be used to undermine the claim that these beliefs are capable of motivating. We need to treat beliefs and desires consistently, I argued.

Recall what SR&N (2010, 93) have to say about Parkinson disease: “Parkinson disease is a disorder in which intrinsic desires slowly lose their capacity to influence motivation.” The clear implication of this is that even in the more extreme cases, patients with this disease may still have desires that entirely fail to influence action. This point is clearer in Schroeder's (2004, 173) discussion: “People with severe Parkinsonian symptoms [...] lack the power to move [and] appear to have no motivational capacities, tendencies, or dispositions of any sort. Yet it is not particularly plausible to see such people as lacking in desires.” In fact, even one of the very cases of brain damage SR&N (2010, 94-95) describe seems to undercut their case in this respect.<sup>25</sup> The experiment involved individuals with damage to the ventromedial (VM) part of the prefrontal cortex (i.e. “to structures that are crucial input into the reward system” (SR&N 2010, 95)). Unlike normal agents, these subjects failed to act on their knowledge about what strategy would help to maximize their returns in a card game. From this, SR&N (2010, 95) draw the conclusion that “reason alone does not suffice to guide action independently of

---

<sup>25</sup> The experimental results are from Bechara et al. (1997) and Bechara et al. (2000).

reward information.” The problem with this reasoning is that it is a fair assumption that these subjects *wanted* to maximize their returns on the game. If so, then this experiment equally showed that these subjects were not appropriately motivated by their *desires*. Thus the bulk of the evidence (once again) suggests that desires are not necessarily motivating. For that reason, the cognitivist can legitimately complain that she should not be saddled with the claim that (3) seems to undermine.

It is worth emphasizing, that in this discussion I have not tried to establish the truth of the cognitivist’s hypotheses. I have only been trying to show that it remains an open empirical question.

#### **2.4 The Dialectic to Come**

I have argued that, on a plausible account of desire, the debate between Humeans and cognitivists amounts to an empirical dispute—the Humean theory entails empirical predictions, while the cognitivist offers one of two empirical hypotheses that stand opposed to those predictions.

In the rest of this chapter, I will further develop this cognitivist proposal, by situating it within two more familiar debates between Humeans and cognitivists. In Section 3, I lay out, in rough form, two Humean arguments that tie an agent’s purposes in acting to what she wanted to achieve in acting that way. These arguments provide the background for a common cognitivist retort: there is a key ambiguity in the ordinary concept of desire that renders this argument unproblematic for the cognitivist. In Section 4, I discuss this cognitivist response and tie it to the position that falls out of Sections 1 and 2 of this chapter. In Section 5, I emphasize the *a posteriori* nature of this proposal by

contrasting it with another common variety of motivational cognitivism—the view that all action is pursued under the guise of the good.

### 3. A Humean Objection

In Chapter 3, I argued for the following claim:

***Motivation-Disposition Link***

An agent *S* in circumstances *C* is motivated to perform some action  $\varphi$  only if either *S* is disposed to (try to)  $\varphi$  when *C* (or relevantly similar circumstances) or *S* is disposed in *C* to intend to  $\varphi$  at some later time.

If the *Motivation-Disposition Link* is correct, motivational cognitivism will be true only if the behavioral dispositions that are constitutive of being motivated are sometimes grounded in agents' cognitive states. In other words, according to this view, a moral judgment that  $\Phi$ -ing is right (for instance) is capable of motivating an agent to act because such a judgment can support dispositions to  $\Phi$  and it is capable of supporting such dispositions without the aid of any further state from the passionate side of the mind.<sup>26</sup>

Cognitivism faces an immediate challenge from certain opposing conceptions of desire—in particular, those versions of dispositionalism according to which the possession of certain behavioral dispositions is sufficient, or nearly sufficient, for desire. If being disposed to behave in characteristic ways is (nearly) sufficient for desiring, then

---

<sup>26</sup> As our discussion in the last two chapters indicates, however, it will be a contingent matter whether such a judgment will ground the needed dispositions. For this reason, it will also be a contingent matter whether the judgment will motivate an agent to act appropriately. Like desires, moral judgments are at best contingently motivating states.

the *Motivation-Disposition Link*, along with a few auxiliary premises suggests the truth of the following claim, which seems to be in tension with the view I have been sketching:

***Purpose-Desire Link:***

Whenever an agent, *S*, performs some action,  $\phi$ , in order to achieve some purpose or end, *E* (which could be no more than to  $\phi$  itself), *S* wants *E*.

Whenever an agent acted in some purposeful way, this is explained as the manifestation of some underlying disposition—the disposition that is (according to the *Motivation-Disposition Link*) entailed by the fact that the agent was motivated to act in that way. However, as we mentioned earlier, the kind of explanation in question is not just a one-off description of how that person was disposed to act in this well-defined situation. Understanding someone as having a particular purpose or goal in mind that explains their behavior in one situation requires us to see that purpose or goal as constraining and determining her actions in other situations too, so that she would behave similarly in a variety of relevantly similar circumstances and behave differently in a variety of relevantly different situations, depending on how that goal is to be achieved. When we have a purposeful explanation of an agent's action in terms of some end *E*, there will likely be a sufficiently extensive set of dispositions to engage in *E*-productive behaviors. Thus if productive behavior is the only or the primary characteristic manifestation of desire, this would suggest that whenever an agent performs an action to achieve some end, then she desires that end. This is just what the *Purpose-Desire Link* claims.

I have already offered some reason to think that the account of desire on which this version of the challenge rests is inadequate; but this does not matter since the

*Purpose-Desire Link* can also be supported in a different way, by appealing to the way in which our common practices for identifying an agent's purposes for performing some particular action involve pointing out something she wanted to achieve in performing it. Given this way of identifying an agent's purposes, it seems to be a straightforward *corollary* of the fact that someone intentionally or purposefully  $\phi$ -ed that there is something that she wanted—if only to  $\phi$  itself.<sup>27</sup>

To help clarify this idea, we can compare normal cases of intentional action with non-purposeful behaviors. Suppose Kaylee spills a cup of hot coffee on the museum guard. One way to explain Kaylee's behavior would be to point out some end that was to be achieved or promoted by the action: her behavior may have been meant to get revenge for some past offence, to distract the guard so that her partner could swipe some precious artifact, or just to satisfy an urge she has been feeling to pour hot coffee on him. In such cases, the action is explained by reference to its purpose.

Purposeful actions like these can be contrasted with non-purposeful actions—say, if Kaylee spilled the coffee on the guard because she tripped, sneezed, had a muscle spasm, or was not paying attention to where she was going. To explain these behaviors, we do not point out some thing at which Kaylee was aiming. For, she was not obviously aiming at anything.

It is natural to talk about what an agent was aiming at in terms of what she wanted to achieve. This point has been noted by, among others, Elizabeth Anscombe (1957) who famously argued that intentional actions are those for which a certain version of the

---

<sup>27</sup> This final clause—the 'if only to  $\phi$  itself' is meant to reflect the fact that we sometimes engage in an activity for its own sake, or for no further purpose. For the sake of ease, I will often omit this clause in the following discussion. It should nonetheless be understood as implicit.

question “Why did you do that?” applies—the version of the question that is to be answered by providing the reason for which the act was done. As Anscombe (1957, 62) pointed out, there are other questions that in common language are effectively interchangeable with this one. When inquiring into an agent’s reasons or purposes, we could just as easily ask “What did you want to do that for?” Or, to phrase it slightly differently, we might ask “Why did you want to do that?” The fact that these questions are all easily substitutable for one another suggests that when one applies, so does the other. When Kaylee spilled the coffee because of a sneeze or inattention, she could reject the “Why did you want...” question by responding “I didn’t!” She could similarly reject Anscombe’s original question by insisting, “I didn’t want to!” Or, she could reject either question with: “I didn’t mean to!” In contrast, Kaylee could not sincerely reject either question in one of these ways when she spilled the coffee to distract the guard or to retaliate.

These points about our common explanatory practices suggest that, by uncovering the purpose of an agent’s  $\phi$ -ing or the goal that was to be served by her  $\phi$ -ing, one is *thereby* pointing out what she wanted to achieve by her action, and vice versa. If Kaylee spilled the coffee to distract the guard, then *that* is what she wanted—she wanted to distract the guards and that is why she did what she did. Thus, it stands to reason, all action aimed at some purpose or goal must be grounded in the agent’s desires.

## **4. Purpose-Giving Desires vs. Desires Proper**

### **4.1 Nagel’s Complaint**

A common line of response to this objection is to insist that “want”, “desire”, and related terms are ambiguous in a way that makes the preceding objections unobjectionable to the motivational cognitivist. In a key passage, Thomas Nagel (1970, 29-30) makes the point in the following way:<sup>28</sup>

[I]t may be admitted as trivial that [some considerations] cannot motivate me to act without a desire being present at the time of action. That I have the appropriate desire simply *follows* from the fact that these considerations motivate me; if the likelihood that an act will promote my future happiness motivates me to perform it now, then it is appropriate to ascribe to me a desire for my own happiness. But nothing follows about the desire as a condition contributing to the motivational efficacy of those considerations. It is a necessary condition of their efficacy to be sure, but only a logically necessary condition. It is not necessary either as a contributing influence, or as a causal condition. [Nagel’s emphasis.]

In the most general terms, Nagel’s insight in this passage is that although there is a sense in which the *Purpose-Desire Link* is true, its truth in this sense does not entail the Humean theory. Nagel contends that in one sense of “desire”, to say that someone desires *x* is just to mark the promotion of *x* as the motive or purpose for some action she performed. It follows trivially from the fact that an agent willingly acted in some way for some reason that we can properly attribute to her a desire in this sense. It follows automatically from the fact (if it is a fact) that Kaylee purposefully spilled her drink on the guard that she wanted something or other. In general, whenever an agent’s behavior is guided in the appropriate causal way by an instrumental belief about what such

---

<sup>28</sup> For extensive discussions of Nagel’s argument, see Schueler (1995, Ch. 1) and Jonathan Dancy (1993, Ch. 1-2).

behavior would promote, it will be appropriate to attribute to her a desire (in this sense) for that object. To provide a label, I will call these “purpose-giving desires”.<sup>29</sup>

Nagel’s thought is that, insofar as they are attributed to the agent merely to mark her behavior as purposeful, purpose-giving desires need not play a causal role in the production of action. Attributing a purpose-giving desire to an agent leaves it open whether the state at the source of the action is a non-cognitive state, as the Humean would have it (following G.F. Schueler (1995), we will call these Humean desires “desires proper”), or something else—it leaves it an open possibility, for instance, that the agent’s action had a moral belief as its source. In fact, Nagel’s discussion in the quoted passage leaves it unanswered whether there must be any state whatsoever that plays such a role.<sup>30</sup> It is for these reasons, on Nagel’s view, that the fact that we identify an agent’s purposes with what she wanted to achieve in performing an action does not support the Humean theory.

#### **4.2 Why Nagel’s Complaint (As Is) Won’t Quite Work**

While Nagel’s main concern is correct—that there is a sense of “desire” on which the *Purpose-Desire Link* is true but unproblematic for the cognitivist—our picture requires us to understand this point in a slightly different way than he does. If Nagel is right about the purpose-giving use of “desire” and related terms, then we may have been premature

---

<sup>29</sup> One could use Donald Davidson’s (1963) term “pro-attitude” as a label for this use, as many other philosophers do. (See, for instance, Schueler (1995, Ch 1).) My reason for not doing so is that the label is not uniformly used in this way; it often used with a stipulated definition as another label for conative states (see, for example, Michael Smith (1994, 117)). To avoid needless confusion, I prefer to use a new term.

<sup>30</sup> See, for instance, Dancy (1993, Ch 1-3), who defends such a “pure” cognitivist theory.

in our rejection of behavioral dispositionalism. For, there seems to be something in the neighborhood of the following claim that provides a sufficient condition for the purpose-giving sense of ‘desire’ to apply:

***The Behavioral Condition***

For all agents  $S$ , and all propositions  $p$ , “ $S$  desires that  $p$ ” is true in the purpose-giving sense if (i)  $S$  performed some action  $\phi$ , (ii) this action was the manifestation of an underlying disposition to  $\phi$  when presented with a suitable information-bearing state to the effect that  $\phi$ -ing would promote  $p$ , and (iii) there is a *sufficiently extensive* class of behaviors,  $\psi$ , such that  $S$  is disposed to  $\psi$  when presented with a suitable information-bearing state to the effect that  $\psi$ -ing would promote  $p$ .<sup>31</sup>

The problem is that dispositionalists, by and large, think that desires construed as sets of dispositions still get to play a role in the production of action. The dispositionalist wants to say that when a disposition constitutive of an agent’s desire that  $p$  manifests in her  $\phi$ -ing, she  $\phi$ -ed *because* she desired that  $p$ , where this is “because” means that the desire contributed in some way to the action being produced. Yet Nagel thinks that it is

---

<sup>31</sup> *The Behavioral Condition* is only a sufficient condition for purpose-giving desires, not a necessary condition. For this reason, *The Behavioral Condition* remains silent about many of the cases that are taken to be objections to behavioral dispositionalism. For instance, this condition says nothing about cases in which agents do not attempt to satisfy their desires, and for this reason, the account allows for a range of such desires—including desires that have necessary or impossible contents. Consider the desire to conjure a round square. If Marvin wants to do this, realizes that there is nothing that can be done about it, and therefore does nothing to promote that end, *The Behavioral Condition* does not apply. Insofar as this is an account of purpose-giving desires, this is as it should be.

At the same time, this analysis can account for an agent’s attempts to influence things beyond her control. If we see Marvin staring intently at his hands as he repeatedly snaps his fingers, it could turn out that this is his attempt to conjure a round square—this is the goal at which his finger-snapping is aimed. This is compatible with *The Behavioral Condition* so long as, say, Marvin is convinced that he has supernatural powers and could conjure up a round square by snapping his fingers, this was the belief that got him snapping his fingers, and he is disposed to do a variety of other things if he thought that they were means to conjuring round squares. The fact that it is impossible to conjure a round square is irrelevant to the question of why he is doing what he is doing (although it is relevant to the question of why his finger snapping isn’t working). All we need is the belief that this might work. (The real problem for Stalnaker’s (1984) proposal was the need for worlds in which Marvin’s instrumental belief is true—i.e. worlds in which he could conjure a round square. The immediate proposal does not take this step.)

because purpose-giving desires contribute nothing to the action that this sense of ‘desire’ makes the *Purpose-Desire Link* unobjectionable for the cognitivist. There seems to be some tension between (a) Nagel’s claim that it follows from the fact that an agent purposefully  $\phi$ -ed that it is appropriate to attribute a desire to her, (b) the natural dispositionalist interpretation of this claim, and (c) Nagel’s insistence that purpose-giving desires play no role in the production of action. One of these pieces has got to give.

### 4.3 Refining Nagel’s Complaint I: The Model

In following Nagel’s thought that there is a key ambiguity in the ordinary concept of desire, the cognitivist does not need also follow Nagel’s thought that the ambiguity is between causally-efficacious desires and so-called “desires” that are not. Cognitivism can be reconciled with the claim that purpose-giving desires play a role in producing action so long as the roles they play are not always suitably independent of the agent’s cognitive states. To explain how this works, we will first develop a model by examining the relationship between desires proper and purpose-giving desires in common cases of motivated action. We will then see how this model can be extended to cognitive motivation, assuming that our cognitivist’s positive empirical hypothesis holds.

To begin, consider the normal case in which a Humean desire-based story seems readily available: Rory has been feeling the urge to hit a round of golf all afternoon and for this reason he heads off to his favorite course after work. According to our commonsense practice of desire attribution (and from *The Behavioral Condition*), it follows from the fact that Rory intentionally went to River’s Edge Country Club that it is appropriate to attribute a desire to golf to him. This is just a purpose-giving desire. In

this case, it seems like we also have something else—Rory’s *urge* to hit a round of golf. This “something else” is clearly part of the causal process that results in Rory’s going to River’s Edge. This is arguably a Humean passionate state—a desire proper.<sup>32</sup>

It is reasonable to wonder about the relationship between the purpose-giving desire that we attribute to him as a formal matter and the desire proper that we point to when are talking about his urge. It is clear that the two are not completely distinct. Without the urge, Rory would likely not have gone to River’s Edge or likely would not have done so for the purpose of hitting a round of golf. Other things being equal, if he didn’t have the urge to hit a round, it would not be appropriate to attribute to him a purpose-giving desire to do so. The urge that gets him to go also supplies the purpose or goal of the trip. If it was instead an urge for a drink or a Reuben sandwich that got him to go to River’s Edge, then his purpose for the trip would have been very different, and the purpose-giving desire we attribute to him would need to mirror this fact. The purpose-giving desire thus depends crucially on the desire proper.

According to dispositionalism, even though the purpose-giving desire depends crucially on the desire proper, it still figures in to the correct causal explanation of his behavior. The key point is that the purpose-giving desire does not contribute to the production of the action in a way that goes over and above the desire proper. The roles they play are not really distinct. If the hybrid account of desire I sketched above adequately captures the notion of a desire proper, then we can see why this is so. The

---

<sup>32</sup> Or at least I am assuming so for now. As I have been pointing out, there is a version of cognitivism that denies the existence of desires proper—for, on this view, all desires are ultimately grounded in the cognitive side of the mind. I briefly discuss a more modest version of this view in Chapter 5, according to which paradigmatic desires (including *urges*) are actually forms of perception, rather than desires proper.

desire proper is an HPC that is related to a range of affective, hedonic, phenomenological, and behavioral dispositions (and perhaps some others). The dispositions that are constitutive of the purpose-giving desire are a subset of the dispositions that are part of the desire proper—in particular, those dispositions that manifest in productive behavior. It follows straightforwardly, then, that whenever a desire proper for  $p$  is activated in the appropriate way to produce some productive behavior, there will be a purpose-giving desire for  $p$  that is activated as well. But this purpose-giving desire is not independent of the desire proper.

#### **4.4 Refining Nagel's Complaint II: Extending the Model**

On the assumption that the cognitivist's positive empirical hypothesis discussed in Section 2.2 holds, this model can be easily extended to cover the possibility of wholly cognitive motivation. The key thought is that sometimes the dispositions constitutive of an agent's motivation may not be suitably distinct from her cognitive states. That is to say, purpose-giving desires are sometimes constituted by dispositions that are themselves grounded in the cognitive natural kinds posited by the cognitivist's positive empirical hypothesis.

According to our model, these purpose-giving desires contribute causally to the action in the same way that Rory's purpose-giving desire contributed causally to his action. The key point is that even though purpose-giving desires grounded in the agent's cognitive states contribute causally to the action, they are not suitably independent of the agent's cognitive states in the way that the Humean theory requires—these purpose-giving desires would not be desires proper.

To make the idea slightly more concrete, the cognitivist, following Nagel, accepts that there is a perfectly good way of talking according to which we can say that Martha wants to go to the meeting. It is not like she headed toward the Administration Building by accident or was literally dragged there against her will. She left with the purpose of going to the committee meeting and she *ipso facto* wanted to go. The cognitivist also thinks it plausible that a normative judgment initiates the causal process that leads to her action and determines its purpose. The judgment plays both of these roles by grounding a set of behavioral dispositions—dispositions that are constitutive of the purpose-giving desire. The purpose-giving desire in this case is thus, by hypothesis, ultimately grounded in her motivating belief. But as our model clearly shows, this is not to say that the purpose-giving desire plays no role in the production of Martha’s behavior. The point is just that Martha’s purpose-giving desire does not play a causal role that is distinct from the role played by her motivating belief. The desire is not really independent of the belief any more than Rory’s purpose-giving desire was independent of his urge to hit a round of golf.

## **5. The Guise of the Good**

### **5.1 The *Purpose-Value Link***

Following Nagel and others, I have argued that there are likely two different concepts of desire. On my version of the view, one of these concepts is associated with a hybrid account and is tied to the natural kind that supports several of the characteristic manifestations of desire. On my view, the other concept—i.e. the concept associated with purpose-giving desires—is tied to the *Behavioral Condition*. I argued that if the

cognitivist's positive empirical hypothesis holds, then the cognitivist can exploit the difference between these two senses of desire to account for the truth of the *Purpose-Desire Link*.

In order to highlight the a posteriori nature of this account, it will be useful to contrast it with a more radical version of motivational cognitivism. The more radical version of cognitivism argues that all ends at which purposeful actions are directed are pursued “*sub specie boni*”, or under “the guise of the good”. As Immanuel Kant (1788, 5:59) describes this view that he calls the “old formula of the schools”: “*nihil appetimus, nisi sub ratione boni; nihil aversamus, nisi sub ratione mali* [We desire nothing except under the form of the good; nothing is avoided except under the form of the bad].”<sup>33</sup> According to this view, all desires are ultimately reducible to cognitive states that represent certain actions as good or worth doing.<sup>34</sup>

For our immediate purposes, let us understand this view as a commitment to the following:

***Purpose-Value Link:***

Whenever an agent, *S*, performs some action,  $\phi$ , in order to achieve some desired purpose or end, *E* (which could be no more than to  $\phi$  itself), *S* must see something valuable about *E*.<sup>35</sup>

---

<sup>33</sup> The translation is from Mary J. Gregor (1996).

<sup>34</sup> For a critical discussion of this view, see J. David Velleman (1992). For a recent defense of this view see Sergio Tenenbaum (2007). For further discussion—both sympathetic and critical—see the essays in Tenenbaum (2010).

<sup>35</sup> The “see” in this context is not meant to be factive—in the sense in which I am using it, one can see something that is not really there.

We might try to motivate the *Purpose-Value Link* in the following way. As we try to explain the behavior of others, we automatically modify our interpretations of what they are up to in light of new knowledge regarding the kinds of things they value. If an agent does not see anything worthwhile about a particular result of her behavior, that result is not considered the thing she wanted to achieve—we assume that the purpose of the act is to be found elsewhere. This practice indicates that the fact that an agent intentionally or purposefully  $\Phi$ -ed entails that there is something of value she saw in doing it—if only the intrinsic value of  $\Phi$ -ing itself.

It will be useful to return to Kaylee’s case. Let us say that Kaylee spilled her drink on the museum guard with some further purpose in mind—say, as a distraction, so that her partner could steal an artifact from the collection. If we are told that she sees no good in the thing that we have identified as her purpose, we would rightfully be puzzled. Suppose that she does not see anything good about having the artifact in question, the reasonable question then would be “Why try to steal it?” Kaylee may respond to this question by pointing to some further purpose—say, to sell it for loads of cash. However, if we are told that she sees no good in having the money, we should feel like we have not yet reached the bottom of things.

It appears as though any satisfying explanation of her behavior—and any satisfying answer to Anscombe’s question “Why did you do that?”—must land on *something* she sees value in.<sup>36</sup> Similarly, without an answer grounding out in something

---

<sup>36</sup> Note that we can now add something to our earlier contrast between the cases in which Kaylee acts purposefully, and those in which she spilled the drink because she tripped, sneezed, had a muscle spasm, or was not paying attention to where she was going. We can now say that in these latter cases where we do not explain her behavior by reference to the fact that it was meant to promote some desire, her action is not the pursuit of value either.

the agent sees value in, there will be no satisfying answer to the question, “Why did you want to do that?” There appears to be another version of Anscombe’s original “Why?” question that can now be made explicit: “What *good* did you see in (doing) that?” It seems as though all purposes or goals at which purposeful actions are aimed will be things that the agent sees value in.

## 5.2 Refining the *Purpose-Value Link*

The alternative version of cognitivism faces some well-known objections—at least on some interpretations of the opaque phrase “see something valuable about” that is at the heart of the *Purpose-Value Link*. Consider, for instance, the following gloss:

***Moral Judgmentalism:***

Whenever an agent, *S*, performs some action,  $\phi$ , in order to achieve some desired purpose or end, *E* (which could be no more than to  $\phi$  itself), *S* believes that *E* is morally valuable.

*Moral Judgmentalism* is subject to a range of famous counterexamples, several of which involve agents who are overcome by urges to do things they know not to be morally valuable. One such case is an addict who is overcome by her addiction—as she is pursuing her next fix, she need not be under any illusion that doing so is morally good.<sup>37</sup> Another case, offered by Gary Watson (1975, 210), involves a new mother who is feeling overwhelmed at her child’s constant crying and feels a momentary urge to drown the baby in the bathtub. If she gives in to her urge and drowns the child, she needn’t—even in the heat of the moment—believe that this is morally acceptable. A

---

<sup>37</sup> This example is introduced in a related context by Harry Frankfurt (1971). See also Smith (1994, 134).

further counterexample involves the Satan described in Milton's *Paradise Lost* who pursues certain ends and performs certain actions in part *because* they are morally bad.<sup>38</sup> Two of the cases discussed earlier can be seen as more mundane counterexamples to *Moral Judgmentalism*. Kaylee may have spilled hot coffee on the guard so that her partner could steal the artifact while knowing full-well that this is morally bad. Less dramatically, it seems plausible that Rory might simply lack any belief about the moral status of a round of golf after work—the thought may have never crossed his mind.<sup>39</sup>

Proponents of the *Purpose-Value Link* can amend *Moral Judgmentalism* in a number of different ways. Drawing on Kant's (1788, 5:59-60) insight that the notion of "good" involved in the "old formula of the schools" is ambiguous, one strategy is to weaken or alter the content of the belief, so that it is not moral value, but some other type of value in play. Robert Noggle (1997, 90) offers one such view—he suggests that desires are to be understood as beliefs with the content "that some state of affairs ... *prima facie ought to be brought about*" (Noggle's emphasis). About this property of being *prima facie ought to be brought about* Noggle (1997, 91) tells us that this is not to be understood morally, but in a more general practical sense:

Indeed, [the "ought property"] should be thought of as a marker of "bare normativity"—the thinnest possible notion of ought. Actually, the best rendition into English of the ought predicate would be something like: "I have *prima facie* reason to bring it about that **P**."

---

<sup>38</sup> For discussion of this case see, for instance, Anscombe (1957, 74-75) and J. David Velleman (1992, 18-19).

<sup>39</sup> The counterexamples to the various forms of judgment internalism discussed in Chapter 2 will also give us further counterexamples to *Moral Judgmentalism*, in that in each of these cases, the agent believes that a particular action is required, but nevertheless does something else.

Melissa Barry (2010, 213) similarly contends that purposeful action, if it is to be intelligible, must have its source in “the agent’s belief that there is something choiceworthy about her action (even if this consideration has been outweighed).” It is clear that Barry understands this belief as being about a *pro tanto* or *prima facie* normative property, rather than an all-things-considered property. In this respect her view is similar to Noggle’s. Yet there may also be some differences between the two views. In particular, it is not clear that “choiceworthy”, as Barry understands it, is supposed to be as thin as Noggle’s bare normative “ought”; it may be that Barry’s view is at least somewhat more moralized than is Noggle’s. Despite this possible difference, we can call a view that falls in the same range as Noggle’s and Barry’s *Pro Tanto Judgmentalism*.<sup>40</sup>

A second way to amend *Moral Judgmentalism* is to alter the psychological modality of the cognitive state involved. Instead of an all-out state like belief, some philosophers have suggested that the attitude in question is some sort of weaker representational state—something akin to an appearance or seeming.<sup>41</sup> On this view, although Kaylee and Rory may not *judge* their respective behaviors (or the ends at which these behaviors are aimed) to be good, these behaviors (or ends) might still *seem* good, at least in some respect or from some perspective.<sup>42</sup> *Perceptualist* views still face a choice as to the sense of “good” at issue. *Moral Perceptualism* is the position that when an

---

<sup>40</sup> Eve Garrard and David McNaughton (1998) are also sympathetic to this view.

<sup>41</sup> See Tenenbaum (2007, 38-42), Stampe (1987). For a discussion of Aristotle’s similar view, see Jessica Moss (2010).

<sup>42</sup> Tenenbaum (2007, 42-51) explicitly introduces the notion of an evaluative perspective from which the object in question seems good.

agent acts intentionally, the action (or end at which it is aimed) must seem morally good to her—at least in some respect or from some perspective. *Pro Tanto Perceptualism*, in contrast, deploys only the thinner sense of “good”.

*Moral Perceptualism* faces many of the same counterexamples as *Moral Judgmentalism*. Getting the next heroin fix or drowning the crying child needn’t even seem morally valuable, to the addict or the overwhelmed mother. The other case is even clearer—evil needn’t seem morally good in order for Satan to be attracted to it.

When it comes to *Pro Tanto Judgmentalism* and *Pro Tanto Perceptualism*, things are not quite so clear. Watson (1975, 210) denies of the distressed mother that she assigns “an initial value [to the action] which is then outweighed by other considerations.” Indeed, he insists, this action is “not even represented by a positive entry, however small, on the initial ‘desirability matrix’” (Watson 1975, 210).

One could say a number of things in response to Watson’s claim. One response on behalf of *Pro Tanto Judgmentalism* is to reject Watson’s description of the mother’s psychology as question-begging. It is not obvious that we should accept, without argument, that she sees absolutely no *pro tanto* value in drowning her child—after all, if she acts on her urge, she seems to be acting with the purpose of getting her child to stop crying, and in her desperate state this is something the overwhelmed parent would take to be a very good thing. In retrospect, of course, it will be obvious to her that the value of getting the child to stop crying in this way is far outweighed by other considerations, so that it is safe to treat this value as *effectively zero*.

*Pro Tanto Perceptualism* has a slightly different response available. According to the perceptualist, it might still *appear* to be good (or *pro tanto* good) to the woman in

question at the time—it is just that from a more reflective perspective, this appearance is immediately rejected as obviously misleading or illusory.<sup>43</sup> It is a common fact that things can appear to be a certain way from one perspective and another way from a different perspective. If she acts on her urge, then it is even more plausible that drowning her child really seems good from her immediate perspective—she unfortunately gives in to that illusory appearance; this can be so even if she would reject this appearance from a more reflective perspective. So Watson’s case does not provide a knock-down counterexample to either of the *Pro Tanto* views.<sup>44</sup> Similar responses can be made to the case of Satan and other purported counterexamples.

### **5.3 *The Purpose-Value Link* and Motivational Cognitivism**

Although proponents of the *Purpose-Value Link* can offer a refinement of the view that avoids the purported counterexamples, they have done this at the expense of offering a full explanation of why their favored cognitive states should be required for motivated action. As an autobiographical matter, to the extent that I can get a grip on the thought the favored cognitive states are supposed to be states of a piece with other types of judgment or perception, I find myself puzzled as to why they should be necessary in this way; and the more I feel the force of the idea that the drug addict, the overwhelmed mother, Satan, and all the rest are supposed to have the relevant states (so that these cases are not

---

<sup>43</sup> Tenenbaum (2007, 41) makes precisely this point in response to another of Watson’s counterexamples.

<sup>44</sup> There is another response available, which is to reject the mother’s behavior as a case of intentional action directed toward a genuine goal of hers. On this line of thought, insofar as the mother’s behavior seems to be the product of a compulsion that forces itself on her against her will, it is not really aimed at purposes or goals that are genuinely hers. See, for instance, Barry (2010, 213).

counterexamples to the view), the more I lose my grip of the thought that these are the independently identifiable judgments or perceptions housed on the cognitive side of the mind. I find it difficult to pin down both sides of the metaphor at once.

There is another way of making much the same complaint that should sound familiar. In its most basic form, the complaint is that while there is perfectly good way of talking according to which (for instance) *Pro Tanto Perceptualism* could be understood as true, its truth in this sense does not establish motivational cognitivism. There is a significant ambiguity when we talk about something *seeming good* to the agent, the objection continues, the sense in which it is true to say that *x* seemed good to the agent at the time is just the sense that marks *x* as the motive or reason or purpose for some action she performed. In this sense, it follows automatically from the fact that an agent willingly acted in some way for some reason that we can properly attribute such a take on the good to her.

To give this use a label, I will call these appearances of the good (or judgments regarding the good) “purpose-giving evaluations”. The Humean may object that insofar as they are attributed to the agent merely to mark her behavior as purposeful, purpose-giving evaluations need not be the kind of cognitive perceptions of value that the cognitivist has suggested is the source of purposeful action. Attributing a purpose-giving evaluation to an agent leaves it open whether the state at the source of the action is a cognitive state or something else—it leaves it an open possibility, for instance, that her action had a *non-cognitive* desire as its source. To say that something seemed good to her may not be to say that the agent had a cognitive perception-like state with the

appropriately thin sense of “good” as part of its content; it may suffice if the agent has a desire proper.

This is, of course, nothing more than Nagel’s argument, directed at the move from the *Purpose-Value Link* to motivational cognitivism, instead of the move from the *Purpose-Desire Link* to the Humean theory. It is difficult to see why it should be any worse here, than it was when the cognitivist was championing it. The problem is that if the cognitivist wants to suggest that desires are really just perceptions or judgments of value (or whatever), then it would behoove her to allow that they are of a kind with other types of perception or judgment—this is what would assure that they are cognitive, in the relevant sense. This is nothing more than a strong version of the cognitivist’s positive empirical hypothesis discussed in Section 2.2 above. If the cognitivist wishes to be free of this empirical baggage, she must provide an alternative account of what it is to see value in something. If it is not a perceptual state or a kind of judgment, then what is it and why is it a cognitive (as opposed to a conative) state? How is it that actions motivated by such states have their source in the cognitive side of the mind?

## 6. Review and Preview

I have tried to make it plausible that much of the debate between Humeans and cognitivists can be cast as an empirical dispute. If desire is a natural kind, then it makes sense to understand the Humean Theory as making a claim about some lawful role that kind plays in the production of motivated action. This is clearly an empirical question. Moreover, the cognitivist suspects that one of their leading cases—the begrudgingly

responsible agent—lends plausibility to the thought that the empirical question will fall their way.

I then tried to explain how this version of cognitivism is supposed to fit into some standard moves in the dialectic between cognitivists and Humeans. I argued that this version of cognitivism can help us make sense of Nagel's complaint that the ordinary concept of desire is ambiguous in a way that makes the Humean Theory seem more attractive than it really is. Moreover, the version of cognitivism I have developed here can be contrasted with another dominant version of cognitivism, according to which it is an *a priori* truth that all action is performed for the sake of pursuing some apparent good. I have suggested that if the guise of the good thesis is true, it is true as an empirical, rather than *a priori* matter.

One implication of the picture I have been developing is that some beliefs are capable of playing the same motivational role of desires proper. In fact, given the version of Nagel's point that I endorsed, there is a way in which it makes perfect sense to talk of them as desires—as purpose-giving desires. In the next, and final, chapter of this dissertation, I take up some arguments against these so-called “besires”. I contend that these arguments fail on their own terms, in that they fail to show that besires are a bad thing. I then offer a more general response that is related to the issues I have been discussing in this chapter: I contend that the argument proves too much, by ruling out on *a priori* grounds, a plausible *a posteriori* hypothesis about the nature of some motivational states.

## Chapter 5

### What's So Bizarre About Besires?

#### 1. Introduction

I have been trying to make it plausible that some cognitive states are capable of serving as sources of motivated behavior in their own right, without the aid of additional conative states from the passionate side of the mind. In particular, I have been defending the idea that moral beliefs or some other species of normative judgment are able to move agents to act as they direct.<sup>1</sup> We might say that moral judgments span the rational and motivational side of the mind, being at once both capable of playing the same informational role as other beliefs and playing the same motivational role as desires. This would be to say that these judgments are *besires*.<sup>2</sup>

---

<sup>1</sup> Instead of moral judgment, many motivational cognitivists choose to make their case on the basis of a different set of normative attitudes—in particular, on all-things-considered judgments about what one ought to do. (See, for instance, Ralph Wedgwood (2007, Ch. 1) and Sarah Buss (1997; 1999).) It thus bears emphasizing that the dispute between Humeans and motivational cognitivists is actually much broader than the debate regarding whether or not moral judgments can be sources of motivation; the Humean theory would still turn out false if any cognitive state could serve as independent sources of motivated action. For this reason, if one prefers to think in terms of some other species of normative assessment, what I say here is easily translatable. At the same time, it is worth recognizing that this shift to other normative judgments—including the all-things-considered normative judgment—is often the product of a bad assumption: that to make the case against the Humean theory, one must find some cognitive state that is necessarily motivating. This assumption I tried to undermine in Chapter 2, above.

<sup>2</sup> The term “besire” is owed to J.E.J. Altham (1986). The label is somewhat unfortunate in that it tends to mislead detractors into thinking that besires are a philosopher’s invention bearing no connection to ordinary folk psychological explanations. As Margaret Olivia Little (1997, 77 n. 7) has forcefully stated this concern:

In one sense, of course, the terminology is irrelevant as long as we are clear about what is asserted. In an important sense, though, this sort of rhetoric clearly imparts a substantive slant to the ensuing discussion. Everyday words belong to everyday conceptions, which are now claimed as Humean from the start. That

Humeans frequently suggest that this cognitivist proposal is ontologically extravagant. According to this objection, besires seem to be some unruly—possibly even incoherent—category of propositional attitudes, distinct from the beliefs and desires recognized by folk psychology and best cut adrift like other bits of wanton philosophical excess. At times, even motivational cognitivists have shared these concerns; some have thus tried to find ways of resisting the Humean theory without appealing to besires.

In this final chapter I argue that the widespread objection to besires is misplaced. I begin by laying out the main lines of objection to besires and explaining why those objections fail. There is nothing bizarre about besires. In fact, I contend that a few relatively modest assumptions commonly endorsed by Humeans actually lend plausibility to the thought that moral judgments and other normative states could be besires in the relevant sense. Unless we are going to deny the existence of such beliefs, we are stuck with besires whether we like it or not. I then provide a more general reason to reject the Humean argument. I contend that it is an open *empirical* hypothesis that certain paradigmatic motivational states are actually besires. For this reason, arguments that would rule out this possibility *a priori* are flawed.

---

the non-Humean needs to import words to describe her theory reinforces the sense that she is importing new, unfamiliar sorts of mental states.

We should resist the Humean's easy rhetoric. Besires are no more than cognitive states that can be sources of motivation. I have yet to see why the possibility that such beliefs exist is supposed to be a *radical* departure from folk psychological explanation of moral motivation. So I will go ahead and accept the term and use it without shame.

## 2. The Stalemate

It is commonly accepted by both Humeans and cognitivists that moral judgments influence our behavior in significant ways, and it is a widely-recognized starting point that typical moral agents are often moved by their thoughts about what morality requires of them. As a case in point, suppose that Earl has a long-standing belief that he (and everyone else) has a moral responsibility to promote environmental sustainability. Suppose further that Earl has recently become convinced that buying locally-produced food is a crucial step toward sustainability. Given that Earl believes that his best option for obtaining locally-grown food is at a farmer's market, he decides to go to the one downtown to pick up some groceries.

As described, the case should sound perfectly mundane. The debate between Humeans and cognitivists involves how best to interpret it. The cognitivist contends that we have already been provided all the materials we need to explain why Earl went downtown to the farmer's market. He did this because it was a part of (or a step toward) promoting environmental sustainability—something he thinks he is morally required to do. This seems to be a straightforwardly purposeful explanation of Earl's behavior, and one that did not invoke any state from the passionate side of the mind. As it stands, this explanation did not cite a single desire.

Humeans insist that explanations of motivated action are often elliptical for some larger explanations in which reference to desire is made explicit. For instance, we might say that an agent went to the Post Office because he thought it closed at 4PM. Implicit in this explanation is the fact that he *wants* to get to the Post Office before it closes. Similarly, if we explain his trip to the store by pointing out that he thought he was out of

eggs, we are clearly leaving out some important details—not least of which is the fact that he *wanted* more eggs and thought she could get some at the store. The explanation of Earl's behavior is elliptical in much the same way, according to the Humean. The complete explanation would make explicit that there is some further desire in which his actions are in service—for instance, the desire to satisfy his moral responsibility to live a more sustainable lifestyle.

Since Humeans can explain them away, cases like Earl's cannot provide any decisive counterexample to the Humean theory. At the same time, it is reasonable to ask why the Humean should be so confident that their position is immune to all such potential counterexamples. We can safely grant the Humean's claim that passionate states like desires are capable of moving agents to act.<sup>3</sup> The only question is whether there are any cognitive states that are also capable of playing this role. Insofar as the Humean wishes to rule out an entire class of alternative motivational states, we must demand an account. It cannot be an inexplicable, brute fact that desires can be sources of motivated action while cognitive states cannot. All parties should agree that the debate between Humeans and cognitivists cannot be settled by insisting that the truth of one view or the other is a brute fact.

We will therefore proceed under the assumption that desires have some identifiable feature that enables them to play this particular role. The dispute between the

---

<sup>3</sup> Technically, motivational cognitivists are free to either take or leave this assumption. What we might call *pure* motivational cognitivists deny this, insisting instead that only cognitive states can be sources of motivated action. Jonathan Dancy (1993, Ch. 1-3) seems to forward a version of pure motivational cognitivism. The view defended here, however, allows for a plurality of sources of motivation, including both cognitive and non-cognitive (or passionate) states.

Humean and the motivational cognitivist is thus about whether or not beliefs and other cognitive states could possess this feature—whatever it may be.

The most prominent *a priori* argument for the Humean theory proceeds by locating just such a feature in the fact that intentional action is purposeful, or aimed at a goal. Desires, according to Michael Smith (1987; 1994) are uniquely capable of accounting for this fact. In order to be a source of motivated behavior, an attitude must somehow provide an agent with a goal or end at which the action is to be directed. However, to provide an agent with a goal, an attitude must have a particular “direction of fit” with respect to the object that is to be the goal or end. In order to provide or constitute a goal, a mental state must make some demands on the world—it must in some metaphorical sense be such that the world is to comply, to match or *fit* it. But only desires (and not beliefs or other cognitive attitudes) make such demands. Only desires are states that the world is to fit. Only desires are thus capable of providing agents with goals; only they can be sources of motivated action.

### **3. The Fuss About Fit**

Stated at this level of abstraction, the Humean argument is really no more than an argument schema, in need of an interpretation of the key ideas of a “direction of fit” and of “being such that the world is to fit.” It is extremely controversial how these crucial notions should be spelled out and it is not my goal to resolve this particular dispute. Instead, I wish to make it plausible that this controversy doesn’t really matter—at least as far as the present argument is concerned. Regardless of how we spell out the relevant notion, the argument fails.

Even if we cannot resolve the dispute, it might help us fix on the relevant ideas if we briefly look at a famous example discussed by G.E.M. Anscombe (1957, 56) on which all of the fuss about fit is based. Anscombe's case involves a man on a shopping trip. For our purposes, we will just imagine that it is Earl at the farmer's market. With a list in hand, Earl travels from stand to stand, purchasing items from the list as he goes. Gray (or so we'll call her) is a detective following Earl around the market, recording his every move. In particular, Gray is paying very close attention to the items Earl purchases.

One thing about which everyone can agree is this: even if Earl's list and Gray's notepad contained *exactly* the same entries, they would still be importantly different and stand in importantly different relations with the world. As an extremely rough statement of the difference, Earl's purchases are based on (or in some sense supposed to be based on) his list, while Gray's list is based on (or in some way supposed to be based on) Earl's purchases. This difference between the lists, then, corresponds to the two different directions of fit, and is supposed to help us explain why only desires are capable of motivating.

According to Smith's argument, desires are supposed to be like Earl's list in the relevant respect—we will say that these are states *that the world is to fit*. Beliefs, on the other hand are more like Gray's record, and stand in relation to their contents in much the same way as Gray's notepad stands in relation to the facts it is supposed to represent. We will say that these are states *that are to fit the world*. Beliefs—even moral beliefs—are incapable of moving agents to act because they have the wrong direction of fit to make

demands on the world. Any state that is supposed to match the world cannot make demands that the world is to fit it.

Cognitivists should find this argument unconvincing. For, they think that it is easy to make sense of an action as directed toward a goal that has been provided directly by an agent's moral judgments.<sup>4</sup> To see why, remember what the cognitivist wants to say about why Earl is shopping at the local farmer's market: he is doing this because he thought that this is part of (or a step toward) living sustainably, which is something that he thinks he should do. This is straightforwardly an explanation of Earl's action in terms of the purpose or goal the action is supposed to promote. However, according to this cognitivist explanation, the goal is not provided by any desire. Instead, according to the cognitivist proposal, the goal was provided directly by his belief that he has a moral responsibility to live an environmentally sustainable lifestyle. We did not need to point to any additional desire to provide him that goal.<sup>5</sup>

To be clear, the cognitivist's point is not to insist that Earl couldn't have some suitably related desire lurking somewhere in the background. Nor is the cognitivist

---

<sup>4</sup> Not all cognitivists think that this is necessary. According to Russ Shafer-Landau (2003, 135), for instance, some cognitivists would like to deny that all action is goal-directed in this way. Moral motivation in particular might be non-teleological. Even if Shafer-Landau is right about this, my argument is meant to show that goal-directed actions pose no special problems for the cognitivist.

<sup>5</sup> As I discussed in Chapter 4, even though we did not cite any passion among the antecedents or grounds of Earl's action, there is still a perfectly good sense in which it is fine to say that he went to the farmer's market because he wanted to buy locally-produced food or because he wanted to be more environmentally friendly. The motivational cognitivist insists that this usage serves to mark the action as purposeful and to distinguish it from cases in which, for instance, his behavior was the result of a muscle spasm or the triggering of a chip in his head. In other words, we think, this usage of "want" and related terms entails only the existence of a purpose or goal, and not necessarily the existence of any *distinct* psychological state from the passionate side of the mind. This point is made originally by Thomas Nagel (1970, 29-30), and is forcefully defended by G.F. Schueler (1995).

denying the possibility that Earl's action could have had its source in such a desire. The point is simply that we didn't *need to* appeal to such a desire in order to render his action intelligible as the pursuit of a particular goal. We could equally well understand this goal as provided directly by his moral judgments. To put the point somewhat differently, it is up to the Humean to tell us what citing a desire adds to this explanation. Once again, it cannot be a brute fact that desires can play this role while beliefs and other cognitive states cannot.

We are working under the assumption that the direction of fit metaphor is supposed to provide us a picture of the kind of states that are capable of playing this type of role. If one is convinced that an agent's goals must have their source in states that the world is to fit, and if one is also convinced—as is the cognitivist—that those goals need not be determined by anything over and above certain cognitive states, it should be natural to explore the hypothesis that such cognitive states might be states that the world is to fit. Yet since these states are also supposed to be beliefs, they are states that are supposed to fit the world. If this is right, there would be states that have both directions of fit—states that are in some respects like other paradigmatic beliefs, while also having the same direction fit that enable desires to play their explanatory role in the production of motivated behavior. In short, the cognitivist should be open to the hypothesis that moral judgments are besires.

According to common philosophical lore, we may have just presented a *reductio* of the apparently intuitive idea that moral judgments can be independent sources of motivation. *Being such that the world is to fit* it is incompatible with a moral judgment's

belief-like nature.<sup>6</sup> *Besires* would have *contradictory* natures. On a more modest version of the lore, cognitivism has been saddled with a great cost. It demands that we posit a strange new type of propositional attitude, distinct from the beliefs and desires recognized by folk psychology.<sup>7</sup> But I find it very difficult to see what all of the fuss is about. The knee-jerk reaction to the thought that mental states could have both directions of fit seems to me to be quite puzzling.

To take an analogy, suppose that Tweedledee and Tweedledum are arguing about whether or not one could haul a refrigerator in a car. Tweedledee takes it to be possible, but Tweedledum does not. In support of his view, Tweedledum points out a feature that would be necessary to accomplish this—say, a place to put the payload—and contends that no cars have this feature. Tweedledee, in turn, suggests that some cars might. An El Camino, he points out, might have sufficient room. If he agrees that El Caminos are cars, it should be puzzling if Tweedledum were to respond by insisting that having enough space to haul a refrigerator is incompatible with the El Camino’s car-like nature—that since any vehicle with enough space to haul a refrigerator would be a truck, and since El Caminos are cars, El Caminos must lack the necessary space. It should also be puzzling if he proceeds by insisting that Tweedledee is excessively inflating our automotive ontology by positing strange car-truck hybrids. These are not reasonable ways to argue that cars are incapable of hauling refrigerators.

---

<sup>6</sup> Smith (1988a, 250), for instance, seems to be making this claim when he insists that “Beliefs are thus *in their nature* unsuited to embodying the having of goals” (Smith’s emphasis).

<sup>7</sup> Smith (1994, 210-211 n. 10), for example, explicitly characterizes the dispute in this way.

The debate between Humeans and motivational cognitivists is at bottom about whether or not any beliefs or other cognitive states are capable of doing a certain kind of work. This is no more a dispute about the existence of mental states unrecognized by folk psychology, than Tweedledee and Tweedledum's dispute was about the existence of some new class of vehicle distinct from the cars and trucks recognized by *Popular Mechanics*. If both parties agree that El Caminos exist, and if both parties agree that they are cars, all that's left to disagree about is whether or not a refrigerator could fit in one. The way to settle this dispute is to take a measuring tape and check things out. Similarly, Humeans and anti-Humeans agree that we make moral judgments and agree that these judgments are beliefs.<sup>8</sup> The only remaining question is whether these beliefs are attitudes that the world is to fit—whatever that amounts to.

The way to resolve this dispute is to figure out what it is to be such a state, and to determine whether beliefs are like *that*. In the next two sections, I proceed to do just that. As our earlier rough statement of the difference between Earl's and Gray's list hinted, attempts to characterize the distinction between the two directions of fit come in two general flavors: those marking the divide in broadly normative or evaluative terms and those marking the divide in causal or counterfactual terms. The former will be the topic of Section 4; the latter will be the topic of Section 5. In both cases, my main argument is simple. On the best available accounts of what the distinction between the two directions of fit comes to, there would be nothing bizarre if it turned out that some mental states have both. Indeed, on the accounts, moral judgments arguably *do* possess both.

---

<sup>8</sup> At least the Humeans I am interested in agree to this. This is, after all, why these Humeans are so eager to deny that moral judgments can be independent sources of motivated action.

Therefore, if the direction of fit metaphor is supposed to tell us something deep or important about the nature of motivating states, this same metaphor also lends plausibility to the cognitivist's suggestion that moral judgments are such states.

#### 4. Normative Characterizations

In her original discussion of the example, Anscombe (1957, 56) draws the distinction in terms of different mistakes or ways that things might go wrong:

[The difference] is precisely this: if the list and the things that the man actually buys do not agree, and if this alone constitutes a *mistake*, then the mistake is not in the list but in the man's performance (if his wife were to say: 'Look, it says butter and you have bought margarine', he would hardly reply: 'What a mistake! we must put that right' and alter the list to 'margarine'); whereas if the detective's record and what the man actually buys do not agree, then the mistake is in the record. [Anscombe's emphasis]

This characterization does not actually square very well with the Humean argument under consideration. Any attempt to use the direction of fit metaphor to defend the Humean theory must point to a feature of desires that explains why they alone are capable of being motivational. In order to accomplish this, the characterization of the relevant direction of fit must, at a minimum, actually apply to desires. Anscombe's characterization does not satisfy even this minimal condition.

If Anscombe's account provides us the feature that the desire's direction of fit consists in, then when an agent's desire is left unsatisfied, the mismatch between the desire and the world constitutes a mistake in her performance. This is clearly not always the case. Consider someone's desire that the Minnesota Twins win the World Series. There need not be any mistake in *the desirer's* performance when the Twins fail to win—

chances are the problem lies exclusively with *the team's* performance. When the object of desire is beyond the agent's control, there is no sense in which its failure to come about constitutes a mistake in her performance.

Nor does Anscombe's account straightforwardly apply to desires about things within our reach. Desires can compete with one another. For this reason, we must sometimes sacrifice one thing in order to achieve something else. Suppose Rachel would like to stay inside where it's warm but also has a hankering for some Santa Fe chowder. If she doesn't have any at hand (or the necessary ingredients), satisfying the latter desire would require Rachel to venture out into the cold. She can't have it both ways and may be torn between her two options. But choosing to satisfy the urge for Santa Fe chowder at the expense of the desire to stay in need not involve any mistake in performance on her part. In deciding this way, she may even be performing ideally. Desires simply don't make this kind of demand on us.<sup>9</sup>

While ultimately rejecting the picture, Mark Platts (1979) offers a suggestion in the spirit of Anscombe's, more easily applied to the case of desire. According to Anscombe, the important difference between Earl's and Gray's lists is about where we would locate a mistake when there is a mismatch between one of the lists and Earl's purchases. Locating the mistake in Earl's performance may have just been an artifact of her focus on intention; the rest of the idea might still stand. Platts (1979, 256-7) suggests that the distinction may be drawn in terms of whether there is a failing *in the attitude* or *in the world* in the case of such a mismatch and which of the two should be changed to

---

<sup>9</sup> This isn't to accuse Anscombe of making any mistake here. For, her discussion of the case was not meant to tell us anything about the nature of desire; Anscombe was concerned primarily with the nature of intention.

remedy these mistakes. This way of locating mistakes and remedies is then grounded in the particular aims of the attitude:

Beliefs aim at the true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit with the world, not *vice versa*. Desires aim at realization, and their realization is the world fitting with them; the fact that the indicative content of a desire is not realized in the world is not yet a failing in the desire; the world, crudely, should be changed to fit with our desires, not *vice versa*.

This characterization avoids the issues related to Anscombe's focus on the case of intention. While it is no mark against an agent's performance when she does not get what she wants for reasons beyond her control, there is a perfectly good sense in which there is something wrong with how things turned out—at least as far as her desire is concerned. As far as her desire is concerned, the world would be somewhat shoddier if the Minnesota Twins lose the World Series. Similar thoughts apply, I think, to the case in which the agent satisfies one desire at the expense of another. At least as far as her desire to stay inside is concerned, not all is right with the world when the agent ventures out into the cold for a bowl of Santa Fe chowder.

Of course this talk about things not being all right with the world or somehow shoddier as far as one's desire is concerned doesn't really add much to the analysis of desire's direction of fit; it is really no more than another metaphorical way to articulate the same general idea. More work is needed to spell out this alternative way of putting the metaphor in non-metaphorical terms. Fortunately for us, however, we can still put pressure on the objection to desires even in the absence of such an account.

Besires are bizarre only if the two different directions of fit cannot be embodied by the same mental state—only if these two directions of fit are somehow in tension with one another. On at least the metaphorical way of articulating the normative account, that would be to say that there are never two different ways that things can go wrong with respect to a single mental state. This is plainly wrong. Consider a clarification that Anscombe (1957, 56) makes in her initial discussion of the case:

In the case of a discrepancy between the shopping list and what the man buys, I have to introduce the qualification: If this and this alone constitutes a mistake. For the discrepancy might arise because some of the things were not to be had and if one might have known they were not to be had, we might speak of a mistake (an error or judgment) in constructing the list. If I go out in Oxford with a shopping list including ‘tackle for catching sharks’, no one will think of it as a mistake in my performance that I fail to come back with it.

This passage suggests that Earl’s list may itself bear more than one normative relation to the world. Anscombe focuses on one particular case: the case in which Earl puts something on the list that he knows to be unavailable. Maybe this makes for defective intentions and goals, but it doesn’t obviously make for defective desires. (Is there anything wrong with wanting to win the lottery when you know this will never happen?)

Yet this needn’t be the *only* way that goals or intentions (and maybe even desires) might be “mistaken”. We would not be straying too far from commonsense if we thought that things could go wrong in other ways too. Dr. Evil’s performance might be off if he chooses insufficient means toward his goal of world domination. However, commonsense dictates that there is something wrong *with his goal* too. Hume’s famous insistence to the contrary notwithstanding, this is not the type of goal that Dr. Evil should be having and it is a failing of judgment on his part to hold it.

Many philosophers, Humean and cognitivist alike, have often endorsed just this idea.<sup>10</sup> For example, many of us think that intentions, goals, and even desires that have morally bad objects are defective, and should be changed to better fit with the moral landscape. If the normative distinction is what matters for our two directions of fit, the Humean's argument is very far from where it was supposed to be. The argument was supposed to establish that desires would be incoherent because the two directions of fit would be in tension with one another. However, if there are normative constraints on what kinds of things that can properly be the objects of one's goals, desires, and the like—if goals (and the rest) with morally objectionable objects are defective and to be revised in light of this fact—then on the normative construal of the guiding metaphor, attitudes with both directions of fit are not only coherent, they are looking downright mundane!

We find a similar pattern in the distribution of possible ways things can go wrong in the case of moral judgments. When a moral belief fails to accurately capture the normative contours of the world, it is mistaken. If it turns out that Earl has no moral responsibility to live an environmentally sustainable lifestyle, his belief that he has such a responsibility is faulty in this respect. The belief, roughly, ought to be revised to better fit the moral facts. Yet the possibility of this type of error is also compatible with there being another type of "mistake" as well. To take a rather simple case, one might believe that it would be best if House Republicans quit trying to repeal the Patient Protection and Affordable Care Act. It is intuitive to think that *something* would be amiss (as far as this

---

<sup>10</sup> Indeed, Graham Oddie (1994, 451-452) identifies three interconnected "regulative ideals" that he considers mundane: *harmony* – when one's desires and one's values (seen as beliefs about what is valuable) are aligned; *truth* – when one's beliefs about value aligned with the truth of the matter; and *purity* – when one's desires are aligned with what is valuable.

belief is concerned) when the House Republicans continue to press the issue. This is straightforwardly analogous to the case of an unsatisfied desire beyond one's control.

To take a case more relevant to the issue of action, many philosophers believe that weakness of will or *akrasia* is also a failing of sorts.<sup>11</sup> Many of us cannot shake the feeling that *something* would be going wrong if Earl continues living in obviously unsustainable ways while he is convinced that he has a moral responsibility to do otherwise. We might capture this feeling in the same metaphorical way: when Earl fails to do what he thinks is right, at least as far as this belief is concerned, things are not as they should be—at least as far as the moral belief is concerned, not all is right with the world. Whatever this really amounts to, it isn't obviously so much different from the case of unsatisfied desire. The Humean bears the burden of giving a non-question-begging account of what this difference is supposed to be.

The lesson should be clear. On the normative construal of the guiding metaphor, there is nothing bizarre about the mental states having both directions of fit. Moreover, at least in abstract terms, the normative account of desire's direction of fit actually lends plausibility to the cognitivist claim that moral judgments are capable of motivating agents to act. On the normative construal of that metaphor, moral judgments are *besires*, whether we like it or not.

## 5. Causal and Counterfactual Characterizations

Non-normative characterizations of the guiding metaphor tend to draw the distinction between the two directions of fit in causal or counterfactual terms. To highlight that there

---

<sup>11</sup> See Davidson (1970). For a fuller discussion of this issue, see Sarah Stroud (2008).

is some genuine difference here, it is worth pointing out that when things go as expected, Earl's and Gray's lists bear different causal relations to the world even when they contain exactly the same entries. Gray's notepad says "Organic Tomatoes" because Earl's cart contains organic tomatoes, while Earl's cart contains organic tomatoes because his list says "Organic Tomatoes". One of the lists is an effect, while the other is a cause.

If the two directions of fit are characterized along these lines, it is very difficult to see why motivational cognitivism should cause such a stir. Motivational cognitivism cannot be problematic for insisting that there are mental states that have both causes and effects. That would be no complaint at all. So if this account of the distinction supports an objection, there must be something problematic about the specific combination of a besire's causes and effects.

There is a way of thinking about besires that would indeed be problematic in this way. If we think of besires as a class of propositional attitudes distinct from belief and desire, then it might make sense to ask how a besire that  $p$  would differ from a belief that  $p$  or a desire that  $p$ . A natural idea, on this way of thinking, would be that a besire would be an attitude with both directions of fit regarding its content that  $p$ . On the simpleminded causal characterization of our guiding metaphor, this state would be the cause of the fact that  $p$  while at the same time being itself caused by the fact that  $p$ .

This is clearly a bad result. Yet this bad result is a red herring, and the way of thinking about besires that leads up to it is misleading. I argued above that the debate is not about whether there is a class of attitudes distinct from the beliefs and desires recognized by folk psychology, but about whether a subset of the beliefs (or other cognitive states) recognized by folk psychology could be sources of motivation. The

question is whether moral judgments could play this role and have the desire-like direction of fit that enables them to do so. But when we think of besires as beliefs that can do a particular kind of work, it does not make any sense to ask how it differs from a belief. For it *is* a belief! Earl's judgment that he ought to live a sustainable lifestyle is a belief *that he ought to live a sustainable lifestyle*. End of story.

We are assuming that Earl's moral judgment is a state that is to fit the world. According to the causal story under discussion, the cognitivist should thus be granted that Earl's moral belief is causally explained by the moral facts—that he *believes* he should live an environmentally sustainable lifestyle *because he really should* live such a lifestyle.<sup>12</sup> That is what having that direction of fit consists in, according to this position. The important question is whether these beliefs also have the relevant effects. To answer this question we must first figure out what the relevant effects are supposed to be. The debate between Humeans and cognitivists was supposed to be about whether this belief can get Earl to behave in more environmentally sustainable ways. If it can, then it has the relevant direction of fit in this respect—it has the same feature that enables a desire to live a sustainable lifestyle to play this role, which is what the cognitivist has been claiming all along. On the misleading way of thinking about besires, however, the relevant effect is to make it the case *that he should live a more sustainable lifestyle*. That is not, and never was, what the cognitivist thinks these beliefs are capable of doing. Nor is this what the Humean is trying to resist. The objection that falls out of the misleading

---

<sup>12</sup> Admittedly, there is something a little bit puzzling about the idea that moral beliefs are caused by moral facts. But this is something that can be safely set aside for my purposes here. This would pose a problem for anyone (Humean or not) who wished to claim that moral judgments are beliefs while at the same time insisting that beliefs are states that are to fit the world *in this respect*. Similar puzzles also arise for other attitudes that are uncontroversially beliefs—say the belief that the square root of 3,336 is 56.

way of thinking about besires attacks a straw man and completely mischaracterizes the debate.

I have dwelled on this point because Smith's leading argument against motivational cognitivism makes just this mistake. Smith (1994, 115) provides an alternative non-normative characterization of the distinction in counterfactual terms:

Very roughly, and simplifying somewhat, it amounts, *inter alia*, to a difference in the counterfactual dependence of a belief that *p* and a desire that *p* on a perception with the content that *not p*: a belief that *p* tends to go out of existence in the presence of a perception with the content that *not p*, whereas a desire that *p* tends to endure, disposing the subject in that state to bring it about that *p*.

Using this characterization, Smith (1994, 118) provides a direct argument against the possibility of besires.

[T]hough it might sound like a coherent possibility that there be such a state, it isn't really, at least if we take the suggestion quite literally. For, as we have understood the concept of direction of fit, the direction of fit of a state with a content that *p* is determined *inter alia*, by its counterfactual dependence on a perception with the content that *not-p*. A state with both directions of fit would therefore have to be such that, both, in the presence of such a perception it tends to go out of existence, and, in the presence of such a perception, it tends to endure, leading the subject who has it to bring it about that *p*. Taken quite literally, then, the idea that there may be a state having both directions of fit is just plain incoherent.

Smith's argument is that the possibility of besires "taken quite literally" is incoherent. This follows, he contends, from the thought that the two competing directions of fit place contradictory demands on the state when there is an apparent mismatch between its content and the world. Smith's argument presumes that motivational cognitivism, "understood quite literally" would entail one of two conjunctions:

- (A) In the presence of a perception *that he is not morally required* to live a sustainable lifestyle, Earl's besire (i) would tend to go out of existence, while also (ii) tending to endure, motivating him to bring it about *that he should live a sustainable lifestyle*. Or,
- (B) In the presence of a perception *that he isn't* living a sustainable lifestyle, Earl's besire (i) would tend to go out of existence, while also (ii) tending to endure, disposing him *to make the relevant lifestyle changes*.

Smith's argument fails because this is not, and never was, what motivational cognitivism implies. As far as I know, no motivational cognitivist has ever tried to defend either (A) or (B).<sup>13</sup> The only reason to think that the motivational cognitivist *would* be committed to either of these conjunctions is if Smith is thinking about besires in the misleading way I mentioned earlier. If we take besires to be a distinct type of attitude, then Earl would either have a besire *that he should live a sustainable lifestyle* or a besire *to live a sustainable lifestyle*. Only if we think about besires in this way would it make sense to ask how the besire would differ from a belief (and a desire) that he should live such a lifestyle or a belief (and a desire) that he does live such a lifestyle. And only then would it make sense to think of Earl's state as having both directions of fit regarding one and the same thing (i.e. to one of these propositions or the other).

The motivational cognitivist can, however, accept the following conjunction:

- (C) (i) In the presence of a perception (say, in the presence of some compelling proof) *that he is not morally required to live a sustainable lifestyle*, Earl's belief that he should live such a lifestyle would tend to go out of existence (and would cease motivating him to do so), but (ii) when presented with evidence that things *are not as they should be* (i.e. that his lifestyle is not sustainable) Earl's belief that he should be living such a lifestyle would endure, disposing him to make the relevant changes.

---

<sup>13</sup> Huw Price (1989, 120) and Margaret Olivia Little (1997, 63-64) make much the same point.

Insofar as moral judgments are beliefs, and relatively open to revision in light of new evidence, the first conjunct of (C) is unproblematic. Indeed, relatively few Humeans would want to deny this. Nor is the first part of the second conjunct objectionable: our moral judgments do not tend to go away when we come face-to-face with our less-than-perfect reality. (Few of us actually think that this is the best of all possible worlds.) Thus Earl's moral belief is similar to a desire to live sustainably in at least this respect. So *all* that is left to argue about is whether these enduring states also dispose the subject to make things better. Smith's argument does nothing to show that *this* idea—even “taken quite literally”—is incoherent. While one might dispute this second conjunct of (C), it is clearly *not at all* in tension with the first conjunct. We can take the motivational cognitivist's key intuition “quite literally” without falling into incoherence.<sup>14</sup> Once again, as far as I can tell, on such non-normative accounts of the guiding metaphor, there is nothing bizarre about besires.

---

<sup>14</sup> Smith (1994, 118) recognizes the need to respond to this “more subtle” version of motivational cognitivism that implies only (C). (He does, however, still characterize the “more subtle” view as holding that besires are some third kind of mental state—not wholly beliefs, not wholly desires, but like both.) Against this view, Smith argues that, it entails an implausibly strong version of motivational internalism, which he argues is false:

***Judgment/Motivation Internalism:***

Necessarily, if an agent judges that an action,  $\phi$ , is right (good, best, obligatory, etc.) she will be motivated to  $\phi$ , while if she judges that  $\phi$ -ing is wrong (bad, inferior, impermissible, etc.) she will be motivated to not- $\phi$ .

Following Michael Stocker (1979, 144), among others, Smith rejects *Judgment/Motivation Internalism* because it is incompatible with the phenomena of weakness of the will or debilitating depression. In Chapter 1 I argued that motivational cognitivism is not committed to such strong varieties of internalism. This is, essentially, because similarly strong internalist theses about paradigmatic motivational states (like desire) are similarly false. To put the point in the context of the present argument, while having the desire-like direction of fit may be necessary for an attitude to motivate, it is not sufficient.

It may be objected at this point that even on this picture, there would still be something strange about the *way* in which these besires have both directions of fit. For, on both the non-normative characterizations we are presently discussing and the normative characterizations discussed Section 4, these attitudes have the belief-like direction of fit with respect to one state of affairs, and the desire-like direction of fit with respect to another. Earl's moral judgment is supposed to relate in the belief-like way to the moral state of affairs that he ought to live a sustainable lifestyle, and in the desire-like way to the descriptive state of affairs that he lives such a lifestyle. Thus, it may be objected, if they are coherent in the way that I have described, besires are even stranger than we have imagined—they are, unlike any states recognized by folk psychology, propositional attitudes with multiple contents.<sup>15</sup>

In a footnote, Russ Shafer-Landau (2003, note 11 p. 137) lodges precisely this complaint:

One might claim that evaluative beliefs can have two directions of fit in virtue of having two contents, one aimed at describing the value attaching to some course of action, the other aimed at characterizing the end state being sought. But this seems ad hoc, at least if one is willing in all other contexts to distinguish belief from desire in terms of direction of fit. Further, it seems a plausible governing assumption of intentional attitude ascription that we individuate mental states by virtue of their contents. A unitary state with two contents (never mind their competing directions of fit) violates this plausible assumption.

But this objection misses the mark. I have been insisting that besires are simply motivational beliefs. As such, they have only a single content. The content of Earl's belief is nothing other than the moral proposition that he should live a more

---

<sup>15</sup> This is precisely how Smith (1994, 118) characterizes the “more subtle” version of cognitivism that endorses (C).

environmentally sustainable lifestyle. My view is that it is in virtue of having this belief (with this propositional content) that he is related in the desire-like way to the state of affairs that he lives such a lifestyle. The descriptive proposition representing this state of affairs is not a distinct content of his attitude. We can say this so long as we deny what we might call *The Content Condition*:

***The Content Condition:***

In order for an attitude *X* to have the desire-like direction of fit with respect to the state of affairs *that p*, *X* must have the proposition *that p* as its propositional content.

I do not, have any *a priori* proof that this condition is false. But we are certainly not forced to accept it. Suppose that we introduce a technical term to talk about the thing that an attitude like a goal or a desire is directed towards—we will call it the attitude's *object*. On the assumption that the direction of fit metaphor captures the important relation that these attitudes bear to their objects, the following condition should be true:

***The Object Condition:***

In order for an attitude *X* to have the desire-like direction of fit with respect to the state of affairs *that p*, *X* must have the state of affairs *that p* as its object.

As it so happens, a desire's object is also its propositional content. A desire with the content *that p* is directed (in the relevant sense) toward the state of affairs *that p*. The question is whether or not this must be true of all motivational states. I submit that cognitive states might have objects (in the stipulated sense) that differ from the state of affairs represented by their propositional contents.<sup>16</sup> As a case in point, Earl's moral

---

<sup>16</sup> For a defense of a similar idea, see Mark van Roojen (1995, 40-41).

judgment has his living a sustainable lifestyle as its object, but not its content. The content of the belief is simply the moral proposition that he has a certain obligation, while the thing toward which the belief is directed is whatever would satisfy that perceived obligation.

We might try to make this point another way. In Chapter 4, I developed a version of Nagel's complaint that ordinary concept of desire has two different senses—what I called the purpose-giving sense and the desire-proper sense. I argued that if a cognitivist-friendly empirical hypothesis holds, then purpose-giving desires are sometimes cognitive states in disguise. In other words, in such cases, there is a perfectly good sense in which we can talk about a cognitive state as though it is a desire.<sup>17</sup> Given the distinction between the content and the object of the attitude, we can not say that purpose-giving desire attributions are really about an attitude's object, not its content. Whenever an agent acts on the basis of an attitude that has the state affairs *that p* as its object, we can say that the agent desired *that p* (in the purpose-giving sense) even if we do not think that the agent acted on a distinct attitude with the content *that p*.

## 6. Appetites, Needs, and Desires

In the last two sections, I attempted to show that on the best available accounts of what desire's direction of fit consists in, there would be nothing bizarre if it turned out that certain cognitive states have it too. Indeed, moral judgments arguably do have desire's direction of fit, at least as far as these extant accounts are concerned. It turns out, then, that the guiding direction of fit metaphor actually supports the contention that moral

---

<sup>17</sup> For further discussion, see Jonathan Dancy (1993, 7-14), Little (1997, 64), Price (1989, 120-121), and Schueler (1995).

judgments are desires. While this result is important, the argument for it lacks generality. This is because we have been relying heavily on stories that philosophers have told to articulate the main idea behind the guiding metaphor.

In this section, I wish to offer a more general reply to the Humean argument. The response is simply that Humean attempts to use the direction of fit metaphor prove too much—they would rule out *a priori*, a coherent *empirical* hypothesis about the underlying nature of a class of paradigmatic motivational states.

Among the paradigmatic motivational states are a relatively distinctive category of attitudes sometimes discussed as *appetites* or *appetitive desires*. These attitudes are often thought of as the cravings for things like food, drink, and sex involved in hunger, thirst, and other biological urges. While these paradigm cases are generally accepted, it is somewhat controversial what other states should be classified along with them. It is easy to think that this forms a distinctive class of desires solely in virtue of this distinctive type of object—that these desires are ones that are directed toward the satisfaction of the agent's needs (especially her biological needs).<sup>18</sup> This is not exactly right, for a given object could be wanted in a variety of ways. Someone might desire to eat saag because she is craving it, because it is the cheapest item on the menu, because she wants to try

---

<sup>18</sup> Schueler (1995, 10-11), for instance, distinguishes these paradigm cases from other paradigmatic cravings (like the craving for chocolate) and from desires that are more difficult to characterize (like a “craving to see a Cary Grant movie” or a “desire to see my sister and her family next summer”) in part on this basis. In fact, Schueler (1995, 197 n. 1) is even hesitant to categorize sexual desires with the desires associated with hunger and thirst, because it is not clear how sexual desires are connected to genuine needs. Wayne Davis (1986, 65), in contrast, sees appetitive desires as the broader category, while recognizing a proper subset of these desires (what he calls “physical drives”) that includes both those connected to needs as well as other kinds of typical cravings: “Some people have physical drives for alcohol, caffeine, nicotine, and morphine, and everyone has physical drives for urinating, defecating, sleeping, and sex [...] but no one to my knowledge has ever had a physical drive for music.”

something new, or because she wants to make someone else happy. Only the first of these ways of desiring to eat saag is properly seen as part of this distinctive type of attitude that we think of as an appetite or appetitive desire for saag.

Moreover, the canonical appetitive desires are strikingly similar to other attitudes that are directed toward objects that are clearly not among our needs, and may even be detrimental to them. The hypertensive's craving for salt and the diabetic's sweet tooth bear a remarkable resemblance to both to other cases in which someone craves the object of a clearly unhealthy addiction as well as to the extreme thirst one might feel when travelling through the desert. So, while the connection to the agent's needs is an important feature of appetitive desires, this is neither what makes them distinctive, nor is even essential to them.

Part of what is distinctive about these attitudes also seems to be a particular type of associated phenomenology.<sup>19</sup> In Chapter 3, I described this phenomenology as a *felt urgency*. As I mentioned there, it is difficult to describe this phenomenology in any precise way. However, a feeling of this sort is at least part of what makes a craving for saag (or a craving for food in general, or for a cigarette) importantly different from desires the subject might have in other circumstances. While I find it introspectively plausible that different appetites or urges can share some aspect of their phenomenology, this is also supported by the established fact that we are frequently confused about what it

---

<sup>19</sup> Schueler (1992, 10-11; 15) also notes the importance of the phenomenological character in distinguishing different types of desires from one another.

is that we are craving. It is widely noted, for instance, that we can be very bad at distinguishing hunger and thirst from one another.<sup>20</sup>

Almost nobody—and certainly no Humean—wants to dispute that appetitive desires can be sources of motivation. As I said, these attitudes are among the paradigmatic motivational states. In fact, these are the very states that are often used to illustrate the need for motivational states in the explanation of action:

It is, of course, quite possible to have two identically trained rats that, on particular occasions, behave quite differently. Reggie eagerly presses the bar when he sees the light; Ronnie, his littermate, remains indifferent when he observes the same stimulus. What is the difference? Reggie is hungry; Ronnie isn't. [...] In explaining Ronnie's failure to press the bar (despite his training) when the light comes on by saying that he isn't hungry, we are merely identifying one factor in the cause of movement—the motivational factor—that distinguishes him from Reggie. (Fred Dretske 1988, 115)

When Earl has a craving for saag, there is no puzzle about why he stops at the Indian restaurant on the way home. He stops to get some saag—to scratch this particular itch. This is the goal served by his stopping, in just the same way that receiving a food pellet is the goal served by hungry Ronnie's lever-pressing behavior. *Ex hypothesi*, such attitudes must be states that the world is to fit.

That appetitive desires are attitudes that the world must fit seems right to me, and I will not try to dispute this. What I do want to dispute is that this fact resolves all of the questions regarding their underlying nature. One hypothesis regarding this nature is that an appetitive desire constitutes a special type of experience or awareness of one's own

---

<sup>20</sup> Indeed, as psychologist A.H. Maslow (1943, 373) pointed out, “any of the physiological needs and the consummatory behavior involved with them serve as channels for all sorts of other needs as well. That is to say, the person who thinks he is hungry may actually be seeking more for comfort, or dependence, than for vitamins or proteins.”

needs. Earl's craving for saag, for instance, ultimately amounts to an appearance that he needs saag or an experience as though saag is needed.<sup>21</sup> He is, quite literally, feeling a need for some saag.

For the sake of showing that this account isn't flatly absurd or incoherent, it is worth making the following points. It is plausible that creatures like us could have evolved an innate mechanism that automatically tracks information about our needs. The hypothesis at hand, then, could be understood as claiming that the faculty of "appetitive desires" effectively presents information about our needs automatically, in much the same way that information about our environment is represented in sense perception, or that we are able to track information about goings-on in our bodies via proprioception and other forms of interoception.<sup>22</sup> This hypothesis provides a nice way to explain why these states have a distinctive phenomenology in the first place—their felt urgency is tied to their representational function. It also provides us a tidy explanation of the regular connection that the distinctive phenomenology bears to our needs.

Appetites are similar to perception in other striking ways as well. Like visual or olfactory experiences, our appetites impress themselves on us automatically. As Thomas Nagel (1970, 29) describes it, these attitudes "simply assail us." Moreover, they play a key role in belief formation. We tend to move very quickly from desiring something in this way to forming an explicit judgment that we need it. The easiest cases in point are

---

<sup>21</sup> Dennis Stampe (1986, 167-169) floats something like this as an account of desires generally. I am not sure about this as an account of desires generally, and I want to float this as a hypothesis about appetitive desires, supporting this account by facts that are relatively distinctive about these desires. See also Stampe (1987).

<sup>22</sup> Stampe (1987, 371-376) also draws an analogy between desires, generally, and proprioception.

hunger and thirst—we move quite freely from these states to the thought “I need some food” or “I need some water” just as we freely infer facts about the world from our visual experiences. The cases of hunger and thirst are easy, I think, only because the “inferences” these states license are the most common and stable ones, and the ones most likely to be endorsed on later reflection. While other judgments may be less common or stable, we clearly do (at least in our non-reflective moments), say and think that we need things like coffee or cigarettes or excitement. When we say or think these things, our only basis for our judgment is that we are feeling the need. Moreover, other things being equal, the more vivid or intense the phenomenology, the harder one finds it to resist making the “inference” that the object is really needed. When faced with intense cravings, an individual can judge that she needs the craved object even though all of the other evidence runs in the other direction—as in the case where the heroin addict has become convinced that she *does* need another hit, despite all the contrary evidence.

To avoid misunderstanding, it is important to recognize that the hypothesis is not that when Earl is craving saag, he *will* believe or judge that he needs some. For, he might know better—he might know that appearances can be misleading and may fail to accurately represent the world as it is. This is no different from when our eyes play tricks on us or when one experiences pain in limbs that one knows are not there.<sup>23</sup> But when

---

<sup>23</sup> See e.g. Stampe (1987, 355-359). David Sobel and David Copp (2001, 47) make a similar point about vision: “Consider the apparent puddles one often sees when driving on a hot dry pavement. An experienced driver [may continue] thinking the road is dry when confronted with such an appearance.” Mary Clayton Coleman (2008, 130-131) likewise notes that “Someone familiar with the [Mueller-Lyon] illusion [might have] no tendency at all to stop believing that [...] two lines are the same length, even though it [...] appears to her that they are different lengths.”

Earl recognizes that he doesn't really need the saag, he is refusing to buy in to the appearance—he recognizes it to be mistaken.

For these reasons, it isn't flatly implausible that the distinctive class of states that we have been calling appetitive desires could actually be a rudimentary type of experience or awareness. If this hypothesis turned out to be true, then these attitudes would resemble beliefs and other paradigmatic cognitive states in an important way—an appetite would seem, in Hume's (1888, 415) terminology, to have a "representative quality, which renders it a copy of [some] other existence."<sup>24</sup> An appetitive desire would be like Gray's record in an important way—it, would turn out to be a state that is to fit the world. This is despite the plausible thought that it can be (and often will be) an independent source of motivation and thus, *ex hypothesi*, a state that the world is to fit.<sup>25</sup> Unless this hypothesis would undermine the consensus that these states are potentially motivational, it would support the claim that appetitive desires are states with both directions of fit—these states would be besires.

While I find this account of appetitive desires independently attractive, I want to rest here only with the fact that this is ultimately a coherent *empirical* hypothesis—and an ontologically benign one at that.<sup>26</sup> There is nothing obviously incoherent with the idea

---

<sup>24</sup> *Pace* Hume who, in the passage from which this quote is taken uses thirst as an example of something that lacks this type of "representative quality."

<sup>25</sup> Notice that here, once again, the two directions of fit are related to two different states of affairs. The content of such a perception is *that I need this* and these states are supposed to match the actual needs of the subject; the *object* of these states (in the technical sense I defined in the previous section) is the state of affairs in which the perceived need is satisfied.

<sup>26</sup> If one doesn't find this particular hypothesis compelling, it is worth pointing out that cognitivist (and perceptual) accounts are taken to be live views about the nature of the emotions—another type of paradigmatic motivational state. For a discussion of such views, see Jesse J. Prinz (2004).

that appetitive desires are of a kind with other recognized forms of perception, and this is a conclusion that cognitive scientists could arrive at after examining all of the psychological, neurological, and evolutionary evidence. Nor would it be some bizarre inflation of our mental ontology if appetitive desires turned out to be besires in this way. This hypothesis does not force us to posit strange new mental states distinct from those recognized by folk psychology. Everyone is already committed to the existence of appetitive desires. I have only suggested that there might be an interesting story to be told about these states and about the roles they play in our mental lives. This story cannot be ruled out *a priori*, as the Humean direction of fit argument would have it seem.

### 7. Motivated Desires, Unmotivated Desires, and Besires

A final objection to besires comes not from the Humean, but from other motivational cognitivists. Drawing on the work of Nagel (1970), many have suggested that if the motivational cognitivist distinguishes two different types of desires she can account for the goal-oriented nature of motivated action without positing besires.<sup>27</sup> Besires are simply unnecessary.

Nagel draws a distinction between what he calls “motivated” and “unmotivated” desires. The latter, like hunger or a sudden urge to go for a walk, are those that “simply assail us” while the former are those that do not assail us in this way but are “*arrived at* by decision and after deliberation” (Nagel 1970, 29; original italics). When it comes to motivated desires, Nagel (1970, 29) contends, “[r]ational or motivational explanation is just as much in order [as it is] for the action itself.”

---

<sup>27</sup> See, e.g. Shafer-Landau (2003, 137-138). This strategy is also discussed by, among others, Dancy (1993, 9-12) and Smith (1994).

Many have interpreted Nagel as, in effect, partially marking a distinction between *proximal* and *distal* (or *ultimate*) sources of motivation.<sup>28</sup> Some sources of motivated action, on this account, are ultimately explicable in terms of some antecedent set of mental goings-on that are themselves the ultimate (or *distal*) sources of motivated action. As a case in point, suppose that while at the farmer's market, Earl heads in the direction of the tomato stand. This action might be explicable in terms of a desire to pick up some fresh tomatoes. The desire to pick up some fresh tomatoes might in turn be explained in terms of some *further* motivational state—perhaps Earl wants to make paella for dinner. We might say that the desire to pick up some fresh tomatoes is motivated by the desire to make paella for dinner. It is thus also natural to think that he is heading over to the tomato stand because he wants to make paella for dinner—this is the desire that got him to act in this way, and it did so by producing a new desire in him.

This chain of explanation could continue. For instance, the desire to make paella for dinner might be the product of some further desire—for instance, the desire to please one's partner or to eat more seafood. On the other hand, the desire to make paella for dinner may not be explicable in this way. It may be a brute urge or craving that simply strikes Earl; it may be an unmotivated desire.

Nagel (1970, 29-30) suggests that a key question here is whether motivated desires always have an unmotivated desire among the causal or explanatory antecedents.

---

<sup>28</sup> Nagel is not as clear as he should be about this. One point of confusion is what relationship there is between the purpose-giving desires discussed in Chapter 4, above, and the present distinction between motivated and unmotivated desires. Nagel (1970, 29-30), as you might recall, is pretty clear about the fact that he thinks that purpose-giving desires often play no causal role in the production of motivated action. For this reason, it is odd to think of motivated desires as both proximal causes and the trivially ascribed desires (i.e. purpose-giving desires) as Nagel thinks about them.

According to some cognitivists, this explanatory chain may sometimes end in cognitive states (like moral judgments), rather than in unmotivated desires. According to the case we have been discussing, Earl's decision to go to shop at the farmer's market has its source ultimately in the belief that he ought to live a more environmentally sustainable lifestyle. For all we have said, it could have done this by getting him to want to go to the farmer's market. So long as a judgment like this can serve as an ultimate source of motivated action, and so long as an individual could be moved to act without the aid of any conative states among her ultimate sources of motivation, then the Humean conclusion can be resisted. For, actions produced in such cases would be the product of purely cognitive forces.<sup>29</sup>

This much seems right in the motivated-desire strategy: our moral judgments frequently *do* seem to produce further desires within us. We wouldn't be surprised if, after coming to judge that it would be really good (morally) for us to do something, we find ourselves genuinely wanting to do it. Nor is it implausible to think that, when this happens, this further desire could play some role in the production of action. Yet nothing in this account provides us a genuine alternative to the view that I have been defending here—nothing in this account, I think, allows us to avoid positing desires.

If we are to understand this view as offering a distinct alternative to the account I have been presenting, cognitive states must be capable of serving as sources of motivated action not in virtue of having both directions of fit, but in virtue of their ability to produce

---

<sup>29</sup> Given the discussion of the previous section, we need to find an alternative way to redraw Nagel's distinction. For, if the cognitivist story of appetites is correct, it would turn out that some important cognitive states also "simply assail us" in the relevant way. But there are independent reasons for the motivational cognitivist to redraw that distinction. It would be naïve to think that moral judgments never "simply assail us" but are instead always based on something like deep reflection or rational deliberation.

the necessary (i.e. motivated) desires. But the motivated-desire strategy gains its credibility as an alternative to the besire strategy only by ignoring the underlying concerns that led us to endorse the possibility of besires in the first place.

Even if we accept that proximal sources of motivated action *can* be desires, we have been provided no reason to think that they *must* be. Once again, this cannot be a brute fact; desires must have some feature enabling them to play this role. If this feature is their direction of fit, we are right back where we started—we must find a reason to think that no cognitive states could have that direction of fit. Without such an account, we have no reason to think that only desires can be proximal sources of motivation.

In much the same vein, it would be nice to have a unifying story about what it would take to be an ultimate source of motivated action. Assuming that unmotivated desires can be ultimate sources of motivation, our unifying story should tell us what feature enables them to play this role. If it is their direction of fit, then we have, once again, gotten no further. It is not obvious how this view can do without besires.

It is important to recognize that the standard motivation for endorsing the motivated-desire strategy as an alternative to positing besires is nothing more than the tired assumption that besires would be bizarre. I have been urging that this assumption is baseless; I have urged that there is nothing at all strange about besires, and so there is no reason to insist that they do not exist. In short, even if the motivated-desire strategy could (*per impossible*) be employed to show that we can do without besires, my arguments establish that there is nothing to be gained by doing so.

## 8. Concluding Remarks

In this dissertation, I have tried to make plausible the thought that there is an important sense in which moral agents are sometimes motivated in the absence of suitably related desires. More precisely, I argued that cognitive motivation—especially motivation that proceeds directly from an agent’s moral beliefs—is an a posteriori possibility. Insofar as the Humean Theory of Motivation is typically posited as either a conceptually necessary or metaphysically necessary truth, the arguments I have offered give us good reason to reject this theory.

It is worth highlighting a key thread that runs throughout several of the arguments of this dissertation. Cognitivists frequently react to the Humean’s insistence that the passionate mind plays a special role in the production of action by arguing that it is on the cognitive side of the mind where all the real action is. My version of cognitivism does not insist on this. At several turns, I have tried to take desire-based motivation as a model for cognitive motivation. At each of these turns, I have argued that the model fits. Thus my version of cognitivism is neutral on the question of whether there are any Humean desires-proper—even though I have been arguing that we can do without such desires, I have been supposing that we could do with them too.

The fact that desire-based motivation provides us such a neat model for cognitive motivation might give the radical cognitivist hope that the mind is not divided between reason and passion in the way the Humean Theory assumes. Indeed, by offering a hypothesis regarding the cognitive nature of certain paradigmatic desires, I have shown how the cognitivist might check for chinks in the wall that is said to divide the mind. But caution here is warranted. It bears repeating that the possibility that moral beliefs or paradigmatic desires are really desires—and indeed, the possibility of cognitive

motivation, more generally—is an *a posteriori* hypotheses for which I have provided only modest support. One should be wary when defending such a hypothesis from the comfort of the armchair.

This dissertation's defense of cognitivism is, by necessity, incomplete in another important respect. My main concern has been to vindicate the possibility of motivation that proceeds directly from one's moral judgments. I have not said much to address a further central question: What is so special about those beliefs? Why can a belief that it is good to help people motivate someone to help, while the belief that it is time-consuming to help people cannot? I have, at times, suggested that moral beliefs are probably not unique in their motivational ability—there are likely other desires out there. This too cries out for further explanation. What do all of these motivational cognitive states have in common? Why can they move us to act? Someone might suggest that some contents are *normative* or *practical* or seen as such to agents. That is probably correct, but it is not very helpful—this gives us a label for the explanandum, rather than providing the explanans.

I have provided one sort of answer to these questions: certain cognitive states can be motivational in virtue of the fact that they can (or perhaps should) support classes of behavioral dispositions that are sufficiently extensive to constitute a purpose-giving desire for some object or another. A benefit of this explanation is that it is neutral about which cognitive states, if any, will ultimately satisfy this condition. But insofar as it is neutral in this way, this structural explanation just kicks the can down the road. Why, if it is true, do some cognitive states give rise to such dispositions while others do not? Here, philosophical explanations may run out. I suspect that if there is any deep story to

tell it is more likely to come from cognitive science, evolutionary biology, and cognitive ethology than it is to come from purely philosophical reflection on the concept or nature of motivated action.

## References

- Altham, J.E.J. 1986. The Legacy of Emotivism. In Macdonald, Graham and Crispin Wright (eds.). 1986. *Fact, Science and Morality: Essays on A.J. Ayer's Language, Truth and Logic*. Oxford: Basil Blackwell. 275-288.
- Anscombe, G.E.M. 1957. *Intention*. Oxford: Blackwell.
- Aristotle. 2000. *Nicomachean Ethics*. Translated and Edited by Roger Crisp. Cambridge: Cambridge University Press.
- Audi, Robert. 1973. The Concept of Wanting. In Audi, Robert (ed.). 1993. *Action, Intention, and Reason*. Ithaca: Cornell University Press. 35-55.
- \_\_\_\_\_. 1989. *Practical Reasoning*. London: Routledge.
- Barry, Melissa. 2010. Humean Theories of Motivation. In Shafer-Landau, Russ (ed.). 2010. *Oxford Studies in Metaethics: Volume 5*. Oxford: Oxford University Press. 195-223.
- Bechara, Antoine, Hanna Damasio & Antonio R. Damasio. 2000. Emotion, Decision-Making and the Orbitofrontal Cortex. *Cerebral Cortex* 10: 295-307.
- \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_, & Daniel Tranel. 1997. Deciding Advantageously Before Knowing the Advantageous Strategy. *Science* 275(5304): 1293-1295.
- Bird, Alexander. 1998. Dispositions and Antidotes. *The Philosophical Quarterly* 48(191): 227-234.
- Boyd, Richard. 1988. How to Be a Moral Realist. In Sayre-McCord, Geoffrey (ed.). 1988. *Essays on Moral Realism*. Ithaca: Cornell University Press. 181-228.
- \_\_\_\_\_. 1991. Realism, Anti-Foundationalism and the Enthusiasm of Natural Kinds. *Philosophical Studies* 61(1/2): 127-148.
- \_\_\_\_\_. 1999. Kinds, Complexity, and Multiple Realization: Comments on Millikan's 'Historical Kinds and the Special Sciences'. *Philosophical Studies* 95: 67-98.
- Brink, David. 1986. Externalist Moral Realism. *Southern Journal of Philosophy* 24: 23-42.
- Buss, Sarah. 1997. Weakness of Will. *Pacific Philosophical Quarterly* 78: 13-44.
- \_\_\_\_\_. 1999. What Practical Reasoning Must Be if We Act for Our Own Reasons. *Australasian Journal of Philosophy* 77(4): 399-421.

- Carver, Charles & Michael Scheier. 1990. Origins and Functions of Positive and Negative Affect: A Control Process View. *Psychological Review* 97(1): 19-35.
- Coleman, Mary Clayton. 2008. Directions of Fit and the Humean Theory of Motivation. *Australasian Journal of Philosophy* 86(1): 127-138.
- Collins, John. 1988. Belief, Desire, and Revision. *Mind* 97(387): 333-342.
- Dancy, Jonathan. 1993. *Moral Reasons*. Oxford: Blackwell Publishing.
- \_\_\_\_\_. 1995. Why There is Really No Such Thing as the Theory of Motivation. *Proceedings of the Aristotelian Society* 95: 1-18.
- Davidson, Donald. 1963. Actions, Reasons, and Causes. In Davidson, Donald (ed). 1980. *Essays on Actions and Events*. Oxford: Oxford University Press. 3-19.
- \_\_\_\_\_. 1970. How is Weakness of the Will Possible? In Davidson, Donald (ed). 1980. *Essays on Actions and Events*. Oxford: Oxford University Press. 21-42.
- Davis, Wayne. 1986. The Two Senses of Desire. In Marks, Joel ed. 1986. *The Ways of Desire: New Essays in Philosophical Psychology on the Concept of Wanting*. Chicago: Precedent Publishing, Inc. 63-82.
- Dretske, Fred. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT Press.
- Fara, Michael. 2005. Dispositions and Habituals. *Noûs* 39(1): 43-82.
- Frankfurt, Harry G.. 1971. Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68(1): 5-20.
- Garrard, Eve & David McNaughton. 1998. Mapping Moral Motivation. *Ethical Theory and Moral Practice* 1: 45-59.
- Gibbons, John. 2009. You Gotta Do What You Gotta Do. *Noûs* 43(1): 157-177.
- Greenspoon, Joel. 1955. The Reinforcing Effect of Two Spoken Sounds of the Frequency of Two Responses. *The American Journal of Psychology* 68(3): 409-416.
- Gregor, Mary J. (trans. & ed.). 1996. *Immanuel Kant: Practical Philosophy*. New York: Cambridge University Press.
- Hare, R. M.. 1964. *The Language of Morals*. New York: Oxford University Press.

- Heuer, Ulrike. 2004. Reasons for Actions and Desires. *Philosophical Studies* 121: 43-63.
- Hume, David. 1888/1978. *A Treatise of Human Nature*. Oxford: Clarendon Press.
- Jacobson-Horowitz, Hilla. 2006. Motivational Cognitivism and the Argument from Direction of Fit. *Philosophical Studies* 127: 561-580.
- Kant, Immanuel. 1785. *Groundwork of the Metaphysics of Morals*. Translated by Mary J. Gregor. In Gregor, Mary J. (ed.). 1996. *Immanuel Kant: Practical Philosophy*. New York: Cambridge University Press. 41-108.
- \_\_\_\_\_. 1788. *Critique of Practical Reason*. Translated by Mary J. Gregor. In Gregor, Mary J. (ed.). 1996. *Immanuel Kant: Practical Philosophy*. New York: Cambridge University Press. 137-271.
- Kenny, Anthony. 1963. *Action, Emotion, and Will*. London: Routledge & Kegan Paul.
- Knowlton, Barbara J., Jennifer A. Mangels & Larry R. Squire. 1996. *Science* 273(5280): 1399-1402.
- Korsgaard, Christine M.. 1986. Skepticism about Practical Reason. *The Journal of Philosophy* 83(1): 5-25.
- Kripke, Saul A. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.
- Lewis, David. 1988. Desire as Belief. In Lewis, David (ed.). 2000. *Papers in Ethics and Social Philosophy*. Cambridge: Cambridge University Press. 42-54.
- \_\_\_\_\_. 1996. Desire as Belief II. In Lewis, David (ed.). 2000. *Papers in Ethics and Social Philosophy*. Cambridge: Cambridge University Press. 55-67.
- \_\_\_\_\_. 1997. Finkish Dispositions. *The Philosophical Quarterly* 47(187): 143-158.
- Little, Margaret Olivia. 1997. Virtue as Knowledge: Objections from the Philosophy of Mind. *Noûs* 31(1): 59-79.
- Manley, David and Ryan Wasserman. 2007. A Gradable Approach to Dispositions. *The Philosophical Quarterly* 57: 68-75.
- Martin, C.B. 1994. Dispositions and Conditionals. *The Philosophical Quarterly* 44(174): 1-8.
- Maslow, A.H. 1943. A Theory of Human Motivation. *Psychological Review* 50(4): 370-396.

- McDowell, John. 1978. Are Moral Requirements Hypothetical Imperatives? In McDowell, John (ed). 1998. *Mind, Value, and Reality*. Cambridge: Harvard University Press. 77-94.
- \_\_\_\_\_. 1979. Virtue and Reason. In McDowell, John (ed). 1998. *Mind, Value, and Reality*. Cambridge: Harvard University Press. 50-74.
- \_\_\_\_\_. 1980. The Role of *Eudaimonia* in Aristotle's Ethics. In McDowell, John (ed). 1998. *Mind, Value, and Reality*. Cambridge: Harvard University Press. 3-22.
- McNaughton, David. 1988. *Moral Vision: An Introduction to Ethics*. New York: Blackwell.
- Mele, Alfred R. 1995. Motivation: Essentially Motivation-Constituting Attitudes. *The Philosophical Review* 104 (3): 387-423.
- \_\_\_\_\_. 1996. Internalist Moral Cognitivism and Listlessness. *Ethics* 106: 727-753.
- Milgram, Elijah. 1995. Was Hume a Humean? *Hume Studies* 21(1): 75-93.
- Miller, Christian. 2008. Motivation in Agents. *Noûs* 42(2): 222-266.
- Moss, Jessica. 2010. Aristotle's Non-Trivial, Non-Insane View that Everyone Always Desires Things Under the Guise of The Good. In Tenenbaum, Sergio (ed.). 2010. *Desire, Practical Reason, and the Good*. New York: Oxford University Press. 65-81.
- Mumford, Stephen. 1998. *Dispositions*. Oxford: Oxford University Press.
- Nagel, Thomas. 1978. *The Possibility of Altruism*. Princeton: Princeton University Press.
- Noggle, Robert. 1997. The Nature of Motivation (And Why It Matters Less to Ethics than One Might Think). *Philosophical Studies* 87(1): 87-111.
- Nesse, Randolph. 2001. Motivation and Melancholy: A Darwinian Perspective. In French, Jeffrey A., Alan C. Kamil & Daniel W. Leger (eds.). 2001. *Evolutionary Psychology and Motivation: Volume 47 of the Nebraska Symposium on Motivation*. Lincoln, NE: University of Nebraska Press.
- Oddie, Graham. 1994. Harmony, Purity, Truth. *Mind* 103(412): 451-472.
- Pettit, Philip. 1987. Humeans, Anti-Humeans, and Motivation. *Mind* 96(384): 530-533.
- Platts, Mark. 1979. *Ways of Meaning*. London: Routledge & Kegan Paul.

- Price, Huw. 1989. Defending Desire-as-Belief. *Mind* 98(389): 119-121.
- Prinz, Jesse J.. 2004. *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.
- Prior, Elizabeth. 1985. *Dispositions*. Aberdeen: Aberdeen University Press.
- Quinn, Warren. 1993. Putting Rationality in Its Place. In Quinn, Warren. 1993. *Morality and Action*. Philippa Foot (ed). Cambridge: Cambridge University Press. 228-255.
- Raz, Joseph (ed.). 1978. *Practical Reasoning*. Oxford: Oxford University Press.
- Railton, Peter. 1986. Moral Realism. *The Philosophical Review* 95(2): 163-207.
- Rosati, Connie S. 2008. Moral Motivation. *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*. Edward N. Zalta (ed.). URL= <<http://plato.stanford.edu/archives/fall2008/entries/moral-motivation/>>.
- Roskies, Adina. 2006. Patients With Ventromedial Frontal Damage Have Moral Beliefs. *Philosophical Psychology* 19(5): 617-627.
- Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Schiffer, Stephen. 1976. A Paradox of Desire. *American Philosophical Quarterly* 13(3): 195-203.
- Schroder, Mark. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.
- Schroeder, Timothy. 2004. *Three Faces of Desire*. Oxford: Oxford University Press.
- \_\_\_\_\_, Adina Roskies & Shaun Nichols. 2010. Moral Motivation. In Doris, John (ed.). 2010. *The Moral Psychology Handbook*. New York: Oxford University Press.
- Schueler, G.F. 1991. Pro-Attitudes and Direction of Fit. *Mind* 100(2): 277-281.
- \_\_\_\_\_. 1995. *Desire: Its Role in Practical Reasoning and the Explanation of Action*. Cambridge: MIT Press.
- \_\_\_\_\_. 2009. The Humean Theory of Motivation Rejected. *Philosophy and Phenomenological Research* 78(1): 103-122.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defense*. Oxford: Clarendon Press.

- Sinhababu, Neil. 2009. The Humean Theory of Motivation Reformulated and Defended. *The Philosophical Review* 118(4): 456-500.
- Smith, Michael. 1987. The Humean Theory of Motivation. *Mind* 96 (381): 36-61.
- \_\_\_\_\_. 1988a. Reason and Desire. *Proceedings of the Aristotelian Society* 88: 243-258.
- \_\_\_\_\_. 1988b. On Humeans, Anti-Humeans, and Motivation: A Reply to Pettit. *Mind* 97(388): 589-595.
- \_\_\_\_\_. 1994. *The Moral Problem*. Oxford: Blackwell.
- \_\_\_\_\_. 1998. Galen Strawson and the Weather Watchers. *Philosophy and Phenomenological Research* 58: 449-454.
- Sobel, David and David Copp. 2001. Against Direction of Fit Accounts of Belief and Desire. *Analysis* 61: 44-53.
- Stalnaker, Robert. 1984. *Inquiry*. Cambridge: The MIT Press.
- Stampe, Dennis W. 1986. Defining Desire. In Marks, Joel ed. 1986. *The Ways of Desire: New Essays in Philosophical Psychology on the Concept of Wanting*. Chicago: Precedent Publishing, Inc. 149-173.
- \_\_\_\_\_. 1987. The Authority of Desire. *The Philosophical Review* 96 (3): 335-381.
- Stocker, Michael. 1979. Desiring the Bad: An Essay in Moral Psychology. *Journal of Philosophy* 76 (12): 738-753.
- Strawson, Galen. 1994. *Mental Reality*. Cambridge: The MIT Press.
- \_\_\_\_\_. 1998. Replies to Noam Chomsky, Pierre Jacob, Michael Smith, and Paul Snowdon. *Philosophy and Phenomenological Research* 58: 461-486.
- Stroud, Sarah. 2008. Weakness of Will. *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2008/entries/weakness-will/>>.
- Svavarsdóttir, Sigrún. 1999. Moral Cognitivism and Motivation. *The Philosophical Review* 108: 161-219.
- Tenenbaum, Sergio. 2007. *Appearances of the Good: An Essay on the Nature of Practical Reason*. New York: Cambridge University Press.
- \_\_\_\_\_. ed. 2010. *Desire, Practical Reason, and the Good*. New York: Oxford University Press.

- Thorndike, Edward L. and Robert T. Rock Jr. 1934. Learning Without Awareness of What Is Being Learned or Intent to Learn It. *Journal of Experimental Psychology* 17(1): 1-19.
- van Roojen, Mark. 1995. Humean Motivation and Humean Rationality. *Philosophical Studies* 79(1): 37-57.
- \_\_\_\_\_. 2002. Should Motivational Humeans be Humeans About Rationality? *Topoi* 21: 209-215.
- \_\_\_\_\_. 2011. Moral Cognitivism vs. Non-Cognitivism. *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*. Edward N. Zalta (ed.). URL = <<http://plato.stanford.edu/archives/spr2011/entries/moral-cognitivism/>>.
- Velleman, J. David. 1992. The Guise of the Good. *Noûs* 26(1): 3-26.
- Wallace, R. Jay. 1990. How to Argue about Practical Reason. *Mind* 99(395): 355-385.
- Watson, Gary. 1975. Free Agency. *The Journal of Philosophy* 72(8): 205-220.
- Wedgwood, Ralph. 2007. *The Nature of Normativity*. Oxford: Clarendon Press.
- Williams, Bernard. 1980. Internal and External Reasons. In Williams, Bernard (ed.). 1981. *Moral Luck*. Cambridge: Cambridge University Press. 101-113.
- Young, Sera L. 2011. *Craving Earth: Understanding Pica: The Urge to Eat Clay, Starch, Ice, and Chalk*. New York: Columbia University Press.
- Zangwill, Nick. 2003. Externalist Moral Motivation. *American Philosophical Quarterly* 40: 143-154.
- \_\_\_\_\_. 2008. The Indifference Argument. *Philosophical Studies* 138: 91-124.