

# UC Berkeley

## UC Berkeley Electronic Theses and Dissertations

### Title

Reason, Desire, and the Will: In Defense of a Tripartite Moral Psychology

### Permalink

<https://escholarship.org/uc/item/2hf590r3>

### Author

Carey, Jeremy

### Publication Date

2017

Peer reviewed|Thesis/dissertation

Reason, Desire, and the Will: In Defense of a Tripartite Moral  
Psychology

By

Jeremy Carey

A dissertation submitted in partial satisfaction of the  
requirements for the degree of

Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Hannah Ginsborg, Co-Chair

Professor R. Jay Wallace, Co-Chair

Professor Kinch Hoekstra

Summer 2017



Abstract

Reason, Desire, and the Will: In Defense of a Tripartite Moral  
Psychology

by

Jeremy Carey

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professor R. Jay Wallace, Co-Chair  
Professor Hannah Ginsborg, Co-Chair

Which aspects of our psychology are most central to explaining our intentional actions, and how should we conceive of them in light of their abilities to play those explanatory roles? These are key questions in moral psychology, and my dissertation tries to answer them, or at least to provide a beginning. As with much else in philosophy, the basic contours of this debate first came to us in Plato. Though I am not primarily concerned with the historical details, the initial argument of the dissertation and its distinctive approach reflect these Platonic origins in an interesting way.

In Plato's *Protagoras*, he presents Socrates as having an intellectualist moral psychology; that is, as claiming that all intentional action is motivated by a belief about what is best. This leads him to argue against the possibility of weakness of will. Plato himself later rejected this view, most notably in the *Republic*. There he argued that properly accounting for psychological conflict required dividing the soul into three "parts" - rational, spirited, and appetitive. I argue for what can be seen as a contemporary analogue of this later Platonic view; in particular, I argue that making sense of important aspects of our agency requires thinking of

ourselves as having three distinct motivational capacities, which I identify as reason, desire, and the will.

I begin by arguing that we can only act akratically, that is, against our judgment about what we ought to do, if there is a way to act *for* a reason besides our *faculty* of reason. This is because in such cases we have already formed a judgment about what we have most reason to do, and there is nothing besides this that can plausibly be attributed to an act of reasoning. And yet we do act akratically, and when we do so we act on the “weaker reason”. So there must be some other way to act for a reason than *through* reason. An obvious suggestion is that we act in such cases through desire. However, once we introduce this division between reason and desire, I argue that we will need a third, which I call the will - a capacity to consciously *decide* to act a certain way when reason and desire are insufficient or in conflict.

Though the appeal to desire in the initial argument in some ways seems obvious, to speak of desire, especially as a motivational capacity *alongside* reason, is to be embroiled in controversies that have dominated philosophical thinking about action since the early modern period. So I dedicate three chapters to defending the approach to desire appealed to by my tripartite theory. In the first, I argue against the “Humean” dispositionalist account of desire in favor of an evaluative conception. In the second, I argue that most recent evaluative accounts either fail to show how desire is really evaluative or else are overly intellectualized. In the third, I argue that we should think of desire as a form of responsiveness to apparent reasons, and show how this fits with both animal as well as human desire.

The last part of the tripartite view to be discussed is the will, which I introduce by way of a discussion of intention. I start by arguing that intention is a type of commitment to action that cannot be reduced to beliefs or desires, though in many cases beliefs and desires lead directly to intentions. Once we have this distinction between beliefs and desires and intentions, we can say more explicitly what the will is - it is a capacity to *decide*, where that involves the deliberate act of forming an intention. The basic claim of the tripartite view, that there are three distinct sources of motivation, can now be put in terms of intention, i.e., as saying that there

are three distinct routes to an intention - evaluative judgment (a conclusion of practical reasoning), desire, and decision.

In the final chapter, I briefly apply the view to the problem of identification, an issue in moral psychology which might initially seem to be particularly difficult to account for on the tripartite view. On the contrary, I argue that the tripartite view has certain advantages over other approaches, and sketching what I call the character view of identification allows me both to review the basic shape of the tripartite view and demonstrate once more its wide appeal.

# Table of Contents

Table of Contents .....	i
Preface .....	iii
<b>Chapter 1: <i>Akrasia</i> and the Diversity of Motivation, or Why Plato was (Basically) Right: A New Defense of Tripartition .....</b>	<b>1</b>
Introduction .....	1
Acting on the Weaker Reason .....	3
Two Responses.....	6
Towards Tripartition.....	9
Reason and Desire .....	9
The Will.....	13
Putting It All Together .....	15
<b>Chapter 2: Desire and Reasons for Action: In Defense of the Evaluative Conception .....</b>	<b>17</b>
Introduction .....	17
Desire and Reasons on the Humean Approach.....	18
Desire and Evaluation: Against Humeanism.....	23
Desire, Rationalization, and Quinn’s Radio Man.....	23
Objections and Replies.....	28
Copp and Sobel on Quinn and Scanlon .....	28
Response .....	31
Desires and Anti-Psychologism about Reasons .....	35
<b>Chapter 3: Two Constraints on an Evaluative Account of Desire, and Some Failed Attempts to Meet Them .....</b>	<b>40</b>
Introduction .....	40
The Two Constraints.....	41
Scanlon on Desire .....	45
Stampe and Tenenbaum on Desire.....	47
Moving Forward .....	53
<b>Chapter 4: Desire as Responsiveness to Reasons .....</b>	<b>54</b>
Introduction .....	54
Animal Action and Animal Desire: Preliminary Thoughts .....	55
The Structure of Reasons, the Structure of Desire.....	57
Reasons, Goodness, and Aristotle on the Nature of Animals.....	60

Responsiveness to Reasons and Responsiveness to Reasons <i>As Such</i> .....	63
Conclusion on Desire.....	69
<b>Chapter 5: Intention and the Will .....</b>	<b>70</b>
Introduction .....	70
Intention for the Future.....	72
Intentions in Action .....	74
Intention and Agency .....	76
Decision and the Will .....	82
What does the Will have to do with Willpower? .....	86
Putting it All Together.....	89
<b>Chapter 6: Agency, Identification, and Character.....</b>	<b>92</b>
Introduction .....	92
The Problem of Identification.....	93
Frankfurt’s Desire-Based View .....	97
Judgment-Based Views .....	100
A Character-Based Approach .....	103
The Need for a Distinctive Concept of Choice .....	109
Conclusion .....	112
<b>References.....</b>	<b>114</b>



## Preface

This dissertation is an extended essay in moral psychology. As I understand it, moral psychology is concerned in the first instance with a philosophical investigation of those aspects of human psychology which are relevant to moral evaluation, and in particular with those which are connected to an understanding of intentional action. Historically, such investigation has been preoccupied with questions about the psychological basis, or bases, of motivation. Through the early modern period, the central debate concerned the relative importance of reason and “the passions” in explaining action. The contemporary heirs of this dispute tend to talk rather in terms of a contrast between judgment and desire. My concern here is with this contemporary debate, though, as will be seen, I am often deeply influenced by the ancient Greeks.

What is at stake in this debate, and why is it worth thinking about at such length? There are several things to say. In the first place, of course, there is an intrinsic interest in gaining a deeper self-understanding and knowledge of the workings of our own mental lives. There are also, as might be guessed, implications for our understanding of morality itself. For one thing, moral evaluation is often thought only to attach to *intentional* actions, and one thing an adequate moral psychology should do is help us to understand what differentiates such actions from either unintentional actions or mere movements. In addition, no moral theory will seem adequate unless its dictates are possible for ordinary agents to put into practice. And this possibility will depend significantly on what human motivational psychology is like.

It is this last connection, put to use in a sort of skeptical argument, that has provided much of the urgency of the recent philosophical debates. A simplified version of the argument goes like this: A person is only morally obligated to do something if it is possible for her to be motivated to do it. However, what a person can be motivated to do depends on what desires she happens to have at the time of action or deliberation. Therefore, whether a person is morally obligated to do something depends on what desires she happens to have at the time. While full-blown moral skepticism obviously does not follow from this conclusion, it does seem to

involve a limiting of the universal rational authority of morality that many will find deeply troubling. And it depends centrally on a claim about moral psychology, about how human beings come to be motivated to act.

I do not directly address this argument in what follows, though I do briefly discuss it in the second chapter. I hope that the account of desire I give there and in the following two chapters, as well as the attractiveness of the overall picture, will undermine its appeal. Nevertheless, my direct concern in what follows is more descriptive than ethical. I try to give a big-picture account of the psychological bases of intentional action. I call this picture the tripartite view, insofar as I argue for three distinct and possibly conflicting sources of motivation: reason, desire, and the will.

The point of this little introduction has just been to put this big picture project into an even bigger picture, to say something about its wider appeal. Nearly every major moral philosopher has had something to say about human nature and action. While my philosophical aims are ultimately practical, it seemed best to me to start at the beginning.

# Chapter 1: *Akrasia* and the Diversity of Motivation, or Why Plato was (Basically) Right: A New Defense of Tripartition

## Introduction

A weak willed agent is one who freely and intentionally acts in a way contrary to his or her judgment about what it is best to do. While weakness of will seems an all too common feature of our lives, philosophers have had a notoriously hard time so much as accounting for its possibility. Most discussions of weakness of will occur within the context of some presupposed or previously given understanding of the psychology of action. In this chapter, I take the opposite tack, asking the question: What moral psychology is suggested by paying attention to weakness of will itself? This approach approximates that taken by Plato in the *Republic*, and I argue for what can be seen as a contemporary analogue to the Platonic view.

In his *Protagoras*, Plato attributes to Socrates an intellectualist moral psychology, according to which all intentional action is motivated by the agent's belief about what is best. This leads to the famous claim that no one errs willingly, and the equally famous denial that genuine *akrasia* ever occurs.<sup>1</sup> In spite of the counterintuitive implications of the account, the basic view can seem quite attractive. It is at least initially plausible that what distinguishes fully intentional actions from mere movements has to do with the agent's conception of what there is reason to do, and there is something very puzzling in the suggestion that one could see some course of action as unqualifiedly the best and yet fail to do it. At the very least, such a situation calls out for special explanation.

Yet Plato himself seems to have come to reject this view, most notably in the *Republic*. There he argues that the soul is divided up into three

---

<sup>1</sup> I will use "*akrasia*" and "weakness of will" interchangeably to refer to intentional action against one's better judgment. While this is a traditional way of speaking, it has recently come under fire. I discuss why briefly in chapter five.

distinct parts: the rational, the spirited, and the appetitive. In contemporary terms, we could say that he argues for three distinct sources of motivation, each to some degree independent and capable of conflicting with the others. He does this by pointing to certain examples of motivational conflict that we can be subject to and arguing that they are best explained by dividing the soul.

I will not be interested here in the details of Plato's arguments, nor in exactly his way of dividing the soul.<sup>2</sup> But I think he was right to distinguish between different sources of motivation, and that he was right in finding three. Most philosophers writing about action and practical reason today acknowledge only one predominant source of motivation - desire for the Neo-Humeans, and evaluative belief for the New Rationalists.<sup>3</sup> This strikes me as a regression from Plato's insight, even if I think that we should express that insight in different terms.

This dissertation as a whole is a defense of what can be considered a contemporary analogue of Plato's tripartite view of the soul. In this chapter, I want primarily to argue for the need for some basic division like Plato's, at least if we want to make sense of weakness of will. The key claim I want to argue for is this: given the reality of akrasia, we must have some capacity to act for reasons other than reason. Here I will only gesture at what else I think is involved - I think that Plato was right to find three motivational capacities, but whereas he divided the soul into reason, spirit, and appetite, I would divide basic human motivational psychology into reason, desire, and the will. Some time will be spent on each of these three parts in the chapters ahead. The goal of this first chapter is to defend

---

<sup>2</sup> Though as will be seen, the first part of my argument will actually be reminiscent of that in the *Protagoras*.

<sup>3</sup> Prominent examples of what I'm calling Neo-Humeanism, though diverse in other ways, include Simon Blackburn, Harry Frankfurt, and Michael Smith. The label "New Rationalists" is from (Johnston, 2001), who applies the label explicitly (and it seems to me rightly) to Parfit, Scanlon, and Raz. (Though Raz is a special case, since he also believes in the will in a sense very similar to the one I will defend. He does seem to think, however, that all motivation requires evaluative belief. Cf. (Raz, 1999) and (Raz, 2010).

and start to sketch the big picture. I would like, therefore, to present a new argument for a tripartite theory of motivation.

What follows will have three main parts. In the first, I will give an argument for why we cannot see akratic action as an expression of practical reason. In the second, I will show how the argument applies to alternative approaches from Michael Bratman and Pamela Hieronymi. In the third, I will briefly say how I think we should respond to the argument in developing an adequate psychology of action.

### **Acting on the Weaker Reason<sup>4</sup>**

On the standard way of understanding *akrasia* or weakness of will, the agent judges that he has most reason to do one thing, but does another instead. I display weakness of will, for example, when I judge that I ought to start grading a batch of papers this evening but instead watch a basketball game on television. It is important for genuine, “clear-eyed”, weakness of will that the agent both really believes, at the time of action, that he has most reason to do something else, and yet that he acts fully intentionally and free from compulsion.

If we make a standard connection between acting intentionally and acting for a reason, then it follows from the last requirement for genuine *akrasia* that the weak-willed person acts for a reason. This implication is so generally accepted that it is sometimes taken to be in part definitive of the phenomenon. Michael Bratman, for example, begins a paper on the topic by describing weakness of will as involving an agent acting “freely, deliberately, and *for a reason* in a way contrary to his best judgment” (Bratman, 1979, p. 153, my emphasis). And in a survey article on *akrasia*, Arthur Walker notes that

With the notable exception of Wiggins, all the [28] philosophers we have discussed accept (i) that whatever reason the akrates has for performing his akratic action is overridden by his better judgment and (ii) that the

---

<sup>4</sup> The argument of this first part of the paper was (to my knowledge) first given by Agnes Callard in her dissertation work (Callard, 2008). Though the basic line of argument was arrived at independently and I shall express the conclusion in rather different terms, my discussion here has been influenced and helped by hers.

akrates fully recognizes that his reason is thus overridden. Yet they maintain that the akrates acts on this reason (Walker, 1989, p. 670).

This way of spelling out what is involved brings out well what makes weakness of will so puzzling, and indeed why many philosophers have thought it in the end impossible. I don't want to deny that weakness of will happens, and neither do I want to deny that the weak-willed agent acts on a weaker reason. What I do want to argue initially, now that the basic ideas are on the table, is that weakness of will *would be* impossible, if we supposed that the akratic agent acts *on* the weaker reason *from* his capacity to reason. In other words, if we want to say that the agent acts for the weaker reason, we have to think that when he does so he is not *guided by* his *faculty* of reason. There must be another way, besides reason, to act for a reason.

The basic argument is very simple. When I reason about what to do, I am at least aiming to discover what it is best, all things considered to do.<sup>5</sup> I won't always arrive at an answer to this question. Sometimes good enough has to be, well, good enough. But in paradigm cases of weakness of will, reason *has* answered the question: there is a single thing, from my own perspective, that it is best to do. Therefore, insofar as I am acting *out of* that capacity to reason, I must act in the way that is judged best. If I do not, then my action cannot be motivated by reason.

This argument is in some interesting ways analogous to that given by Socrates in the *Protagoras* against the possibility of weakness of will. Socrates gets Protagoras to agree both (a) that weakness of will occurs when one does not pursue what one knows to be good because one is overcome by pleasure, and (b) that goodness and badness are to be defined in terms of pleasure and pain. But given the principle of the substitutability of definitions for what they define, this would allow us to

---

<sup>5</sup> Some may doubt this, thinking that I am conflating *theoretical* with *practical* reason at this point. The former, it may be said, tries to figure out what one *ought* to do, while the latter figures out *what to do*. I will return to this worry in the next section. For now, I will only say that the only sense in which it seems possible to *reason* about what to do is by figuring out what one ought to do. If there is a difference here, I do not think it can be made clear at this point without begging the question.

say that in weakness of will one does not pursue what one knows to be good because one is overcome with goodness (or, alternatively, that one does not pursue what one knows to be most pleasurable because one is overcome by pleasure). But Socrates takes this to be absurd - how could one be overcome by goodness and *thereby* choose something known to be less good? I want to say a similar thing about reason, though without depending on such a controversial substantive claim about value. If the akratic is guided by reason in his action, then he is motivated by reason to act against reason. And that is absurd. So he must be guided by something other than reason.

The problem here is not just normative, but constitutive. The claim is not just that being motivated in the way the akratic is would be irrational, but that it's just not possible to think of such a thing as being an outcome of reasoning. In practical reasoning, we *weigh* alternatives, and weighing is in at least some ways like counting. When I judge that I ought to start grading papers tonight, rather than watching the basketball game, I have fully taken the value of watching the game into account in making the judgment. It has already been counted, we might say. And unless I get new information, or fear I have made some mistake, it cannot be counted again, so long as I can be considered to be reasoning. As far as reason is concerned, this consideration can have no *additional* force beyond that which went into the original act of weighing.

Consider a scenario where you come across someone looking intently at some stones. You ask what he is doing, and he replies: "Counting how many stones there are." There are four stones in front of him, and as you watch, he points to each one, saying out loud "1...2...3...4..." Then he briefly pauses, looks a bit unhappy, and then points to the first one again and triumphantly says "5"! You point out to him that he has already counted that one, to which he replies, "Oh, I know I've already counted it, so that there were four stones, but I counted it again so that there are five". We could not conclude in such a case that the person is simply bad at counting. We would have to conclude that, whatever he was doing

(perhaps counting to five with the help of some stones), he was simply not counting the stones in front of him.<sup>6</sup>

The akratic appears to be in a similar position. Though he has fully considered all of his beliefs about what he has reason to do, and though he has already counted the reason for which he acts, finding it outweighed and insufficient to motivate action, he is motivated to act on that reason anyway. This invites the analogous conclusion that, whatever such an agent is doing, he is not doing something that is recognizable as an expression of practical reason.

## Two Responses

Resistance to this line of argument is natural in the absence of any alternative picture of how we act for reasons. For we do not want to deny the possibility of weakness of will, and it is an important feature of the phenomenon that the agent both acts on a reason and *sees* himself as doing so, even though he believes that the reason he is acting on is only *prima facie* or outweighed.<sup>7</sup> So it might be thought that there must be *some* form of reasoning leading to action.

This is the basic argument given by Bratman in one of his early treatments of the subject (Bratman, 1979). He proposes as a basic principle of practical reasoning that one can move rationally from a belief that one has *prima facie* reason to do *A* to the practical conclusion “I shall do *A*”, so long as that belief is not undermined or overridden by anything else one believes (or that is entailed by what one believes). The akratic agent follows this “potentially-justified inference pattern” in a way that is not actually justified, since the reason *is* taken to be overridden. Since making this inference conflicts with the principle, it is irrational. But the irrationality at least seems understandable - it doesn’t involve, for example, having contradictory beliefs or violating any constitutive norms.

---

<sup>6</sup> This example is from (Callard, 2008, p. 22).

<sup>7</sup> I use “*prima facie*” here to fit with the usage of Bratman discussed below. However, he almost certainly meant for it to be understood as “*pro tanto*” instead.



The problem is that this does not in any way address the actual argument I've given above. Bratman recognizes there is a problem here - "...how could this (faulty) transition be a piece of *reasoning*?" he asks (p. 168). His only response is to point to the fact, mentioned above, that a weak-willed agent can still give *some* justification or reason for his action, and so seems to be (and to see himself as) acting in accordance, albeit imperfectly, with the inference pattern Bratman gives. This fact does seem to be a requirement for thinking that the agent acted intentionally. However, it does not follow from either of these facts that the agent actually *makes* the practical inference above, and the argument was supposed to show that he *couldn't*, at least knowingly. The issue is not just understanding how someone could make a faulty inference - we do that all the time. The problem is understanding how someone could make a faulty inference *while simultaneously being aware that it is fallacious*. This was supposed to be the analogy with counting the stones. If a person double counts a stone without noticing, then he has made an error but he still counts as counting. But if he is fully aware that he has double counted or is about to double count and then does so anyway, *then* I think we lose our grip on how he is even engaged in an activity we could call "counting the stones".

Now, I've assumed throughout the above that when one is reasoning practically about what to do, the initial conclusion of such reasoning is a judgment about what *ought* to be done, or about what one has most reason to do.<sup>8</sup> To many, that may have seemed like a bad move right from the start. It may be claimed that I have confused *theoretical* reasoning *about* practical matters with *practical reasoning*. Pamela Hieronymi, for example, says that these are two distinct types of reasoning insofar as they aim at different types of answers to two different sorts of questions. What I discussed above she would call theoretical reasoning: an attempt to answer the question of what one *ought* to do, which concludes in a

---

<sup>8</sup> I say "initial conclusion" because I think that in normal cases such a judgment will lead directly to an intention, and so in some sense the intention may be seen as the ultimate conclusion. For more on this, see chapter five.

judgment. Practical reasoning, on the other hand, she calls an attempt to answer the question of what *to do*, which concludes in an intention.<sup>9</sup> She thinks this way of looking at things both allows us to see how action is governed by reason, and so is attributable to us in a distinctive way, as well as how weakness of will can be a possibility - in such cases, we have answered the *theoretical* question, about what we ought to do, one way, but have answered the *practical* question, about what to do, another way.

I don't think we can draw these distinctions in these ways. The main problem is that there is no sense that can be given to an attempt to answer the question of what to do, conceived of as an exercise of reason, that does not just collapse into what she would call theoretical reasoning. The only way we can *reason* about what to do is by considering the reasons there are to do the various things which are open to us. But reasons to do things are also reasons why we at least potentially ought to do them, and such reasoning is concluded when we make a judgment about what we have most reason to do. She insists that this cannot be the conclusion of practical reasoning because the practical question of what to do may be left open by such a conclusion, noticeably in just those sorts of cases I have been discussing. I want to agree that this gap *may* arise, but deny that it supports her way of distinguishing between practical and theoretical reason.

First of all, it is wrong to think that the gap between judgment and action is a general feature of practical judgments; rather, it is a special case. We don't always form explicit judgments about what we ought to do, but in very many of those cases in which we do, the judgment that one ought to do something will lead directly to one's doing it. In other words, and to use Hieronymi's terminology, answering the question of what one ought to do at least sometimes also constitutes an answer to the question of what to do. Indeed, it seems like it must be the normal case, for a failure to do what one judges one ought calls out for special explanation. This is why weakness of will is a puzzling phenomenon, and why in cases of

---

<sup>9</sup> See, e.g., (Hieronymi, 2009, pp. 206–7).

weakness of will we can almost always say that a conflicting desire at least in part explains the failure.

Second, and more importantly, when there is a gap between judgment and intention, it does not seem that what is required to close it can be described as reasoning. And thus it's unclear why we should consider whatever closes the gap as reason. The reasoning has been done. One knows, or at any rate has conclusively judged, what one has most reason to do. Now one must choose whether to do it or not. So there is a question that may be left open. But the answer comes by a choice, and that choice is not itself an exercise of reason. Hieronymi claims that it is, because she defines reasoning as the attempt to answer a question, when that is "the kind of activity done for reasons". But this, I think, begs the question. She is right to think that there is a question that needs settling, and right to think that in settling it one must in some sense be acting in accord with one's take on reasons. But neither of these implies that it is reason that settles the question, that one settles it by *reasoning*. The challenge for her account is to say what the agent must do in such circumstances and how it counts as reasoning. I am doubtful that this challenge can be met.

I've just said that what seems needed in a case where some gap arises between judgment and action is a *choice*, where that seems something like an irreducible act of will. But this is getting ahead of myself. I must now say something about how I think we should respond to the argument above in building a moral psychology. To do that, we must begin by understanding weakness of will as involving a *conflict*. I will then return to some of the considerations above in discussing the will.

## **Towards Tripartition**

### **Reason and Desire**

If there were no other alternative to acting on a reason than by way of practical judgment and *reasoning*, then we would have to deny the possibility of clear-eyed *akrasia*. Or so I argued above. Since it seems obvious that clear-eyed *akrasia* happens, and therefore must be possible, we need to find an alternative. Thankfully, there is an alternative, and it is at least frequently a part of the description of the phenomenon itself: in cases of weakness of will, the agent does not act according to reason, but

according to desire. The best way to allow for weakness of will, then (and the first step in defense of a tripartite view), is to accept that rational animals like us have two distinct basic sources of motivation: reason and desire.

There are two main sources of resistance to this move, one broadly Humean, and one broadly Rationalist. On the one hand, there are those who think that all motivation comes from desire, anyway, and that reason itself is never practical in its issue. This response could hardly be justified based on the argument above, since it does not help to solve the problem. In fact, it might only make it worse. On the simplest version of such approaches, what one has most reason to do depends on what one most desires, and desires are conceived of as intrinsically motivating states. Assuming that strength of desire correlates to strength of motivation, this seems to imply that one will always do what one most desires to do, at least if one knows that some way of satisfying the desire is possible and feasible. But in that case, one can never fail to do what one judges one has most reason to do, at least if that judgment is accurate. And that seems wrong.<sup>10</sup>

There is also good pre-theoretical reason to think that we can be motivated by our value judgments, even against our desires, at least in any ordinary sense of “desire”. We can all think of cases when we’ve done something out of duty or prudence even though we “really didn’t want to”. Of course, there is a large literature on this “Humean”, desire-centered, approach to practical reason, that I cannot get into here. I will just say for now that, by my lights, the arguments for such an approach fail, and there are good arguments against it.<sup>11</sup>

---

<sup>10</sup> For more sustained argument to this effect, see (Korsgaard, 1986) and (R. J. Wallace, 2006b).

<sup>11</sup> For modern loci classici in defense of important aspects of the Humean position, see (Williams, 1979) and (Smith, 1987). For a clear interpretation of influential criticisms of the Humean arguments, see (R. J. Wallace, 2006a). The argument against this position which I will develop comes originally from (Quinn, 1993).

On the other hand, those of a more rationalist bent will worry whether action motivated by desire, especially against one's normative judgments, could count as fully intentional. If a desire motivates in spite of being overruled in deliberation, then it seems to look more like an alien force than something which we can attribute to the agent. Consider David Pears sinister-sounding description: "...in a typical case of last-ditch akrasia an unruly desire takes over the control of intentional action and constitutes the will. This usurpation...occurs after the desire has been defeated in deliberation..." (Pears, 1982, p. 56)<sup>12</sup>

Both of these worries, however, depend on two substantive claims, neither of which can be adequately supported. The first is that desire motivates on its own only by being something like a brute causal push. The second is that the agent's standpoint on things is to be identified with the reflective point of view constituted by his or her judgments. Though I can't do full justice to either of these issues in this chapter, I will try to say briefly why I think that there are good reasons to reject both.

In the first case, there has been an increasing dissatisfaction with the dispositional, "direction-of-fit", account of desire, according to which a desire is nothing else but a disposition to behave a certain way given certain circumstances. There has also been a corresponding increase in the discussion and defense of the view, nearly universally held until modern times, that desire is in some sense "aimed at the good". The basic argument for this latter approach is simple: Intentional action is action which can be rationalized from the agent's perspective. Desires are able to rationalize action. An action is rationalized when we understand what made it seem good or reasonable to the agent. Therefore, desires must be things which make actions seem good or reasonable to those who have them.<sup>13</sup>

This argument is perhaps most well known in the negative version given by Warren Quinn (1993). Quinn asks us to imagine that he finds himself with a bare disposition to turn on every radio he comes across, but

---

<sup>12</sup> Perhaps surprisingly, Pears doesn't take this to show that the akratic action is unfree.

<sup>13</sup> I discuss and defend this argument in more detail in the following chapter.

without seeing anything at all good or worthwhile in doing so - he doesn't want to listen to music or to check the speakers or to do anything else. Many philosophers have concluded from this scenario that without any evaluative element, he cannot really be said to desire to turn on radios or to turn them on intentionally when he does so. But if we accept this conclusion, that desires play their motivational role only by being evaluative, the Humean denial that evaluative judgments could play the same role begins to lose its footing.<sup>14</sup>

In response to the second claim, that we should identify the agent's standpoint with their reflective judgments, two things can be said. The first is that there are famous counterexamples to this. Take, for example, Huck Finn. He *judges* that he is acting wrongly in helping to free the slave Jim and sees himself as doing something which will merit hell. Yet we still see his action, based on his more passionate attachments to Jim, as morally praiseworthy. If his normative judgments defined his "true self", then we could only say that he is a (perhaps half-hearted) racist. Instead, we identify him with a deep moral perceptiveness that makes him stand out from his contemporaries, even if he shares their beliefs, and even though that moral perceptiveness is only present at the level of emotion and desire.

So we have some case-based reason to think that the actions which we can attribute to agents in a substantive way are not limited to those which are in line with their reflective judgments. The second thing to note now is that mutual support for this intuitive judgment is provided by the evaluative conception of desire pointed to above. The kernel of truth in the rationalist understanding of identification or attributability is that a *person* is in part constituted by their standpoint on value and what sorts of things provide reasons for what. The fact that we at least sometimes identify a person with their desires over against their judgments is explained, and given justification, by the fact that desires are a distinct *way* of responding to value or reasons.

---

<sup>14</sup> For further defenses of this approach, see (Stampe, 1987), (Tenenbaum, 2007), (Johnston, 2001), (Schapiro, 2009), (Brewer, 2006).

It appears, then, that human motivational psychology is divided into two parts, so to speak, two different ways an agent can respond to things: the reflecting, *reasoning*, part, and the more directly apprehensive desiring part.<sup>15</sup> The purported values or reasons which desire makes us aware of *can* also be considered in reasoning, but desire is its own thing and there is no guarantee that reason can silence desire (or that, objectively speaking, it would be best if it could). What this allows us to say about cases of weakness of will is that they involve a divided agent, and indeed this reflects the phenomenology. In that sense, we cannot *fully* identify the akratic agent with his action, to the extent that we could if he were more unified. But again, this matches the phenomenology, and it does not follow from this that we cannot see the action as intentional. The akratic who acts wrongly may be blameworthy even if his vice is incomplete, just as Huck Finn is praiseworthy though his virtue is incomplete. The division helps us understand why this is so.

## The Will

Once we accept this bipartite division, however, it can be seen that we will need a third. The third part is needed so that, when reason and desire conflict in such a way, the agent can still act. This third part I call the will, insofar as I think of it as a capacity for direct choice or decision. An act of will is an act of deliberate intention formation.

Why is a third distinct part needed? Simply put, because in the sort of conflict I have in mind, desire and judgment have both already played their parts and so there is nothing more they can do. One has already made an all-things-considered judgment about what ought to be done, but this judgment has not silenced the desire to do something else, and so one is continually drawn at the same time to act on that desire. One can deliberate again, seeing if perhaps something has been missed, or try some technique to calm or work up the desire, but any of these moves may leave one in exactly the same spot in which one began. To return to

---

<sup>15</sup> It is an interesting question whether and exactly how this idea fits with the currently-popular dual-process theory in psychology. There is an obvious affinity, but I haven't yet been able to see it through.

my initial example, it may be that I've already judged that the value of getting these papers graded is greater than that of watching the game, but this judgment isn't enough to lead to action, and so I am continually drawn to the value of watching to see if the Warriors can continue their historic winning streak. Neither course seems able to win out over the other and the minutes are ticking away. In such situations, one simply must *decide*, and so it seems that this capacity to decide must be something different from either judgment or desire.

The problem here, it should be noted, is not one of incommensurability or incomparability as these are typically conceived. The issue is not, for example, that there are different types of reason or goodness involved which cannot be compared to one another. The problem is with two distinct *ways* we have of responding to reason or value and their conflict. In most cases of weakness of will, the types of reason *can be* and *are* compared, for the agent has judged that one particular course of action ought to be pursued over others available. There is no other type of comparison or evaluation that can be made; what is lacking is a choice.<sup>16</sup>

Now, I am not saying that in cases of conflict or akrasia the will is *always* used.<sup>17</sup> Sometimes, for example, our desire begins to fade while we are torn between the two options, so that no effort is needed to act in accord with our judgment. Perhaps most often in cases where we act akratically, we are carried forward by the inertia of our desire and act intentionally, but without making any explicit choice or ever consciously *forming* the intention to act in that way. Though I think that such an action can be seen as intentional insofar as it will be rationalized by a desire which is indicative of (part of) the agent's take on what reasons there are for action, more in this case seems needed if we are going to hold the

---

<sup>16</sup> (Raz, 1999) presents an account of the need for the will, in a very similar sense, in order to get around incommensurability. I don't want to deny that the will may *also* be needed to act in cases of incommensurability or underdetermination in practical reasoning; I just don't think that's the problem in the sorts of cases I'm discussing here.

<sup>17</sup> This is one difference between my view and that developed by R. Jay Wallace (2006b) - though in many ways we conceive of the will in the same way, his view is that the will is always involved in genuine rational agency.



agent responsible for their akratic action. What is needed is that the agent could have chosen to resist temptation, that he had the capacity to intend to stick with his judgment - and I think this is part of the phenomenology of *akrasia*.

To perhaps make this more clear, we can return to the terminology from Hieronymi discussed above. Hieronymi distinguished between the question of what one ought to do, which is settled with a judgment, and the question of what to do, which is settled with an intention. The tripartite view as I am thinking of it says that there are three ways in which the question of what to do can be settled. Sometimes it is settled without any explicit reflection, on the basis of desire. Sometimes one settles the question of what to do *in* settling the question of what one ought to do. In these cases, judging that one ought to do something leads straight to an intention to do it, either because there is no conflicting desire or because the judgment silences any conflicting desires. In a third type of case, there is conflict - one has a persistent desire such that the judgment about what one ought to do does not settle for one the question of what to do. In these cases, one has a capacity, the will, to directly choose what to do. Since this capacity decides between desire and an already given conclusion of reason, it must be distinct from either one.<sup>18</sup>

## Putting It All Together

Now that each of the three parts have been discussed, more can be said about how the parts fit together. The main thing I want to point out is the striking similarity with the Platonic view of *The Republic* that was mentioned in the beginning of the paper. In both my view and Plato's there is a fundamental division between reason and desire, with a third part whose role is fundamentally that of a help or an ally to one or the other. There are, of course, some differences. The conception of desire I

---

<sup>18</sup> There is obviously much more that could be said about the will. I've appealed here mostly to the phenomenology and to the space theoretically opened up for the will, and, briefly, to a consideration about moral responsibility. There is also emerging relevant empirical work being done on a concept of willpower. I discuss these issues more fully in chapters five and six.

have appealed to (though admittedly not defended at much length) is broader and more sophisticated than Plato's "appetite". Where Plato has the "spirited" part, I have the will, and where Plato's third part is an ally to reason only, I think the will can side with either reason or desire when the two conflict.

Still, Plato formed his view in response to the intellectualist theory of motivation developed by Socrates, which was undivided and reason-centered. He seems to have developed his own account because of the difficulties the earlier one had in making sense of psychological conflict and other aspects of agency. My main goal here has been to show that he was right to introduce divisions into the soul for exactly these reasons.

I have also tried to begin to say something about the direction a successful positive account should go. The resulting view is relatively messy, and perhaps more complicated than other approaches. But it is messy in the way reality is messy, and I think it fits quite well with folk psychology. Expressed as simply as possible, it says that sometimes we act as we do because we want to, sometimes because we think that we should, and sometimes, especially when these conflict, because we *decide* to, where this involves a distinctive act of will. More importantly, as I have tried to argue above, the view or something like it is *required* if we are to make sense in a plausible way of what happens in important cases of psychological conflict. And this is why Plato was, basically, right.

## Chapter 2: Desire and Reasons for Action: In Defense of the Evaluative Conception

### Introduction

In the previous chapter I gave a series of arguments involving *akrasia* that were meant to help bring into relief and make attractive the approach to moral psychology I want to defend; namely, the tripartite view. The first step towards this distinctive view was the claim that we ought to think of desire as a distinct way of being motivated to act for reasons in addition to reason. I also said that this approach fits best with the ancient idea that desire is itself somehow evaluative or “aimed at the good”.

To make either of these claims about desire is to be embroiled in some of the liveliest ongoing debates in the philosophy of action and practical reason. In particular, two aspects of what I’ve claimed call out for further discussion. First I’ve said that both evaluative judgments as well as desires are able to play some role in motivating, and thus in explaining, action. But it’s not clear what this role is or what sort of explanation is at issue here. Second, I’ve said that to desire something is, in some sense, to see it as good. But, again, it’s not clear what this means, and a straightforward reading of it seems to lead to problems. For example, it is apt to seem overly intellectualized, and not to fit with our practice of ascribing desire to animals and very young children without developed conceptual capacities.

These are the issues I hope to discuss in this and the following two chapters. Here my focus is on the first set of issues, those surrounding desire and the explanation of action. The controversy, at the most basic level, is about how to understand the relationship between desires and the reasons which explain our actions. A series of common positions has emerged. On the one hand, there is the strong Humean position, according to which *all* reasons for action either are or depend in some way on the antecedent desires of the agent.<sup>19</sup> The rationalist, on the other hand,

---

<sup>19</sup> The contemporary *locus classicus* for this position is Bernard Williams’ “Internal and External Reasons” in his (Williams, 1981).

disagrees, claiming that belief, especially moral judgment, is capable of motivating action, even in the absence of a previous desire. The rationalist may go further and claim, as Scanlon does, that desire “plays almost no role in the justification and explanation of action” (Scanlon, 1998, p. 18). Finally, some philosophers want to say that both of the previous approaches are mistaken insofar as they think of reasons in terms of psychological states of any sort, and that reasons are actually best thought of as (typically non-mental) *facts* or *states of affairs*.<sup>20</sup>

In what follows, I’ll try to consecutively map out what I see as the primary motivation for each of these positions and say where I think the right answer lies. Since I believe that reason concluding in judgments can motivate action, I will obviously reject the Humean claim that only desires can motivate action. I’ll also reject the conception of desire that leads to that Humean claim and in doing so argue more directly for the evaluative conception. Whether or not to accept Dancy’s anti-psychologism is a more difficult matter. Though I am attracted to his picture, I don’t ultimately need to take a stand on the issue. What matters most for the purposes of the tripartite view is that we think of desire in a certain way, such that it’s relation to reasons and the explanation of action is in important ways the same as that of evaluative judgments, however it turns out best to think of the latter.

## **Desire and Reasons on the Humean Approach**

I said just now that the main controversy here has to do with the relation between desires and reasons for action. But this is too simplistic, for the notion of a reason is not itself simple. When we explain an action in terms of reasons, there may be more than one thing we are trying to do. It has become common in the philosophical literature to distinguish between the *motivating* reasons I mentioned above, and *normative* reasons. These are supposed to correspond to two different ways we can make an action intelligible: by justifying it on the one hand, and by explaining its occurrence on the other.

---

<sup>20</sup> This claim is forcefully argued in (Dancy, 2000).

The difference is perhaps best seen in cases where the two seem to come apart. Suppose we see Mark walking down the sidewalk and notice that he is carefully and with great difficulty avoiding all of the cracks. Not knowing him well, you ask me why he is doing this. I explain that he believes that stepping on a crack will cause his mother's back to break, which he obviously does not want to happen, and that this is why he's avoiding the cracks. On the one hand, I seem to have given you his reason for exhibiting such odd behavior. On the other, it seems that we can both agree that there is absolutely no reason for him to go on avoiding cracks as he has been. And so it seems that we must be using "reason" in two different ways. We can explain this using the philosophical distinction: he seems to have a motivating reason for what he does (his belief that stepping on cracks will break his mother's back), while he lacks any normative reason (since his belief is false).

Understanding these two ways of speaking about reasons, and especially how they are related to one another, is a key task for philosophy, and it is helpful to think of the basic Humean picture of practical reason as one way of doing so.<sup>21</sup> In particular, we can think of a full Humeanism as I intend to use the label as involving a commitment to two related claims, one about motivation and one about normativity:

1. Motivational Humeanism: Only desires are capable of motivating action.
2. Normative Humeanism: You only have a normative reason to do some action *A* if you desire to *A* or if *A*-ing contributes to satisfying your other desires.

The first of these is a claim about our psychology, about what sorts of mental states are capable of producing action. According to the Humean, only desires can play this role. Why would one accept this claim,

---

<sup>21</sup> The full picture below is most clear in Williams, *op. cit.* Both claims are also present, in a more complicated form in the more recent (Schroeder, 2007). The first, motivational, claim is defended and plays an important role in the moral theories of (Smith, 1994) and (Blackburn, 1998). It is also worth noting that the view I will describe is not identical to Hume's actual view, but the label has stuck. On this point, see (Setiya, 2004) and (Baier, 1986).

especially in light of the fact that we often appeal to beliefs or other states in explaining actions? The answer goes back to Hume's distinction between reason, whose function it is to discover the relations that hold between ideas and the objects of experience, and the passions, which, on his view, are non-representational states whose function it is to cause behavior. Contemporary Humeans will demur from making desires completely non-representational<sup>22</sup>, and will be more likely to talk in terms of belief and desire than reason and the passions, but the basic idea is the same: there are two fundamentally different types of mental state, and the difference consists precisely in the different functions that they play in our psychology and behavior.

This is what Michael Smith has fairly recently called "the standard picture of human psychology", which he explains like this:

there are two main kinds of psychological state. On the one hand there are beliefs, states that purport to represent the way the world is. Since our beliefs purport to represent the world, they are assessable in terms of truth and falsehood, depending on whether or not they succeed in representing the world to be the way it really is. And on the other hand there are desires, states that represent how the world is to be. Desires are unlike beliefs in that they do not even purport to represent the way the world is. They are therefore not assessable in terms of truth and falsehood (Smith, 1994, p. 7).

This way of distinguishing between beliefs and desires is often spelled out in terms of "direction of fit". Beliefs are said to have a "mind to world" direction of fit insofar as a belief is supposed to change in response to how the world is. A successful belief is one that fits the world. Desires, on the other hand, have a "world to mind" direction of fit, insofar as to have a desire is to be disposed to change the world in a certain way. A desire is satisfied or successful when the world is the way that it represents.

---

<sup>22</sup> Desires may be taken by Humeans to be representational in the sense that they may have propositional contents, as in a desire *that I get some ice cream*. But they do not represent things *as being the case*; their contents are not *truth-evaluable*.

Smith develops this metaphor helpfully in terms of the conditions under which beliefs and desires go out of existence:

For the difference between beliefs and desires in terms of direction of fit comes down to a difference between the counterfactual dependence of a belief and a desire that *p*, on a perception that *not p*: roughly, a belief that *p* is a state that tends to go out of existence in the presence of a perception that *not p*, whereas a desire that *p* is a state that tends to endure, disposing the subject in that state to bring it about that *p*. Thus we may say, attributions of beliefs and desires require that different kinds of counterfactuals are true of the subject to whom they are attributed. We may say that this is what a difference in their directions of fit *is* (Smith, 1987, p. 54).

If this picture is accepted, then motivational Humeanism simply follows.<sup>23</sup> Only desires can motivate action because of the type of mental state that they are - anything that motivates action will, it seems, need a world to mind direction of fit, and thus will be, by definition, a desire. So the first tenet of Humeanism depends on a general and, on the face of it, quite plausible psychological distinction between beliefs and desires.

Once the first tenet is accepted, the Humean can go on to use it in arguing for the second. What is needed to bridge the gap between the two is a principle that we can call *Weak Internalism about Reasons*. In perhaps the most well-known statement of this general argument, Bernard Williams puts the principle like this: "If something can be a reason for action, then it could be someone's reason for acting on a particular occasion, and it would then figure in an explanation of that action" (Williams, 1981, p. 106). There is some controversy about how to interpret this, but the basic idea seems to be that there must be *some* connection between what reasons a person has and what they can be motivated to do, or else we would lose the sense of the former's importance and the seeming conceptual link between the two in virtue of which they both involve use of language involving reasons. We can put this principle in the language of our distinction like this:

---

<sup>23</sup> Or so it seems at first glance. For doubts, see the discussion in (Wallace, 2006).

*IR*: If you have a normative reason to *A*, then you must be capable of being motivated to *A*.

Using this principle, we can make an argument from motivational Humeanism to normative Humeanism:

(1) Only desires are capable of motivating action.

(2) You can only be motivated to *A* if you desire to *A*, or if *A*-ing contributes to the satisfaction of your other desires.

(3) You only have normative reason to *A* if you can be motivated to *A*.

(4) Therefore, you only have normative reason to *A* if you desire to *A*, or if *A*-ing contributes to the satisfaction of your other desires.

The first premise is just motivational Humeanism, and the second follows from it, with the addition of an important qualification. The Humean needn't claim that one must always have an antecedent desire for whatever action one does *as such*. It is enough that the action contributes to the satisfaction of one's other antecedent desires. For example, if you want to have something refreshing to drink and remember that there is a beer in the fridge, then you can be motivated to go and get and drink the beer. Although the initial desire was not for beer as such, but only for something refreshing, the belief that drinking the beer is the best way to satisfy that desire is sufficient, other things being equal, to motivate you to get the beer. This disposition to do what is a means to the satisfaction of one's desire is taken to be just part of what it is to have a desire. In cases like these it is common to *identify* the (motivating) reason for action with a belief and desire pair - your reason for getting the beer was that you wanted something refreshing to drink and believed that the best way to satisfy this desire was by getting the beer. Still, given the Humean psychological picture, the motivation in such a case comes from the desire. The belief is simply needed, so to speak, in order to activate the desire. The third premise is just our principle of weak internalism about reasons, and the fourth is normative Humeanism, which follows from (2) and (3).

So we have an argument for what I've called a full Humeanism. Taken together, the two Humean theses imply that desire has some role to play in both the explanation and justification of *all* action. Of course it is possible to be only a partial Humean, by, for example, accepting the claim



about motivation but rejecting the normative claim. The easiest way to do so would be to reject the third premise above, i.e., internalism about reasons.<sup>24</sup> Since moral psychology is my primary concern, I'm most interested in undermining the motivational claim. But since the normative claim seems to depend (as in the argument above) on the motivational claim, undermining the latter will tend to undermine the former as well.

## **Desire and Evaluation: Against Humeanism**

The basic problem with the Humean account in my view is with its conception of desire. And it seems to me that the best way to attack the Humean picture of practical reason is indirectly, following an argument by Warren Quinn: Let's grant to the Humean that desire plays at least some role, at least some of the time, in explaining action. What must desire be like in order to do that?<sup>25</sup> Quinn argues that, conceived in the dispositional direction of fit way given above, desire couldn't possibly play such a role. Rather, he thinks we should think of desire as typically involving an evaluative element, and that it is this element which allows it to play whatever role it does in explaining action (Quinn, 1993). In what follows, I'll explain Quinn's argument and how I think it undermines Humeanism. I'll then address some Humean responses given to the argument by David Sobel and David Brink.

## **Desire, Rationalization, and Quinn's Radio Man**

The Humean picture of desire, as just discussed, is as a state with a distinctive direction of fit, where this in turn is spelled out primarily in terms of one's behavioral dispositions given one's circumstances and other psychological states. This is the essence of what it is to be a desire. Desires may have some other features, like a certain phenomenology, or

---

<sup>24</sup> For a more complicated way of combining the Humean motivational claim with the idea that there are objective moral reasons, see (Smith, 1994).

<sup>25</sup> There are other possible approaches. One could attack the direction of fit approach to desire more directly, as in, for example (Setiya, 2007, pp. 50–51). Or one could grant the Humean approach to desire but deny that either of the Humean theses follow from it, as in, for example, (R. Wallace, 2006).

frequently come along with other types of states, but these will be irrelevant for the sake of making it the case that the state is a desire. Quinn asks us to imagine that he has found himself with a desire, in just this sense, to turn on all and any radios in his vicinity which are turned off. The desire is basic, in that it does not serve to satisfy any other desire he has, and it is unassociated with any belief in the goodness or reasonableness of his doing so. He is not moved to turn on radios in order to hear music or the news, or to test how good the speakers are, or in order to do anything else. Indeed, we may even imagine (though Quinn does not include this detail) that he finds this disposition completely bizarre and stupid. Could the fact that he has such a disposition ever provide a *reason* for him to turn on radios? Would he act intentionally in doing so? Does it seem right to say that he has a *desire* to turn on radios?

Quinn thinks not. Of course the example is puzzling because it is hard for us to imagine a person acting in such a way without acting intentionally, and so we think that he *must* be. But the point is that we begin to lose our grip on this as soon as we begin denying of him all the psychological aspects or concomitants of typical desires that are strictly speaking, from the Humean point of view, mere extras. Once we do that, it no longer seems quite right to say that he wants to turn on radios, since he doesn't see any point in doing so. And without that even minimally evaluative element, we cannot say that he intentionally turns them on.<sup>26</sup> The example works at an intuitive level - it is designed to get us to judge based on a particular case that the Humean conception of desire is too thin

---

<sup>26</sup> For this way of putting the worry, see also the discussion in (Stroud, 2013): "The point is brought out by Warren Quinn's striking example of someone with a strong but apparently unintelligible urge to turn on radios. The man is said to see nothing in turning on radios and not to favor or evaluate positively in any way their being on. He does not seek the music or news or commercial advertising he gets when he turns them on. He does not pursue the goal of having every radio in the world turned on. But he more or less irresistibly moves to turn on every radio he sees. Quinn's main point with the example is that this man cannot be said to have a desire to turn on radios. We cannot explain his turning on a particular radio by saying he wanted to turn it on, or saw something in its being turned on. With that evaluative element missing, we cannot understand him as intentionally turning on the radio" (p. 109).

- it neither captures our ordinary concept nor seems to be the sort of thing that could be or provide a reason for action. A little further theoretical reflection leads to the same conclusions and thus provides further support for our intuitions about the particular case.

It might seem as if the problem here is primarily about whether Humean desires can provide *normative* rather than motivational reasons. And it is true that Quinn's main objection is to a bare instrumentalism about practical reason, which says that one's normative reasons are limited to the reasons one has to do those things which satisfy one's desires. But I think the worry goes deeper than this. Both motivating and normative reasons are meant to be explanatory. Presumably it is in virtue of this explanatory function that we consider them reasons, since we use the term "reason" in other contexts than action where explanation is involved. For example, we can talk of the reason that the tides change as they do, or the reason that seven is a prime number.

So what *kind* of explanation is involved in giving motivating and normative reasons? It might seem that they must involve completely different sorts of explanations, since we have already seen that they may come apart. But this doesn't seem quite right either, since it should be possible, on the face of it, for one's normative and motivational reasons to coincide - to have a genuine reason to act and then to act for that very reason. In either case, what distinguishes action explanations of either sort seems to be the type of intelligibility that they bestow on actions; namely *rational* intelligibility. Davidson makes this point at the very opening of perhaps his most famous paper on action: "What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did? We may call such explanations *rationalizations*, and say that the reason *rationalizes* the action" (Davidson, 1980, p. 3).

It's clear that he is referring here to what we've called motivating reasons. An immediate objection arises, however: What about cases where the agent is in some way mistaken - how can whatever we give as a reason in such a case rationalize his action? Recall the example above of Mark, who avoids stepping on cracks out of a mistaken belief that stepping on one will break his mother's back. It may be denied that such a belief could

really make his actions rational. Nevertheless, it does make them rationally *intelligible*. The idea here is that, in this case, what keeps his action from being rational is just that his belief happens to be false. But we can see how such an action could *seem* rational, *from the agent's point of view*. The right way to draw the contrast between normative and motivating reasons, then, seems to be this: both rationalize actions, but normative reasons do so from an objective and general perspective, from the standpoint of how things actually are, whereas motivating reasons do so from the more limited perspective of the agent who acts.<sup>27</sup>

In this way we see that the notion of an explanatory reason seems to depend conceptually on that of a normative reason. Perhaps surprisingly, Davidson at one point seems to contradict this when he remarks that “The justifying role of a reason...depends on the explanatory role, but the converse does not hold. Your stepping on my toes neither explains nor justifies my stepping on your toes unless I believe you stepped on my toes, but the belief alone, true or false, explains my action” (Davidson, 1980, p. 8). But it is clear in this context that he is talking about some sort of (efficient) causal dependency, and it is true in that sense that I am only able to act on some reason insofar as it “registers” in some way in my psychology. But on the other hand, my belief that you stepped on my toes can explain my action of stepping on your toes, even if it is a false belief, only insofar as there is some rational connection between the two; insofar, that is, as the fact that one’s toes have been stepped on gives, or is thought to give, one at least some reason to respond in kind.

But now we can state Quinn’s point (or at least the point I want to make from his example) again in theoretical, rather than simply intuitive, terms. If desires are to explain actions, then the Humean notion of desire

---

<sup>27</sup> Actually the motivating perspective can be even more limited than this. For example, in cases of *akrasia* like those discussed in the first chapter, an agent may act on a desire even though he judges that the desire gives him no good reason to act. In such a case (I want to say) the action is rationalized from the view of the agent’s desires but not necessarily from the agent’s perspective as a whole. In such a case there is internal conflict - the agent is divided - and so we cannot fully identify his perspective with that of his desire (or of his judgment). More on this in chapter six.

must be incorrect, for it does not explain how a desire could make an action rationally intelligible in either sense. Finding oneself in a brute dispositional state that involves a reliable tendency to turn on radios does not show how doing so would be rational, from some more general perspective, or from the agent's own.

But at least some desires do seem to be capable of playing this role. Given his false belief, Mark's desire not to harm his mom does make his action intelligible in the right sort of way. So what is the difference between everyday desires, like Mark's, and Humean desire? The proposal is straightforward: Mark's desire can help rationalize his action because it is *not* a mere dispositional state, but involves some evaluative element. In desiring not to bring harm to his mother, he is not just *disposed* to avoid harming her, but he also in some sense thinks that harming her is *bad*, and that avoiding causing her harm is *good*. On the face of it, there is nothing strange about this suggestion (though the next two chapters will deal with difficulties involved), and the presence of such an evaluative element seems clearly relevant to the task of making an action rationally intelligible. If a bare behavioral disposition is unable to explain action in the sort of way we think that desires typically do, and if it seems that desires often *do* involve some evaluative element, then we seem to be led to the conclusion that all desires, at least insofar as they are able to explain an action as intentional, should be thought of as involving some evaluative element.

But this creates serious problems for the Humean, and not simply because it undermines his favored conception of desire. Since motivational Humeanism depends on that conception of desire, it's no longer clear why we should accept motivational Humeanism. It seemed credible that only desires can motivate because of the strong distinction drawn between beliefs and desires. But if it is *essential* to a desire's ability to be part of a motivating reason that it contain some evaluative element, then it seems we should conclude that its motivational source either just is, or depends essentially on, that evaluative element. But now it's hard to see why judgments would not also be capable of motivating, since they too can include an evaluative element. I can desire to help someone, but I

can also believe that it would be best to help him, and it's hard to see why the latter should not be just as capable of motivating as the former.

It seems that the best way to distinguish, then, between beliefs and desires is not in terms of direction of fit, but in terms of non-reflective essentially evaluative states (desires) and those that are more reflective and only possibly evaluative (belief).<sup>28</sup> And this is exactly what the tripartite view claims, that practical judgment and desire are both capable of motivating and explaining action, insofar as both can be ways of (in some sense) representing things or states of affairs as good.

## **Objections and Replies**

### **Copp and Sobel on Quinn and Scanlon**

In spite of the fact that several defenders of the evaluative conception of desire appeal to Quinn's example for support, there has been surprisingly little (so far as I can tell) direct engagement or response by the Humean side.<sup>29</sup> One exception to this, fortunately very thorough, is David Copp and David Sobel's "Desires, Motives, and Reasons: Scanlon's Rationalistic Moral Psychology". They are interested in Scanlon's moral psychology more generally, so much of their discussion will be beyond the scope of the concerns of this chapter. Nevertheless, they do spend some time attacking his use of the Quinn example in defending an evaluative conception, and it seems to me that they raise several worries that need to be answered.

The first objection is that the appeal to rationalization above is, in spite of its apparent modesty, still overly intellectualized. I claimed that the disposition to turn on radios cannot be seen as able to properly explain the action of turning on a radio because it cannot be seen as able to help

---

<sup>28</sup> This is meant to be rough - the nature of desire will be discussed more deeply in the following two chapters.

<sup>29</sup> In (Schapiro, 2014), Tamar Schapiro lists several examples of defenders of the evaluative conception, including (Brewer, 2009), (Darwall, 2006), (Scanlon, 1998), and (Tenenbaum, 2007), and says that "nearly all evaluative outlook theorists cite Warren Quinn's seminal article...as inspiration" (p.134).

rationalize that action. And I claimed that it cannot help to rationalize the action because the radio man sees nothing good in either the disposition or in its object, i.e., turning on radios. But, the objection goes, this argument is subject to strong counterexamples if taken to show either that (a) the disposition thereby does not count as a desire, or (b) the disposition, whatever we decide to call it, cannot explain the action as intentional.

In the first case, it is pointed out that we attribute desires to both very small children and to animals, neither of which, arguably, are able to have concepts like that of “the good”. If this is so, then it might be argued that they are in the same position with respect to their “desires” as the radio man - they are incapable of seeing the involved dispositions or their objects *as* good. And if that is enough to call into question their status as desires, then we must deny that either animals or very small children have desires. And that seems implausible.

In the second case, it is claimed that the argument also rules out the possibility of acting on a mere whim or urge. They continue: “But it is not bizarre to have basic, unmotivated desires. Sometimes we just ‘feel like’ doing things...And in many cases we find ourselves unable to articulate any reasons why we desire what we desire” (Copp & Sobel, 2002, p. 259). The argument is this: Sometimes we act intentionally on a whim or an urge. In these cases, we don’t see anything good in what we do; we just feel like doing it. Therefore, an urge or a whim (and thus, a desire) can explain action in a way that shows it to be intentional without involving any evaluative element.

These are important worries that need to be addressed, but they cannot be addressed at this point. Importantly, they don’t directly address the anti-Humean argument given in favor of the evaluative conception, but instead object to (one understanding of) that conception. In particular, they both seem to presuppose that the relevant way of “seeing something as good” must involve at the least a tendency to explicitly apply the concept “good” to what is desired. Therefore, they can’t be fully addressed until we know a bit more about how exactly to understand the proposed evaluative aspect of desire, and this is the main aim of the following two chapters.

The other two objections they raise can be addressed here. In these objections, Copp and Sobel are willing to accept the claim that the radio man doesn't count as having a desire. Instead, they try to argue that accepting that claim neither counts against the Humean picture nor supports the evaluative picture.

It is no harm to the Humean because the Humean conception of desire need not be so sparse as the example supposes:

Functional states can be more or less complex. On any plausible view, having a desire to do something is normally much more complex than simply having a disposition to do it. We do not normally desire to blink, although we are disposed to do so...[O]n a very thin construal of the state that the radio man is in, even a Humean could agree with Scanlon that the radio man's 'urge' is too barren to qualify as a 'desire' to turn on radios. But it does not follow that what is missing from the radio man example is an element of evaluation, a taking something to be a reason for turning on radios. Perhaps what is missing is a thick enough set of dispositions to constitute a desire.

That is the argument that denying the radio man a desire does not count *against* the Humean. Why does it also not provide support for the non-Humean? Because not only is it *possible* that what is missing from the radio man's dispositional state is not an evaluative element, but, more positively, adding in an evaluative element doesn't seem to help. Suppose we grant that his disposition to turn on radios is just an impulsion, something akin to a bodily reflex, and that is all. Then, the argument goes, it's hard to see how adding in an evaluative element like Scanlon's recurrent tendency to think of the reasons that one has to do so will change anything: "On the contrary, we think that adding such a thought pattern to the radio man's impulsion to turn on radios would not manage to turn the impulsion into a desire. It would merely add a second aspect to the man's obsession with radios" (p.263). If this were true, then it seems that the evaluative element in desire, even if it were present, would not play as central a role as the anti-Humean argument alleges.

Taken together, the arguments above by Copp and Sobel can be helpfully put into the form of a dilemma:

1. Either Quinn's radio man counts as having a desire which can help explain his action as intentional, or he doesn't.



2. If he does, then the example does not count against the Humean view, but may count against the evaluative conception.
3. If he does not, then the example does not count against the Humean view, but may count against the evaluative conception.
4. So, in either case, the example does not undermine the Humean conception of desire, though it may at least partially undermine the evaluative conception.

## Response

As I said above, the first horn of Copp and Sobel's dilemma cannot be fully addressed until we have a more developed evaluative account of desire. What about the second? The argument for it grants that thinking of desire as a state involving a bare disposition to act is unsatisfactory and that something more needs to be added. It goes on to make two further claims: that the Humean can say what more needs to be added, and that it won't turn out to be anything evaluative, since adding an evaluative element wouldn't even help. Let's look at each of these in turn.

Whether the Humean can properly supplement the dispositional understanding of desire in terms of direction of fit must, unfortunately, be a matter for speculation. This is because, so far as I know, none have ever tried systematically to do so. I am skeptical that it can be done on the grounds of principle given above; namely, that what we do in appealing to desire to explain action is make clear some positive light in which the agent saw it, which is precisely the element that is missing in this case. We can see this by looking at the one fix that Copp and Sobel mention without really defending.

They say that what is needed is a "thick enough pattern of dispositions", and they suggest that "a more fully articulated functionalist account of desire in the intuitive sense might view the desire for coffee ice cream, for example, as a disposition to seek coffee ice cream, to plan how to get some, to notice it when it's available in ice cream stores, to eat it, to think about it, and so on" (p. 263). Will this set be thick enough? I think not. Of course, it's very difficult to imagine someone having all the relevant dispositions with respect to turning on radios without also desiring to turn on radios, but it's hard to see how simply adding more

complicated behavioral tendencies will get around the original problem. No matter how many we add, it will still be the case, *ex hypothesi*, that the radio man sees nothing desirable or worth doing in acting in accord with his dispositions, and thus doesn't really *want* to turn on radios.

Copp and Sobel admit that there is a contrast between desire and reflex, both of which involve behavioral dispositions. On the basis of the argument above, it seems that they conceive of the difference as a matter of complexity of disposition. But the difference between the two seems qualitative rather than quantitative. It is true that for us being struck with a hammer under the knee cap causes our knee to extend - a reflex constituted by a relatively simple disposition. But I think we can conceive of creatures for whom being struck below the knee cap would instead cause them reflexively to stand up and walk in a circle around the nearest object, something that we could typically only do intentionally. This shows that we can't get to what is distinctive of desire by simply adding more complex behavioral dispositions.

Perhaps a more plausible attempt to supplement the Humean account of desire would involve adding in pleasure somewhere. As a rough statement, it might be claimed that to desire to do *A* is to be disposed to do *A* in virtue of the pleasure that one is disposed to feel when *A*-ing. A benefit of the addition is the fact that there does often seem to be an element of pleasure in the satisfaction of desire. It would also allow the Humean to explain why it is that the radio man does not count as desiring to turn on radios - for presumably he receives no pleasure from doing so or the thought of doing so.

One worry about this suggestion is that it's doubtful whether the connection between desire and pleasure is really quite so strong. There is no contradiction on the face of it in thinking you could desire something which you know will give you no pleasure. Another worry is that it will collapse into an implausible hedonism. But the most serious worry is whether the new account can really be thought of as Humean in the relevant sense.

Quinn himself presses this worry after giving his example. The problem, as he sees it (and so it seems to me), is that "it is a mistake to think of the concepts of pleasure and displeasure as purely descriptive,

psychological concepts. To call an experience pleasant or unpleasant is already to bring it under an evaluative concept" (Quinn, 1993, p. 37). In other words, to put pleasure into one's account of desire is already to sneak in an evaluative element. And then it will seem that to desire something *is* necessarily to represent it as some way good; namely, as pleasant.

Quinn notes that it is possible to give an account of pleasure that is not evaluative. But such a conception will have the same problem rationalizing action as the more bare dispositional account. For example, Quinn discusses the view that "a pleasant experience is, roughly speaking, one whose intrinsic character makes an agent want to prolong it", where wanting to prolong it is spelled out in the more bare dispositional sense. Such a state would seem to lack the required reason-giving force:

...why should anybody want to be in such a state? Suppose I tell you that if you start scratching your ear the experience will strongly dispose you to keep on scratching. Does this by itself give you reason to want to scratch? Conceived as a kind of psychological inertial force, pleasure takes on a somewhat sinister aspect. This is because the account leaves out the salient thing: that an agent wants to prolong a pleasant experience precisely because it is pleasant - because it feels good. Pleasantness is not merely that which brings about a prolonging disposition, it is what makes sense of it (p.37).

To put the response more succinctly: Either pleasure is conceived of as involving some evaluative element or it is not. If it is, then the view is not properly Humean since it would make desire's reason-giving force dependent on whether it picks out some aspect of the desired action or situation as in some way good. And then it would be puzzling why evaluative beliefs could not play the same role. If it isn't, then we have just added another to our growing list of behavioral dispositions while still failing to show how they could shed light on the way that desire-explanations are taken to rationalize action.

What about the claim that adding an evaluative element won't change the intuition that the radio man doesn't count as having a desire, or as acting intentionally to turn on radios? Copp and Sobel say that adding an evaluative element won't change his "impulsion" to a desire, but will only "add...to his obsession". This betrays a misunderstanding of the argument

being made. The problem alleged with the Humean conception of desire is not that its effects cannot be distinguished from compulsion or obsession, but that it cannot fulfill an essential task of intentional explanations by showing the positive light in which the agent saw his action.

That these are different issues is made clear by reflection on cases of legitimate compulsion or obsession. In particular, it seems that compulsive actions will usually still count as intentional. They may not, on particular occasions and depending on the strength of the compulsion, be considered free, but they remain intentional, at least on an intuitive level.

This intuition is supported by a more theoretical understanding. The *Diagnostic and Statistical Manual of Mental Disorders* defines the compulsive aspect of obsessive-compulsive disorder as involving, in part, “behaviors or mental acts...aimed at preventing or reducing anxiety or distress, or preventing some dreaded event or situation” (American Psychiatric Association, 2013, ch. 14). So in these cases there seems to be a clear evaluative element - either that of discharging some urge that seems to the agent problematic, or to prevent some more specific bad event. For example, compulsive hand washing is a relatively common manifestation that is driven by obsessive thoughts about being contaminated or catching disease. We may not understand these obsessive thoughts, or think them irrational (and typically the compulsive hand washer also recognizes them as such), but an action which is motivated by them seems to be intelligible in the sense we’ve been discussing - catching disease is bad and should be avoided.

If this is so, then the problem with the radio man is not that his “desire” to turn on radios seems obsessive or compulsive. The problem is worse than that, because the desires of the obsessive-compulsive still count as desires in the full-blooded sense at issue, while the radio man’s disposition does not. So, in spite of the rejoinders of Copp and Sobel, we seem to be stuck with the conclusion that desires can only play the role that they seem to play in the explanation of action if they involve an evaluative element. And that seems to require rejection of the Humean psychology.

## Desires and Anti-Psychologism about Reasons

So far I have assumed that desires do play some role in the explanation of action, especially when it comes to specifying an agent's "motivating" reasons. In doing so, I see myself as having followed the appearances; after all, we do frequently explain an agent's actions by mentioning his desires. But some philosophers have urged that the appearances here are deceptive. They recommend an anti-psychologistic understanding of reasons for action. In other words, they urge that we should reject (in the words of Jonathan Dancy) "the claim that the reasons for which we act are psychological states of ourselves" (Dancy, 2000, p. 98).

Pressure towards this position comes from two directions. On the one hand, pressure comes from the evaluative conception of desire itself. If desire involves some evaluative element, then it might seem more natural to say that the reason for action is actually the object or state of affairs that is evaluated. This comes out clearly in Scanlon's discussion of his notion of desire in "the directed-attention sense" (which I will briefly discuss in more detail in the following chapter). According to Scanlon, ordinary desires involve one's attention being directed "insistently toward considerations that present themselves as counting in favor of [the object of the desire]" (Scanlon, 1998, p. 39). He illustrates this with a desire to get a new computer. In desiring to get a new computer, one's attention will be insistently directed towards thoughts like that one's current computer is out of date, or too slow, or that the new computer is more aesthetically pleasing, etc. If this is what such a desire in part consists in, then it seems like what provides one's reason in such a case are precisely the considerations that one's attention is drawn toward, and not the desire itself. If my desire for a new computer involves a fixation, say, on the fact that the new computer will make completing my work easier and more enjoyable, and I act on such a desire, then surely the reason for which I act should be understood in terms of those facts which counted in favor of my doing so, and not the fact that I had a desire.

The other approach to this position is the one given by Dancy himself, which depends not on the nature of desire, but on a more general consideration about the connection between motivation and normative

reasons. Interestingly, the consideration is similar to the modest internalism about reasons that drove the argument for normative Humeanism. Recall that the idea there was that one only has a normative reason to do something if one can be motivated to do it. The anti-psychologistic principle is perhaps slightly stronger, but no less intuitively appealing. It says that if one has a normative reason to do something and does it, it must be possible that one's normative reason and one's motivating reason are one and the same. As Garrard and McNaughton put it in a discussion of moral motivation, it must be possible that "the reason why you ought to do an action and the reason why you do it can be the same" (Garrard & McNaughton, 1998, p. 48).<sup>30</sup>

This leads to an anti-psychologism once we reflect on what sorts of things normative reasons typically are. They are not typically psychological states, but facts, or states of affairs, or features of the world. Suppose, for example, that you have the true belief that the building you're in is on fire. Surely you then have a reason to exit the building as quickly as possible. But what is the reason in such a case - your belief, or the fact that you believe in? There are strong reasons for thinking it must be the latter. The reason you have to flee has to do with the danger you are in if you do not. But your belief is no danger to you - what is a danger to you is the building's being on fire! Or consider a moral context: suppose you see a friend in need and desire to help him. On the way of understanding desire I want to advocate, your desire in this case may just be, at least in part, a way of recognizing the reason his being in trouble gives you to help him. If you go on to act on a desire so conceived, then it seems you have acted as you were morally required to do, that your action has moral worth. But we can only understand your action this way if we understand you as being motivated by the fact that your friend was in need, and not by the fact that helping him was a way of satisfying a desire you had.

---

<sup>30</sup> They call this principle "Korsgaard's Constraint" since it plays an important role in her discussion of practical reason in (Korsgaard, 1986).

So we seem led to the following conclusions. Normative reasons are not typically psychological states, but facts about the world.<sup>31</sup> It must be possible for one's normative reasons and one's motivating reasons to coincide. Therefore, we should not think of one's motivating reasons as psychological states, but rather as facts about the world. And this means that desires are typically not reasons for action.

What should we say to such an argument? The most common defense is actually an offense; i.e., to point out that the view which one is left with if one accepts the argument has its own rather serious problem. The problem is that it makes sense to think of one's motivating reason as being a fact or state of affairs in cases where one has a normative reason to act and is then motivated to do so. But what led us to distinguish between motivating and normative reasons in the first place was precisely the fact that they occasionally come apart. Remember Mark, who avoided stepping on cracks in the sidewalk out of the belief that doing so would break his mother's back. Using the understanding of reasons above, we'd have to say that his reason for avoiding stepping on the cracks is *the fact that doing so would break his mother's back*. But of course there is no such fact, and that is why we said that in that case Mark had a motivating reason for avoiding the cracks, but no normative reason to do so. Now it seems as if we must say that he also had no motivating reason. But that seems clearly false, since he acted intentionally. So we seem forced back into saying that his motivating reason was not any state of affairs but his belief. And if his belief is his motivating reason in the cases where he has no normative reason, why shouldn't they also be his motivating reason in cases where he does?

Dancy suggests two ways for getting around this problem (Dancy, 2000, chs. 6-7). The first is to take a move made by some philosophers in the philosophy of perception, where a similar problem arises. In that

---

<sup>31</sup> I say "not typically" because sometimes these facts about the world are facts about one's psychological states. James Lenman gives the example that a climber's belief that the cliff is crumbling may count as a reason not to climb it if he has a tendency towards nervous twitches that would make climbing with such a belief much more dangerous (Lenman, 2011, p. 19).

context, the worry arises due to the possibility of perceptual illusion. Many philosophers want to say that, in seeing some object, the object itself is part of the content of our visual experience, rather than that there is some state whose content can be specified independently of the object which in the good case happens to be caused by the object. On this view, when I see a red ball, say, I am directly visually acquainted with the red ball. The problem is that I may, for example, instead be having a very vivid hallucination of a red ball. Such a hallucinatory state may be completely first-personally indistinguishable from the state of veridical seeing, and so it seems *prima facie* plausible to say that the two states have the same visual content. But if that's true, then I am never directly acquainted with the red ball, even in the good case. Instead, in both cases I have an identical psychological state, which in one case matches and is properly caused by the way the world is and in the other case is not. The disjunctivist solution to this problem is simply to deny that indistinguishability implies sameness of content. In the good case, one is in a state that has the red ball as part of its content, and in the bad case one is in a different psychological state altogether, whose content can perhaps only be described as being *as if* one were seeing a red ball. The analogous solution in the case at hand would be to deny that the possibility of error in practical reasoning implies that one's reasons are always psychological states. In the good case, one is motivated by the actual state of affairs which provides one with a normative reason to act. Only in the bad case, like Mark's, is one motivated by one's belief or desire, since there is no normative reason.

The other way out, which Dancy himself favors, is to deny the factivity of reasons-explanation. This is to say, in what Dancy calls an "offending formulation", that sometimes agents act in the light of a reason which is no reason. In other words, the proper way to explain an action is in terms of the state of affairs the agent took to obtain even though the state of affairs is not actual. The way Dancy thinks we should put this in the case above is that Mark avoided stepping on the cracks because, as he believed, doing so would break his mother's back. The point of the appositive phrase is to signal that it was the (non-actual) connection between stepping on cracks and his mother's back breaking that really explains his



action, rather than his belief, but in such a way that doesn't commit us to thinking this belief was true. It may seem strange that a non-actual connection or state of affairs could explain anything at all, but this is only clearly true in the case of *causal explanation*. Insofar as we think of reasons explanation as not primarily about efficient causation, it's less clear that a non-actual state of affairs couldn't play that role.

Obviously there is a lot more that could be said here about these solutions and their relative merits. I happen to think that both are workable and not implausible ways of saving the anti-psychologistic view. The key thing to note, however, is that the tripartite view as such doesn't require taking a stand in this particular debate. The hallmark of the view is that desire and evaluative belief are both, insofar as they are ways of responding to reasons or representing courses of action as good, capable of playing similar roles in our understanding of action explanation, whatever role that happens to be. And I think it is obvious that they must play some role. Even if one grants that normative reasons are states of affairs, we can only be rationally moved by those states of affairs insofar as we in some way cognize them, and one way to do this is to believe that the state of affairs obtains. Similarly, insofar as it seems plausible that beliefs could be motivating reasons, this is just because of the way that belief relates to possible states of affairs which would be normative reasons. So one might try to grant the substance of Dancy's point by saying that to be motivated by the state of affairs which constitutes a normative reason to act *just is* to believe that such a state of affairs is a reason to act and then to be motivated by that belief. In any case, if we think of desire as at least constitutively involving evaluation or reasons-responsiveness, then we will be able to say the same thing about desire; desire plays a role in action explanation insofar as it is a way that an agent can come to recognize, and then come to act on, the desirability of something in the world. And that parity claim is the essence of the first part of the tripartite view.

# Chapter 3: Two Constraints on an Evaluative Account of Desire, and Some Failed Attempts to Meet Them

## Introduction

In the previous chapter, I discussed desire and its relation to explanatory reasons for action. My main aim there was to argue against a common “Humean” conception of desire, which thinks of it entirely in terms of “direction of fit” or dispositions to act. I claimed on the contrary that desire can only explain or motivate intentional action insofar as it involves an evaluative element or represents its object in some way as good. An added benefit to undermining this Humean understanding of desire is that it weakens Humean doubts about the ability of reason to motivate, since reason can also represent things as good.

In recent years, several other philosophers have made similar arguments and tried to defend some or other version of what we can call an evaluative conception of desire.<sup>32</sup> But it turns out that giving the details of a plausible account is quite difficult. In this chapter, I want mainly to say why that is, and to show how several recent attempts have failed. In the next, I will try to give my own account, which I hope will fare better.

The difficulty is in understanding exactly how to interpret the metaphor often used of desire’s “aiming at the good”. I will argue that there are two constraints, in particular, that must be met by a plausible account, and that there is a tension between them, such that clearly meeting one makes meeting the other more difficult. In the next section, I will explain these constraints and the tension between them. Then, after briefly looking at Scanlon’s account, I will examine in detail an influential account from Sergio Tenenbaum (who himself closely follows an earlier account of Dennis Stampe), and show how he fails to satisfactorily meet

---

<sup>32</sup> See, for example, (Stampe, 1987), (Scanlon, 1998), (Brewer, 2006), (Tenenbaum, 2007), (Hawkins, 2008), (Schapiro, 2009).

them. I will finish with a brief remark on the direction in which I think an evaluative conception of desire should be developed.

The aims of this chapter are modest, but necessary. Work on the philosophy of desire is still fairly undeveloped, especially on a “guise of the good” approach. Several accounts have been put forward, but none have been widely discussed or accepted. What I hope to do is provide a way of evaluating prospective accounts. As I hope to show, more work needs to be done. I will try to do some of this work in the next chapter.

## The Two Constraints

What might it mean to say that desire is, in the relevant sense, evaluative or “aimed at the good”? A very straightforward sort of response would connect desire with belief. It might be said, for example, that to desire  $p$  is to believe that  $p$  is good. This would certainly show that desire involves evaluation, and it even seems plausible in very many cases - when I desire my friend’s well-being, or to complete my next paper, or to eat an ice cream cone, I do so believing that there is something good about each.

And yet, the claim seems too strong.<sup>33</sup> I certainly don’t need to *consciously* believe that  $p$  is good to desire it. And since it seems *possible* to have the desire while lacking the belief, the suggestion can seem a bit *ad hoc*. Worse, there are cases of weakness of will where an agent will explicitly deny having the belief that there is anything good about their course of action. Acting on desire against one’s normative judgment can seem compelled if desire is just a causal disposition, but this would not be the case if desire itself somehow represents an agent’s take on value. But to play this role, desire must be distinct from belief.

There is a bigger problem, however. Desire is an attitude that is also had by very young children and by non-human animals. So it seems that any account in which desire involves an attitude toward some evaluative *content* will be overly intellectualized. Nothing without a *concept* of

---

<sup>33</sup> In addition to other possible complaints. (Lewis, 1988), for example, argues that identifying belief and desire leads to inconsistencies in decision theory.

goodness can believe *that* something is good, for example. So this initial account would seem to rule out animals and small children as having desires.<sup>34</sup>

The same problem arises for the most common attempts at weakening the desire-as-belief model. Rather than believing that *p* is good, it might be suggested that desiring *p* only involves the *appearance* that *p* is good, or it *seeming* as if *p* were good. It remains unclear how it could appear that *p* is good if one lacks the concept of the good. I will return to this suggestion in the next section, but for now the first constraint on a plausible evaluative conception of desire should be clear: it cannot be overly intellectualized. It must explain how desire is distinct from evaluative belief and still attributable to non-human animals.

The attempt to meet this constraint may cause us to reconsider the metaphor of “aiming at the good”, and weaken the sense in which desire is taken to be evaluative. At the limit of such weakening, we might claim that to evaluate something positively *just is* to desire it. This view, however, seems *too* weak. The sense in which desire is evaluative becomes trivial rather than substantive.

It is also no longer really an analysis of desire, exactly, but an analysis of value by way of desire. We would still need to say more about what it is to desire something, and since we could not non-circularly appeal to value, we would then likely be pressed back towards the causal-dispositional approach. However, the whole impetus for the move towards the evaluative approach was to distinguish desire from mere causal disposition, to fit desire’s motivational efficacy into an understanding of action as in some important way *rationally*, rather than merely causally, intelligible. We are led, then, to the second constraint on

---

<sup>34</sup> For the same basic worry, see (J. D. Velleman, 1992, p. 7). It should also be noted that I am appealing here to one common and natural, but by no means uncontroversial, understanding of concepts, one which connects them with (so far as we know) characteristically human capacities for language and rational thought. Even if one rejects this understanding, one will have to be wary of making desire overly-sophisticated and give some account of what animal desire consists in and how it is connected with more sophisticated judgments about the good - this is what I will try to do in the next chapter.

a plausible evaluative conception of desire: it must show how desire is evaluative in some substantive sense. Not just any urge or disposition can be seen as providing a *reason* for action, and the account of desire should make clear why this is so.<sup>35</sup>

One currently popular way of trying to meet this condition, without running aground on the other constraint of not being overly intellectualized, is to appeal to an analogy with belief. It is claimed that desire is aimed at the good in the same way that belief is aimed at truth. On the face of it, this allows us to think of goodness as a distinct and substantive goal of desire without involving, for example, the objectionably sophisticated requirement of involving some belief about goodness - for belief's being aimed at truth does not involve either reducing truth to belief or making the concept of truth figure into the content of belief.

I will return in the fourth section to one particular use of this idea, but a general sort of worry is I think immediately obvious - namely, that without significant further development, the suggestion is entirely too unclear. It is an attempt to make one metaphor more clear by pointing to a similarly deployed metaphor in a different context. But *that* metaphor is itself controversial and at present poorly understood. While it seems to me that more philosophers are willing to accept that belief in some sense aims at truth than that desire aims at the good, there is still fierce debate about how exactly to understand the former claim. So defenders of this idea, if they hope to understand desire, must first choose among the rival views about the aim of belief and then show how it applies to desire.

While I cannot go into detail about this literature here, some of the more obvious ways of understanding the claim that belief aims at truth actually illustrate, when applied to desire, the importance of the two constraints just given and the tension between them. For example, one thing we may mean in saying that belief aims at the truth is that the truth

---

<sup>35</sup> Think again of Quinn's radio man, or of Anscombe's example of someone who spreads out all the green books in his house carefully on the roof and, when asked why, responds "for no reason" (Anscombe, 1963, pp. 26-27). Tenenbaum discusses this requirement in *Appearances of the Good*, pp. 33-38.

is an explicit aim *of* individual believers. In other words, in trying to decide what to believe, each of us is guided by something like an *intention* to get at the truth.<sup>36</sup> Applied to desire, the claim would be that desire is aimed at the good in the sense that individual desirers have some basic intention of only desiring things that are good. But this seems implausible on its face and is too sophisticated to be a general account of desire; for again, animals and very small children have desires, but it is unrealistic to attribute to them such abstract general intentions.

On the other hand, we may instead interpret the claim that belief aims at truth as a claim about what makes for a successful belief. It might be, for example, to claim that it is in the nature of a belief that it is correct only if it is true.<sup>37</sup> The analogous claim about desire would be that it is the nature of a desire that it is only “correct” if what is desired is good. The notion of correctness here is a bit strange, but the more pressing worry is that this claim, on its own, doesn’t really seem to illuminate the way that desires themselves are evaluative. It tells us something important about desires, viz. how they are *to be* evaluated, but it doesn’t really help us understand them as psychological states, the way in which they themselves count as evaluative, or exactly how to understand their role in leading to action. In other words, I’m not sure whether it fully meets the second constraint of explaining how desire is really an evaluative state. To see this, consider the fact that several Humeans have claimed that this understanding of the “aim of desire” is compatible with Humeanism.<sup>38</sup>

This is not to say that I think this approach should be completely abandoned. In fact, I think that something like the claim in the previous paragraph is true. I just don’t think it is the right place to look in beginning to give an account of desire. My concern so far has just been to show that there are these two constraints on an evaluative conception of desire - that it not be too intellectualized and that it show how desire is

---

<sup>36</sup> For a somewhat similar view, see (Steglich-Petersen, 2006).

<sup>37</sup> See, *inter alia* (Boghossian, 2003).

<sup>38</sup> See (Schroeder, 2008), (Smith, 2013), and (Baker, 2014).

robustly evaluative - and that they are in a tension which makes it difficult to meet both at the same time. Let's turn now to see how these tensions are manifest in some recent accounts.

## Scanlon on Desire

Scanlon gives an analysis of what he calls desire "in the directed-attention" sense, which he thinks is at least very close to our everyday notion of desire. He wants to grant that desire plays at least some role in the genesis of action. But he also wants to hold onto the centrality of judgment about reasons, so important for his general approach. So he claims that we should think that "a desire involves having a tendency to see something as a reason" (Scanlon, 1998, p. 39).

Obviously, more needs to be said here about the *way* that, in desiring, one has a tendency to see something as a reason. Scanlon appeals to a type of emphatic persistence of a thought:

A person has a desire in the directed-attention sense that P if the thought of 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 P...Desire in the directed-attention sense characterizes an important form of variability in the motivational efficacy of reasons, but it does this by describing one way in which the thought of something as a reason can present itself rather than by identifying a motivating factor that is independent of such a thought (Scanlon, 1998, pp. 39-40).

Scanlon seems to capture something of the phenomenology of everyday desire here, and seems to fit an understanding of desire as distinct from belief into a unified account of motivation and agency. However, it is still not clear exactly how desire is evaluative until we know more about what it means to have a "tendency to see something as a reason". How is this *seeing* something as a reason related to or different from *judging* it to be a reason?

In giving a concrete example, Scanlon says that having a desire to buy a new computer involves "a tendency to judge that I have a reason to buy a computer". But how do we understand this tendency? Is it a matter of having a causal disposition to form the judgment, or more like a felt inclination or temptation to judge? Or does the thought "I have a reason to

buy a new computer” simply keep popping up in one’s mind as something to go on to consider? None of these options seems to me particularly true to the phenomenology of desire in general. They also each seem overly intellectualized after all - on the face of it, they require as a precondition of desire that one possess the concept of a reason. Could one see something *as* giving one a reason to do something, or judge that one has a reason to do something, without the concept of a reason? If not, this would rule out our ascription of desires to young children and non-human animals, which is a heavy cost to bear.<sup>39</sup>

On the other hand, while Scanlon’s account shows how desire is evaluative, I do not think it does in a way that could fit with the tripartite view. For it is part of the tripartite view that desire’s evaluative element is distinct from reason, at least to the extent that it can motivate intentional action against reason’s judgment about what it is best to do. Leaving open that possibility was one of the reasons for the move to an evaluative conception of desire. For Scanlon, desires as well as intentions are what he calls “judgment-sensitive attitudes”, meaning that they are, at least in general, responsive to the agent’s judgments about reasons and are irrational when they are not. It is because of this general dependence that we can be held responsible for such attitudes. But what happens when they come apart, as in cases of weakness of will? In such a case, the desire which leads to action is one that is precisely *not* responsive to my judgment about what I have reason to do. How then does it lead to action, and how could I have possibly resisted it? Given Scanlon’s cognitivism and focus on judgment, it seems that at the time of action I must have either really believed myself to have reason to do what I did (and thus not been subject to genuine *akrasia*), or the desire must have acted on me merely causally rather than through it’s giving me the thought of having a reason for action. In that case, it’s hard to see how my action could have been intentional or free.

---

<sup>39</sup> For an expression of similar worries, see Wallace, “Scanlon’s Contractualism”, in his (2006), especially pp. 269-270.



Another way to put the worry is this: If what desires do is something like “suggest” that we have a reason to act, where that means that they present something to our consciousness which we then may or may not actually judge to be a reason, and if they only have a role to play in our intentional actions in virtue of generally being judgment-sensitive, then how could they possibly lead to intentional action without our coming to make the relevant judgment, or even with our having made the opposite judgment? It seems we must either actually endorse them or they must motivate us to act independently of our judgments about reasons. In either case, and given Scanlon’s general picture of agency, real *akrasia* looks impossible.

In light of these worries, I don’t think Scanlon’s account adequately helps us to understand exactly how desire is evaluative in the relevant sense without being overly intellectualized. To be fair to Scanlon, however, his main concern was never to give a complete account of desire, but to say enough about it to show the difference between how desires fit into his approach to moral theory and how they fit in other approaches, especially those which are broadly Humean.

Nevertheless, his idea that desire involves “seeing” oneself as having a reason is appealing, and it is precisely this perceptual metaphor that is taken up and elaborated in the more thorough discussions of desire by Dennis Stampe and, more recently, Sergio Tenenbaum. I will now turn to these accounts, with a special focus on Tenenbaum. As will be seen, I do not think their accounts are much more successful.

## **Stampe and Tenenbaum on Desire**

With widespread recognition that the evaluative aspect of desire cannot be identified with evaluative belief, philosophers have increasingly appealed to a perceptual metaphor to understand the guise of the good approach to desire.<sup>40</sup> This connection with perception has been given its fullest recent treatment by Sergio Tenenbaum, who is himself giving further development to a very similar earlier treatment by Dennis Stampe.

---

<sup>40</sup> In addition to those discussed below, see (R. J. Wallace, 1999), (Oddie, 2005).

Both are explicit in their association of desire with a certain type of perception. Stampe, for example, says explicitly that

...desire is a kind of perception. One who wants it to be the case that  $p$  perceives something that makes it seem to that person as if it would be good were it to be the case that  $p$ , and seem so in a way characteristic of perception. To desire something is to be in a kind of perceptual state, in which that thing seems good... (Stampe, 1987, p. 359)

For Tenenbaum, the key concept is that of an *appearance*: "...the analogs of desire in the realm of theoretical reason are *appearances* [rather than beliefs]", where appearances are "glimpses of certain things from certain perspectives that present them to us as being thus and so" (Tenenbaum, 2007, p. 39).

The main benefit of this shift towards appearances is that it allows us to capture the difference between desires and evaluative beliefs and to show how we are to think about cases where they come apart or are in conflict. Tenenbaum especially calls attention to this. He lists several other ways that we use the notion of an appearance, as when we say "from far above, the car appears very small" or "the raccoon appears to be dead" or "presented this way, the argument appears to be valid, but when we formalize it, we see that it is not" (*ibid.*). Making this connection is enough to defuse what many have taken to be a strong objection to a guise of the good approach, that we often desire and do what we know to be bad, and sometimes do not stop desiring something even as we are consciously and simultaneously aware of its badness. As he points out, such conflicts are not rare in the realm of appearances. The raccoon can continue to appear dead and the argument can continue to seem valid even when we know that they are not. Similarly, even when we judge that there is nothing good about something we desire to do, it can still appear good to us *insofar* as we desire it. From the reflective perspective of judgment, it does not appear good, but from the more limited perspective of the desire it does.

Tenenbaum and Stampe also both combine this appeal to the notion of a perceptual appearance with an appeal, like that discussed above, to the idea of a "formal aim". Stampe, for example, says that there is a *modal* difference between believing that  $p$  and desiring that  $p$  - in believing  $p$ ,  $p$  is

represented as *obtaining*, whereas in desiring  $p$ ,  $p$  is represented as “a state of affairs the *obtaining of which would be good*” (p. 355). So it seems that for him “goodness” plays something of the same role in desire that “truth” plays with respect to belief. Tenenbaum makes this connection even more clear when he says:

The good is supposed to be the formal end of practical inquiry in the same way that truth is the formal end of theoretical inquiry. Thus, one can take conceiving to be good as analogous to ‘conceiving to be true’. To say that desiring is conceiving something to be good is to say that a desire represents its object, perhaps implicitly, as good...Compare this, for instance, with what can be said about imagining. If I imagine  $p$ , I do conceive, at least implicitly, that  $p$  is true (p.21).

This quotation seems to imply that for Tenenbaum the evaluative element of desire is not meant to be part of the explicit *content* of the desire. In believing, or even in imagining, that it is raining, we represent it, in a way difficult to explain, as true. But the *content* of such beliefs or imagining is fully specified by the proposition “that it is raining”. Likewise, in desiring to go to the cinema, going to the cinema is represented as good, but without goodness being part of the content of the desire.<sup>41</sup>

When we combine these two appeals, we have the beginnings of an account that may seem to meet the two constraints above. On the one hand, we have an understanding of the evaluative element as being a part of the “formal aim” of desire, rather than its content. On the other, we

---

<sup>41</sup> Actually, it’s not always clear in the book how central this understanding of the content of desire is to his view, and he occasionally says things that seem to call it into question, for example when he writes on p. 27 that “if I want something, there is a sense in which I seem to be making a judgement (or at least putting forward the content of a putative judgment): I take it that what I want has some value...” Nevertheless, he is clear in later discussions that he intended goodness not to be thought of as a part of the content of desire; e.g., in (Tenenbaum, 2008). There is a similar problem with Stampe, who despite the claim of modal difference given above, and the insistence that on his view the *content* of the desire is the desired state of affairs, also says in two places that the fact that  $p$  would be good can be seen as being part of the “perceptual content” of desire (pp. 368, 378). As I try to show here, these tensions are not incidental to their views.

have a more detailed way of thinking about the analogy between perception and desire that has some intuitive appeal and gives us a model for how to think about the ways that evaluative judgment and desire can conflict. The second shows how desire is genuinely evaluative, and the first helps us answer the worry about over-intellectualization. If evaluation is not explicit in desire, then it may not require possession of complex evaluative concepts as a condition for desiring and so will not rule out animals and young children.

Unfortunately, I don't think the two aspects of the view can be consistently joined together. We cannot think of desire as an appearance of goodness on the model of the "theoretical" appearances Tenenbaum mentions and at the same time deny that goodness is part of the content of the desire. One of the main advantages to likening a desire to a type of perception was that it allowed for an attractive way of understanding the conflict between desire and evaluative belief on the model of the conflicts that can arise between appearance and judgment more generally. To give one sort of vivid example, think of the experience of looking at the Muller-Lyer illusion. We can believe, indeed we can *know* that the lines are equal length. And yet it still appears to us that the one is longer than the other. The same thing is supposed to be made possible by this understanding of desire: "reflection upon the various appearances of the good might allow us to know that something is bad without preventing it from appearing to us, from certain perspectives, to be good" (Tenenbaum, 2007, p. 40).

The problem is that this analogy only seems apt if we reject the idea that goodness is only part of the formal aim of desire. In the Müller-Lyer case, the appearance is *that the lines are unequal*, and the belief is *that the lines are equal*. So what we reject in the belief is the same content that we have in the appearance. It appears that *p* but we believe that not *p*, where *p* is content shared between the states. Things cannot be that way in desire on the present account. The belief which is meant to conflict with desire is a belief about the goodness of what one desires to do. I believe, for example, that smoking a cigarette (which, say, I desire to do) is not good. But the content of the desire, on these views, is not that smoking a cigarette is good, but simply that I smoke a cigarette.

So desire can only be an appearance of goodness on the model of the other appearances Tenenbaum draws our attention to if in desiring we apply the concept “good” to whatever it is we are thereby inclined to do. This seems to be a rejection of the “formal aim” claim, and makes the view once again look problematically intellectualized. If we instead hold onto the formal aim claim, then we must reject any straightforward analogy between desires and theoretical appearances. But then it’s unclear how exactly to understand the claim that desire is evaluative (and how it could explain intentional action against normative judgment). When one belief rationalizes another, it does so through its *contents*. The same seems true with perception. But if we do not think that the content of a desire is in any way evaluative, then it’s hard to see how having a desire could rationalize an action in the right sort of way. For we said that an intentional action, as such, is one about which the agent saw something good.

To see this latter point more clearly, consider a recent account from Karl Schafer that sees itself as a sympathetic further development of Tenenbaum.<sup>42</sup> Schafer argues that it isn’t enough just to distinguish between cognitive and conative states, i.e., those which aim at the truth and those which (allegedly) aim at the good. Even within cognitive states, we must also distinguish between those which have what he calls “assertoric” force and those which do not. For example, in both perceiving something to be the case and in imagining it to be the case, we represent it, in some sense, as true. However, there’s a difference in *force* between the two - perceiving something somehow *really* presents it as being true in a way that imagining does not. Schafer argues that it is this difference that accounts for why we can rationally move from a perception of *p* to a belief that *p*, but not from imagining that *p* to a belief that *p*. Belief and perception have a sameness of both force and content, both of which are necessary for the rational transition between them. The same, he claims, holds true in the case of desire and intention. A desire that *p* and an

---

<sup>42</sup> (Schafer, 2013). He makes the connection with Tenenbaum explicit in footnote 26 on p. 268.

intention that  $p$  both have the same content, which is not evaluative but a description of the desired or intended action, and the same force, which he calls “imperative”. The evaluative aspect of desire is then accounted for in terms of this shared force: “any mental state that presents  $A$  with imperative force...presents  $A$ ...as something that I ought to do” (Schafer, 2013, pp. 276–7).

Note that Schafer quite clearly gets rid of the first part of Tenenbaum’s account of desire. While he still wants to draw an analogy between perception and desire, he no longer presents desire as an appearance of *the good*. Rather, he explicitly removes any implication that goodness may be a part of the content of desire and says that desire is like a perceptual appearance just insofar as it stands in the same force/content relation to intention as perception does to belief. Unfortunately, doing so makes it hard to see how his account is evaluative in the sense that we’re after. He invites us to accept the principle that a state with imperative force presents its content as something we ought to do, but this seems to get things the wrong way around from what we wanted. The notion of imperative force was simply *defined* as being that which desires and intentions have in common, such that one can make a rational transition from the one to the other. But Humean instrumentalists about desire will also agree that one can make a rational transition from a desire that  $p$  to an intention that  $p$ , so it’s hard to see whether there is anything in Schafer’s account that a Humean would not accept, and thus why we should understand him as giving an evaluative account of desire, one that can solve the problems with the Humean view discussed in the previous chapter. In other words, merely calling the force of desire “imperative” doesn’t help much, unless we can clearly say why a mere dispositionalist about desire could not also say the same thing. The whole point of the evaluative conception of desire was to explain *why* such a transition from desire to intention is rational. Schafer gives us a name for the connection between desire and intention, and perhaps illuminates something about the logic of such transitions, but he doesn’t get us much further in our understanding of what desire *is* such that it can play the roles that it apparently does.

Thus we see how the two constraints discussed help us to interpret the approach to desire of Tenenbaum, and those like his, as well as to see the ways in which they fall short. A plausible evaluative conception of desire will have to explain exactly how desire is evaluative and yet not be overly intellectualized. A view which meets one constraint well is unlikely to meet the other, and a view, like Tenenbaum's or Stampe's, which appears to meet both is unlikely to hold together.

## **Moving Forward**

The discussion in this chapter has been for the most part negative. But this is not because I think the evaluative conception of desire should be rejected. Rather, I think that some such "guise of the good" account must be correct, though none of those on offer so far have been, to my mind, satisfactory. Getting clear on the difficulties involved and the criteria for success is the first step in the right direction, and I offer the above as a constructive effort toward that end.

I will say in closing that it seems to me that an understanding of animal desire will be an important part of the proper account. The mark of an overly intellectualized account is that it does not allow for the possibility of animal desire, and yet the "guise of the good" claim is traditionally taken to apply to animals as well. So if we can understand what it means to say that animal desire aims at the good we will have an account which is both evaluative in the right way and not overly intellectualized. We would then have to say what, if anything, is different between animal and human desire. I will turn to these tasks in the next chapter.

## Chapter 4: Desire as Responsiveness to Reasons

### Introduction

In chapter two, I argued that in order to make sense of desire's role in the explanation of action, we need to think of it as in some sense evaluative, hearkening back to the ancient claim that desire is as such "aimed at the good". Then, in chapter three, I formulated two constraints, seemingly in tension with one another, on a satisfactory evaluative conception of desire, and showed how some prominent accounts have failed to meet them. The constraints were that a satisfactory account must both explain how desire is genuinely evaluative and yet not be overly intellectualized. In this chapter, I will try to develop an understanding of desire that can meet these constraints.

I said at the end of the previous chapter that a defining mark of an overly intellectualized account is that it does not allow for the possibility of animal desire and the important role that it plays in explaining animal action. So in this chapter I will begin by looking at how we should understand animal desire. At the same time, I think it is also important to be sensitive to the differences between animal desire and action, and human desire and action. The key, it seems to me, is to try to do justice to the Aristotelian claim that, although humans and other animals both have desires, for rational creatures the desiring part of the soul comes in some way to share in the rational part (*NE* 1102b13-14). And to understand *this* claim, I think it is helpful to appeal to a distinction John McDowell makes, in a different context, between responsiveness to reasons and responsiveness to reasons *as such*.

I'll begin by making some preliminary observations about animal action and briefly discussing what other philosophers have said about animal desire. I will then try to give what I see as a better account of animal desire, and I will finish by discussing what is unique about human desire, and the way that the account arrived at meets the constraints developed in the previous chapter.



## Animal Action and Animal Desire: Preliminary Thoughts

To focus our thoughts in beginning to look at animal desire, consider a relatively straightforward case of animal action. A rabbit is grazing in an open field, when it hears a crunch not far away. Glancing up, it sees the outline of a crouching coyote and as a result quickly turns and flees to safety. The rabbit has performed an act of fleeing, we might say, and we can plausibly explain its action in terms of a desire to avoid or escape the coyote.

There are a couple of other interesting things to note here. The first is that the action seems to be, in a very broad sense, intentional, if we understand an intentional action as one done for a reason. In citing the desire, we seem to be giving a reasons-explanation for the action. Put another way, it seems that there is an answer to Anscombe's "Why?" question with respect to the rabbit's fleeing, even if the rabbit itself cannot give the answer or recognize it as appropriate.

The second thing to note is that there also seems to be something in animal action like intentionality under a description. Suppose in running away, the rabbit indicated to some nearby wildlife researchers that the coyote they'd been trying to trap was nearby. It would then seem that we could describe what the rabbit did as both fleeing from danger and as indicating to the researchers the coyote's whereabouts. But we have the sense that there's an intentionality to the first description that is not there in the second. While these facts do not suffice to make the action fully intentional in the way human action often is, they show a certain continuity with intentional action, which presumably helps to explain the sense that there is genuine *action* here rather than mere movement.

So in many important ways, animal action is similar to human action, and desire seems to play a similar role in explaining it. How should we understand animal desire in order to make sense of this? The immediate thought, of course, is to appeal to the same connection that we've been grasping at between desire and goodness. There seems a particular difficulty here, though, since part of the problem of the previous ways of trying to understand this connection has been a tendency to over-

intellectualize; but there is also a hope, that in trying to understand this connection first in animals, we can avoid this mistake.

But how? In part of a discussion of our duties towards animals, Christine Korsgaard argues, following Aristotle, that what is distinctive of animals (as opposed to, say, artifacts and plants) is that they are capable of perception and voluntary motion: “Animals maintain themselves in part by forming representations or conceptions of their environment and guiding themselves around in the environment in accordance with those representations” (Korsgaard, 2004, p. 6). The representations whose primary function is to guide behavior in animals are desires, or, in Korsgaard’s Kantian language, incentives. She says that an animal acts on an incentive by way of a “primitively normative response, an automatic sense that a certain action is *called for* or made *appropriate* by the representation” (p.7).

Tamar Schapiro, in her treatment of the nature of desire picks up this same idea, which she summarizes like this, using the example of a cat (again using the more Kantian term “inclination” rather than “desire”):

How does the world look to a cat? A cat’s world, presumably, is teleologically organized around her needs and interests. The cat sees this scurrying mouse as to-be-chased, this food in the dish as to-be-eaten, and that big angry dog as to-be-avoided. When she is moved in light of her teleological consciousness, she is not simply subject to a causal disposition. The object does not force her to act in the way that pollen in the air might force her to sneeze. Rather, she sees the object as calling upon her to initiate movement in response to it (Schapiro, 2009, p. 248).

Both Schapiro and Korsgaard focus in on a way of something’s appearing as “to-be-done”, and both think of this sort of representation as involving some “primitive” sense of normativity. They both think of this incorporation of something like normativity as necessary in order to make sense of animal action as legitimately *action* rather than a mere bodily happening. So, for example, the quote above continues: “At least we have to think of her being motivated in something like this way insofar as we are committed to distinguishing between what she does and what happens to her.” Likewise, Korsgaard continues: “I say that the animal responds ‘normatively’ to the incentive, rather than merely that the incentive causes the animal’s movements, because the concept of action is

not adequately captured by the idea of a movement caused by a mental representation" (Korsgaard, 2004, p. 7).

The spirit of this latter claim is something I am clearly inclined to agree with - I made a similar argument against a straightforward dispositionalism about desire in chapter two and argued that desire, if it is to explain action, must be *in some sense* evaluative. But Schapiro and Korsgaard's positive way of understanding this sense seems to me ambiguous in a way similar to Schafer's from the previous chapter. I can only understand what it means to say that a scurrying mouse appears to a cat as "to-be-chased" in three ways: as picking out a behavioral disposition, or as a claim about the content of the thought, or as a description of the phenomenology. But none of these seems very promising. The first is clearly not what they have in mind. The second seems to be too intellectualistic. The third may very well be onto something, and in conversation Korsgaard has indicated it as closest to what she meant. But it remains mysterious and seems to me unhelpful. It is mysterious because I think there is reason to doubt whether we can say anything with any confidence about how things appear to a cat "from the inside". We may be repulsed by something in something like the same way that a cat is repulsed by something, seeing it as "to-be-avoided". But for us, such an experience is conceptually structured in many ways and I don't think there'd be anything left over which we could intelligibly discuss if we subtracted away all the distinctively human elements. It's unhelpful since it's hard to see how the presence of an ineffable quale could make the difference between real action and mere movement - all we seem to be left saying is that animals have experiences in *some relevant way* like our own in desire. It would be nice to be able to say more.

## **The Structure of Reasons, the Structure of Desire**

So the basic problem we are again presented with is this: Is there a way to understand how desire is evaluative without appealing to a mysterious phenomenology or building something evaluative into the content of the state? I think the answer is yes. The first step is to move, at least initially, from talk of goodness to talk of reasons. This is not a change of subject, since, as Korsgaard and Schapiro note, we are trying to build *something*

normative into our understanding of desire, and normativity is primarily discussed in terms of reasons these days. More importantly, nearly everyone accepts that there is some very close relationship between reasons and goodness. For example, Scanlon has famously claimed that “to call something valuable is [just] to say that it has other properties that provide reasons for behaving in certain ways with regard to it” (Scanlon, 1998, p. 96). While he wants to reduce claims about value to claims about reasons, I will soon indicate why I think the explanatory priority goes the other way around. But the important thing at this point is just that there is some essential connection between the two.

The next step is to note the inherently relational nature of reasons, and an isomorphism between this relational structure and the structure of our understanding of desire. Whenever we talk of a reason for action, there at least three things being implicated: some fact or consideration  $x$ , some possible action  $y$ , and the reason-giving relation between them.<sup>43</sup> There is also usually a fourth element, some specific agent,  $S$ . Putting them together, the general schema for a reasons-statement is:  $x$  is a reason for  $S$  to do  $y$ .<sup>44</sup>

And now note that there are several things which are required in order to fully understand a particular desire’s contribution to an action explanation. The first is that which the desire is *for*. Desires of the relevant sort are inclinations towards something, are always desires *to*.<sup>45</sup> To understand what the desire is for in this sense is to understand what is usually picked out as the *content* of a desire. The rabbit in our example has a desire *to run away*. Special emphasis is often, and rightly, given to this element of desire in recognition that desire is in some way teleological and that its primary role in the lives of animals is that of guiding behavior.

---

<sup>43</sup> Compare, from a very different perspective, (Blackburn, 2010, pp. 284–285).

<sup>44</sup> There may be a fifth implied element, the specific circumstances the agent is in.

<sup>45</sup> Anscombe discusses three ways in which we use the word “wanting” - that in which “the primitive sign...is *trying to get*”, with hopes, and with what she calls “idle wishes” (Anscombe, 1963, pp. 68–70). I’m only concerning myself with the first, which seems to me fundamental.

The second element is that which the desire is a response to, what we might call the occasion of the desire. This element helps capture the idea that desire is in some sense passive and receptive, that it is in some way perceptual or like perception. In the example above, the rabbit's desire to run away is clearly occasioned by its perception of the coyote, and we do not fully understand the desire as the desire it is unless we also acknowledge this element. The distinctness of this aspect of desire is sometimes missed by building it into the content of the desire, as when we say the desire is not just to run away, but to run away *from the coyote*. The point here is just that if we spell out the content of the desire purely behaviorally, we will be missing something. The rabbit may have a desire to run which manifests in the same physical movements either to get away from the coyote or as an urge of playfulness or because it has been trained to do so at the sound of a bell. The fact that the present desire to run involves as its precondition a perception of the coyote is essential to our understanding of its role in the action.<sup>46</sup>

The third thing we must understand to really understand a desire is something about the connection between the first two elements. In the case of the rabbit, we immediately grasp the connection. We have no trouble seeing how the perception of the coyote is connected to the rabbit's running away. The latter is *made intelligible* in terms of the former. To see how this connection can fail to be made, consider Nagel's example of the man who, when he feels the sensation of thirst, is disposed to try to put a dime into a pencil sharpener (Nagel, 1970, pp. 33–34). We have a hard time understanding this desire, indeed whether even to count it as a desire, precisely because the connection between a sensation of thirst and putting a dime into a pencil sharpener is unintelligible to us.

But what kind of intelligibility is at issue here? It seems to be precisely that of reason-giving, of counting in favor of. The presence of the coyote

---

<sup>46</sup> Again, I am focusing on desires insofar as they help to explain particular actions. We also, of course, have standing desires, like my general desire to make my wife happy. While it's true to say that I have this desire even when I am not acting on it, I only ever act on it when some aspect of my situation stands out to me as calling for some relevant action.

gives the rabbit a reason to run away, and thus we understand the rabbit's desire in a way we don't understand the man's urge to put the dime in the sharpener. We can also see now that the things required to fully understand a desire are almost exactly the same four things implicated in talk of a reason for action - some fact or consideration  $x$ , some agent  $S$  who perceives that fact, some action  $y$  the agent is thereby disposed to perform, and a relation of at least *prima facie* reason-giving between  $x$  and  $y$ .

In my view, this shows something about the nature of desire. In light of these considerations, what we should say that desire *is* is a responsiveness to reasons for action. The presence of the coyote is a reason for the rabbit to run away, and the rabbit's desire to run away just is a way of responding to this reason.

But isn't this over-intellectualizing things again? How does a rabbit become responsive to so abstract-seeming a thing as a reason without sophisticated conceptual capacities? The next two sections try to address these worries while further filling out the basic account.

## **Reasons, Goodness, and Aristotle on the Nature of Animals**

What explains how rabbits and other animals come to be responsive to reasons, and why are they responsive to some sorts of reasons and not others? I think the best way to answer these questions is to think of living things in general, and animals in particular, along broadly Aristotelian lines.<sup>47</sup>

For Aristotle, all substances, i.e., all *things*, are a combination of form and matter, where the matter is what the thing is made of, and the form is the type of thing it is. What is distinctive of *living* things is that they are in some important sense self-maintaining, that they undergo or undertake processes to keep their form in existence. For plants, this self-maintenance is limited to those things directly involving nutrition and reproduction,

---

<sup>47</sup> For other recent sympathetic appeals to Aristotle in discussing the nature of action which have influenced this discussion, see (Korsgaard, 2004) and (Boyle & Lavin, 2010).

whereas for animals it also involves capacities for perception and self-movement.<sup>48</sup>

This way of looking at living things helps us to distinguish between those things which an animal does and those which merely happen to it. Or, to use the language of reasons given above, it helps us to understand when and how it comes to be genuinely responding to a reason. If an animal (as a living thing) just is something arranged in such a way to reproduce and maintain its being as the type of thing it is, then its movements “make sense”, are intelligible to us as actions, when we can fit them into our understanding of the particular form it has. As Boyle and Lavin explain:

In recognizing a cat as pursuing a mouse or as fleeing in response to a loud noise, we regard processes in which it is presently engaged as organized by general aims that belong to it as a cat. In these sorts of instances, at least, the general idea of processes in which a certain individual figures as an agent pursuing a goal seems to get a grip only against a certain sort of background: only inasmuch as the individual in question is regarded as an instance of a certain kind of thing, a kind with a certain characteristic form or nature, a kind to which certain ends and actives belong as such (Boyle & Lavin, 2010, p. 25).

The basic idea here is that an animal’s actions are only intelligible against a background understanding of the nature of that animal. And as I mentioned above, talk of the intelligibility of actions usually implies an understanding of reasons. I’ve also said that there is some essential connection between reasons and goodness. Aristotle’s picture provides a plausible way of connecting all of these ideas.

The first step is to connect goodness with the nature of a thing. “In recognizing that a certain plant is, e.g., budding - as opposed to, say, developing a cancer - we are relating what is going on with it to a more general conception of how things go in the life of that kind of plant” (*ibid.*). In other words, in understanding what sort of thing a living creature is, we also come to form an understanding of the sorts of things that are good

---

<sup>48</sup> The account above shows why these come together - animals have desires, and desires require both perception and inclination towards movement.

or bad for it. The next step is to understand reasons for action (for those creatures capable of action) in terms of this type of goodness. To turn the claim above from Scanlon on its head, the idea here is that to say that an animal has some reason to do something is just to say that it would contribute to or constitute the realization of some good, where that good is, again, determined by the sort of thing it is. Since a living thing just is one that is disposed in ideal circumstances to realize and maintain its form, and its good is understood in terms of its form, we can understand why animals are responsive to some reasons and not others. Such responsiveness through desire is in part *what it is* to be an animal.

An example may help here, and we can return to our rabbit. The Aristotelian claim is that part of what it is to be a rabbit, *qua* animal, is to be the sort of thing which through perception and self-movement keeps itself in existence through time. It's also true that part of what it is to be a rabbit is for it to be bad when a predator like a coyote is nearby, since the threat of a predator is the threat of non-existence. Thus, when a coyote is nearby, it is good for the rabbit to escape, and since running away is a way to escape, the rabbit therefore has a reason to run away. So to be a rabbit just is to be the sort of thing which in ideal circumstances responds to the perception of the coyote by being disposed to run away from it. To do this is to respond to a reason it has, and my claim is that it does this through desire.

The previous two paragraphs provide, in outline, a general picture of the metaphysics of value and reasons that I find plausible. But accepting those metaphysical details is not necessary for the basic Aristotelian view of animals to be helpful. All that's required is the weaker idea that there is at least some connection between reasons and value, that to talk of goodness is at least to imply something about reasons for action. That would still allow us to say that an animal is, constitutively, a creature who is sensitive through perception to certain sorts of reasons for action - those which are expressive of, and tend towards the maintenance and reproduction of, its form or species. Its way of being responsive to these sorts of reasons is through desire.



## Responsiveness to Reasons and Responsiveness to Reasons *As Such*

I have just tried to say something to make it intelligible that animals genuinely respond to reasons for action; indeed, that this is part of our idea of what it is to *be* an animal. But worries no doubt remain. Most importantly, it remains somewhat unclear exactly *how* animals respond to reasons. How can they respond to a *reason* without having the sorts of abstract conceptual capacities that we possess?

The key here seems to me to be an appeal to the difference between what John McDowell calls “responsiveness to reasons” and “responsiveness to reasons *as such*”.<sup>49</sup> For McDowell, the essential condition for the latter is the ability to “step back” from one’s inclinations and representations and ask whether things really are as they appear, whether we really should do what we are inclined to do. He identifies this capacity with (one sense of) rationality, and also thinks of conceptual abilities as belonging to this capacity.

Perhaps somewhat surprisingly, since he never mentions desire, he begins a paper on “Conceptual Capacities in Perception” by talking about animal action of the same sort that I discussed above. What he says is I think worth quoting at some length, since it reiterates many of the points that I have been making:

...Animals of many kinds are capable of, for instance, fleeing. And fleeing is a response to something that is in an obvious sense a reason for it: danger, or at least what is taken to be danger. If we describe a bit of behavior as fleeing, we represent the behavior as intelligible in the light of a reason for it. But fleeing is not in general responding to a reason as such.

For that idea to be appropriate in this connection, we would need to be considering a subject who can step back from an inclination to flee, elicited from her by an apparent danger, and raise the question whether she *should* be so inclined - whether the apparent danger is, here and now, a sufficient reason for fleeing. If what an animal does flows immediately

---

<sup>49</sup> See especially “Conceptual Capacities in Perception”, ch. 9 in (McDowell, 2009).

from its natural motivational tendencies, with no room for this kind of reflection, its behavior is determined by its nature...In contrast, consider a person who steps back from an inclination to flee that comes naturally to her, and decides that the circumstance that elicits inclination is a sufficient reason for fleeing...

As I said, McDowell isn't concerned here with desire and in fact never mentions it, but the connection with what I've been saying should be obvious. I've argued that on a plausible understanding of desire, it just is how animals respond to reasons in the way he draws our attention to. But the reason an animal is responding to is not itself a part of explicit thought, at least not as such. Making the reason explicit *as a reason* is what our capacity for rationality allows us to do. I said that desire involves three elements - what the desire is a response to, what the desire is an inclination towards, and the reason-giving connection between the two. The human being who is inclined to flee, say, from a distant coyote, is thus able to do what the rabbit cannot - to ask herself in explicit thought whether the distant coyote really is a reason to flee, whether the third element is more than merely apparent, and to have this reflection play a role in determining her action.

In my view, this new ability should be thought of as involving a transformation of the capacity of desire itself. Aristotle draws a distinction between the rational and non-rational parts of the soul. Desire is found in the non-rational part of the soul and is something we share with animals, who do not have a rational part. For rational creatures like us, however, he seems to think desire itself is in a sense drawn up into the rational part: "the vegetative element [the part responsible for growth, digestion, etc.] in no way shares in a rational principle, but the appetitive and in general the desiring element in a sense shares in it" (NE 1102b30). In my view, this transformation, this coming to share in reason, can be understood as the change from desire being a responsiveness to reasons to being a responsiveness to reasons *as such*.

To see the difference, think of an analogy with perception. We can mark a difference between seeing a tree stump and seeing something *as* tree stump, or seeing a blue jay and seeing something *as* a blue jay. To do the second of each contrast requires having the *concept* of a tree stump or a

blue jay. It is plausible to say that an animal can see a tree stump or a blue jay but not see something *as* a tree stump or *as* a blue jay.<sup>50</sup> Suppose a dog is running full speed and approaches a tree stump, sees it, and jumps over it. It seems to me relatively uncontroversial to say that the dog was conscious of the tree stump, though perhaps he was not aware of it *as* a tree stump. What licenses our saying that is the way he responded to it. In the same way, I want to say that the rabbit was conscious of the danger presented by the coyote, of the reason it gave it to run, without being conscious of that reason *as* a reason. What licenses our making that claim is again the way that the rabbit responded. We pick out the rabbit's response of fleeing as an *action*, not a mere movement, and what justifies this is that the presence of the coyote was a reason for the rabbit to run, and the rabbit was responding to this very reason in running. The basic form of responsiveness to reasons is not forming an explicit thought that  $x$  is a reason to  $y$ , but precisely being disposed to  $y$  on the basis of  $x$ . This form of sensitivity to reasons is something that we have in common with other animals, but we can also form the explicit thought that something is a reason.

Now, in the human case this may seem less like a transformation of desire than an additional ability we have with respect to our desires. But McDowell argues that this ability to step back from our inclinations is implicated even on occasions when we do not step back from them. Our consciousness and our action become permeated with rationality:

Let me stress that what matters is the *capacity* to step back and assess whether putative reasons warrant action or belief. If someone actually

---

<sup>50</sup> Importantly, there seems to be something in between here that we have a hard time marking in language. An animal doesn't just see objects, but can mark out objects from the background and from each other, make connections between similar objects through time, and respond appropriately. This is what makes us want to say that the cat doesn't just see a mouse but somehow sees it "as prey". Use of "as" here, however, arguably doesn't mark out use of the concept "prey". Perhaps the additional element here just is the capacity to distinguish objects and respond to the reasons those objects give to do various things. For a relevant discussion of the difference here, cashed out in terms of a difference between "a way of perceiving something" and "a way of perceiving something *as being*", see (Ginsborg, 2006).

steps back, of course that shows she has the capacity to do so. But if the capacity is present without being exercised, we have in view someone who can respond to reasons as the reasons they are. And rationality in the sense I am explaining may be actually operative even though the capacity to step back is not being exercised. Acting for a reason, which one is responding to as such, does not require that one reflects about whether some consideration is a sufficient rational warrant for something it seems to recommend. It is enough that one could (*ibid.* p. 129).

If in acting intentionally we respond to reasons as such even when we act unreflectively, and if desire is a way of responding to reasons by being inclined to act, then it seems that all action through desire should be thought of (in rational creatures like us) as responsiveness to reasons as such. This conclusion is not the claim that desire always involves some particular evaluative thought or concept, but only that desire, like perception on McDowell's view, involves both receptivity and spontaneity, in the loosely Kantian senses of those terms. In desire, we are, like animals, affected by the world and inclined towards some act or movement, but our capacities for rationality are also active, in virtue of which these elements of our desire as well as the relationships between them are open to our rational scrutiny and assent or dissent. This is sufficient for making action out of desire rational and fully intentional even when not backed by reflective judgment on a particular occasion.

This understanding of our rational capacities and their role in desire and perception is not uncontroversial, and I cannot defend it fully here. One thing can perhaps be said in its favor by appeal to phenomenology, however. This is that, although desire does not necessarily involve any explicit evaluative thought, it often does. In desiring some ice cream, for example, one is sometimes precisely and consciously attracted to the *tastiness* of the ice cream or to its *cooling and pleasing* texture. These terms are used in such a context in an evaluative way and one can't fully explicate the desire without reference to them. The prevalence of such desires is what makes Scanlon's account, where desire involves one's attention being drawn to some purported reason for action, plausible in the first place. So sometimes the reasons we are responding to in desire, the goodness we are attracted to, is directly in view and explicit. McDowell's claim, which I find plausible, is that our rational faculties are

still in play even when the reasons to which we are responding are not explicitly thought about, insofar as they can at least be brought into focus for rational scrutiny.<sup>51</sup> This is one way in which human desire differs from animal desire.

The second major difference has to do with the possible objects of desire, where in this instance I mean by “objects” the sorts of reasons that desire can be responsive to. As I tried to argue in the previous section, the sorts of reasons that animals can be responsive to are determined by their natures. But a human being is a *rational* animal, and part of what it is to have this “form” is not simply to be able to perceive and respond to things in one’s environment, but to be able to abstract from particular experiences, form and use general concepts, and reason and deliberate about what is good in a broader and more abstract sense. This rational capacity makes available a new kind of self-movement, one which is directed not only towards those basic goods which our animal nature makes us automatically receptive to, but which is mediated by our explicit thoughts about the good. In the case of desire, this means that we can desire something *because we judge it good*, and this is a type of desire no animal can have.<sup>52</sup>

These capacities also allow us to respond more directly to reasons of a more sophisticated sort than animals are able to. Through our own explicit thought, but especially through habituation and socialization, we learn to be directly responsive to moral reasons. Aristotle’s courageous person, for example, desires directly to do the courageous thing and to avoid the

---

<sup>51</sup> I am not claiming that we never have desire, or something like it, which is not subject to our rational scrutiny. But such a state would seem to us as pathological, and I don’t think action on it could count as intentional.

<sup>52</sup> Cf. Nagel’s distinction between “motivated” and “unmotivated” reasons. It is common in the literature to say that a motivated desire is one which the agent “has for a reason”. This doesn’t seem quite right to me - on my view, motivated desires are desires which are responsive to the agent’s *judgments* about reasons, whereas unmotivated desires are those which are more directly responsive to the reasons themselves. There is a helpful discussion of this distinction (and its relation to another common distinction between desire in the substantive vs the “thin” sense) in (Schueler, 1995, ch.1).

cowardly thing. It is true that he says the virtuous act for the sake of the noble, where nobility might be thought here to be a certain way of conceiving of virtuous actions. But I don't think this means that the desire to do the virtuous thing is necessarily *mediated* by the judgment that it is noble; rather, the virtuous agent is directly responsive to the nobility - his desire is itself in part the exercise of a capacity to non-reflectively pick up on the features of an action that make it noble, and an inclination to act on that basis. Of course, such an agent will also form the belief that the action is noble, is called for, etc., but the desire, while being in line with the judgment, need not be subsequent to or based on the judgment.

Perhaps unsurprisingly, I've once again found that the general idea being developed here about distinctions between objects of desire in rational and non-rational animals has precedents in Aristotle.<sup>53</sup> Aristotle had a general term for desire, *orexis*, which covered all sorts of motivational states directed, in some broad sense, towards the good. However, he divided *orexis* into three types, distinguished by their objects. There is *epithumia*, which is directed towards pleasure, *thumos*, which is directed towards retaliation, and *boulesis*, which is directed towards a more abstract ethical type of good. For Aristotle, all animals have the first two types of desire, but only rational animals have the third. There is some debate about how exactly Aristotle thinks of *boulesis*, and the previous two paragraphs may be thought of as my brief attempt to say what I think are the distinctive objects of desire for creatures like us.<sup>54</sup>

In saying all this, I should stress that my account here should not be taken as identical to Aristotle's, though I do hope it is Aristotelian in spirit. I have appealed to Aristotle at several points where I think his discussion is helpful, and thus have focused primarily on the things that

---

<sup>53</sup> For discussion, which I rely on here, see (Pearson, 2012, esp. Ch.1).

<sup>54</sup> The debate with respect to Aristotle's view is about whether the objects of *boulesis* are plural and singular. If they are taken to be singular, they are usually taken to be connected with the agent's conception of *eudaimonia*. Then there is debate about the extent to which this conception itself should be understood as explicit and reflective. On the latter debate, see (McDowell, 1980). For general discussion, see (Pearson, 2012, ch. 6).

we have in common. There are of course differences, which it would take me too far afield to discuss in detail here. Nevertheless, I do hope to have made it clear that Aristotle's discussion is still relevant and worth taking seriously.

## Conclusion on Desire

The previous three chapters have been concerned with the nature of desire and its relation to reasons for action. I began by arguing for an evaluative conception of desire. I then illustrated that the primary problem for such an approach is being able to explain how desire is really evaluative without making it over-intellectualized. In particular, the evaluative conception faces a difficulty in explaining how animals could count as having desires.

This chapter has tried to give a positive account of desire which meets the two constraints given in the previous chapter. On the view put forward, desire is a responsiveness to perceived or otherwise represented reasons for action, where this responsiveness is understood as an inclination to act on the basis of perception of the fact which is the reason. If we accept some essential connection between reasons for action and goodness, then the view shows how desire counts as substantively evaluative. And insofar as we can draw a distinction between responsiveness to reasons and responsiveness to reasons *as such*, we can understand how animals can count as responding to reasons without possessing sophisticated conceptual capacities.

I have also tried to show how we can make sense of human desire as being continuous with, while still importantly different than, animal desire. I have argued that the difference consists in two ways that our rational capacities affect our desiring capacity - in making our desire a responsiveness to reasons *as such*, and in enlarging the set of types of reasons our desires can be responsive to. I hope thus to have captured something of what Aristotle might have meant when he said that, for creatures like us, desire comes to share in the rational part of the soul.

Now that I have spent some time talking about reason and desire, I will turn in the next chapter to discuss my conception of the will and how

it relates to the capacities discussed so far. Towards that end, I will turn my attention to the important topic of *intention*.

## Chapter 5: Intention and the Will

### Introduction

In this chapter, I will round out my initial exposition of the tripartite view by turning to what I will call the will. In order to talk clearly about the will, however, I think it best to start with a discussion of intention. To see why this is so, it is helpful to step back and remind ourselves of the big picture and place the discussion so far in its philosophical context.

The basic question I am discussing is this: Which aspects of our psychology are most fundamental in explaining our intentional actions, and how should we conceive of them and their roles and relations? Nearly all philosophers think that beliefs and desires play some such fundamental role, although there is some disagreement about how to understand these states and how they help to explain action. On one picture, which has been called by several philosophers “the standard story of action”, beliefs and desires alone are sufficient for understanding intentional action. More specifically, according to this approach, an action is intentional when it is caused in the right sort of way by a desire for some end and a belief that so acting will help to achieve that end. I have begun to tell a different sort of story about how evaluative judgment and desire can each play a role in explaining action, which I hope to complete here.

Donald Davidson gave us the canonical statement of the “standard story”, and yet he himself later came to reject it. The problem, as he saw it, was that it could not account for the phenomenon of intending for the future. And so he came to think that there must be some such thing as a



mental state of intention, without which no theory of action is complete. Many philosophers have followed him in thinking this.<sup>55</sup>

Around the same time that Davidson was making this transition in his views, other philosophers were arguing for the need to return to a substantive conception of the will as a distinctive faculty and to allow for the presence of *volitions*. While the need for volition as a distinct sort of state has not been as widely accepted as the need for intention, this view has also continued to have its adherents.<sup>56</sup>

My aim here is to take a stand on the role that these concepts should play at a big-picture level in a philosophical understanding of action. I agree with both groups of philosophers just mentioned insofar as I believe that we cannot have a full picture of human agency without referring to intentions and volitions as well as beliefs and desires. But I disagree with both, or with influential members of both, about how we ought to conceive of them and their roles and relations. In particular, I object to the idea that it is relation to a state of intention that explains why an action is intentional, on the one hand, and, on the other, that all intentional actions involve or are caused by a volition or act of will. On the tripartite view, evaluative judgment, desire, and the will are conceived as three distinct sources of motivation, each of which may (or may not) play a role in explaining intentional action and be in conflict with the others.

What follows will have five parts. In the first two, I will discuss the concept of intention as a mental state and some of the arguments that have been given for its importance in understanding action. In the third, I argue against the view that what makes an action intentional at the most fundamental level is its relationship to an intention. In the fourth, I introduce my understanding of the will, and in the fifth I spell out the

---

<sup>55</sup> Davidson's initial statement of the standard story was in "Actions, Reasons and Causes". His revised view is discussed in "Intending". Both are reprinted in (Davidson, 1980).

<sup>56</sup> See, for example, (McCann, 1974), (O'Shaughnessy, 1980), (Ginet, 1990), (Zhu, 2004), (R. Wallace, 2006). A helpful overview of how several philosophers have understood volition and its role in action is in (Audi, 1993, ch. 3).

tripartite view in more detail and show how it fits together in a satisfying way the various features of intentional agency discussed in the previous sections.

## Intention for the Future

Why think there is such a thing as *an intention*, thought of as a distinctive sort of mental state? To begin with, the idea has quite a bit of intuitive appeal. Consider the scenario with which Richard Holton begins his recent book on intention, in which you are trying to decide which color to paint your front door. You think that both a dark red or a dark blue would look nice. You can't put off the choice indefinitely and time is pressing, so you come to a decision. You will paint the door blue. It seems now that, as Holton says:

As a result of your choice you have acquired a new mental state. You still think that both colours would be nice; you still think that both are available. In addition, though, you are now in a state that does not look like either a belief or a desire. You have an *intention* to paint the door blue (Holton, 2009, p. 1, emphasis original).

So you now intend to do something in the future, and having this intention seems to involve being in a new and distinctive sort of state. In particular, what is distinctive about this state is the way in which, once you are in it, you are in some sense *committed* to a course of action. Several philosophers have tried to understand intention entirely in terms of beliefs and/or desires, but this distinctive type of commitment, and the rational constraints that come along with it, elude capture by reductive analysis. Or so it seems to me, for some of the following briefly stated reasons.

First, note that you can believe you have overwhelming reason to do something and yet not intend to do it. You can also have a predominant desire to do something, and know this fact, without intending to do it, perhaps because you see it as a temptation which you plan to resist. On the other hand, there seems to be something irrational about having contradictory intentions, but nothing irrational about having contradictory desires. And while it is irrational to have contradictory beliefs about what you have most reason to do, and so they might seem a better fit for understanding intentions, judgments about what one has reason to do will often not settle the matter, since the reasons we have to

act are often equal or incommensurable. If we see two courses of action as being equally reasonable, and we are able to choose between them, and thus to intend to do one rather than the other (as it seems we surely can), then intending could not just be a matter of forming a normative belief.

The most common and initially appealing way of trying to amend a reductive account to get around the worries above is to add in some belief condition, so that agents must not only have a predominant desire to *A* or a judgement that they ought to *A*, they must also believe that they will *A*, on the basis of that desire or judgment.<sup>57</sup> The belief condition helps to account for the rational constraint against contradictory intentions, since it is irrational to believe you will do two contradictory things. It may also be thought to help capture something more of the element of commitment involved, since believing you will do something is at least one way of being committed to your doing it.

Unfortunately for the revised account, however, believing that you will *A* in such circumstances is neither necessary nor sufficient for intending to *A*. Suppose I intend to resist some upcoming temptation, perhaps the temptation to respond with sarcasm to a tedious relative. I really do intend to resist this temptation, but I may not believe that I will be successful, perhaps because I am acutely aware of how many times I've failed in the past.<sup>58</sup> On the other hand, I may find myself really wanting to say something sarcastic and believing that I will ultimately give in, and yet it seems wrong to say that I intend to, especially if I see the desire as a temptation which I intend to resist.<sup>59</sup>

The lesson of these examples, I think, is that the commitment involved in intention is not the commitment of belief. Intention is not fundamentally a doxastic commitment about what *is* the case, like belief, but an irreducibly volitional commitment to a course of action. It is for you to be settled or decided upon doing it. Of course, the brief

---

<sup>57</sup> Wayne Davis defends just such a view with respect to desire in (Davis, 1984).

<sup>58</sup> For a similar example, see Bratman's bike example, (Bratman, 1987, pp. 38–39).

<sup>59</sup> For similar examples, see (Mele, 1997).

considerations just discussed do not amount to anything like a proof of this claim, and there are many ways that defenders of reductive accounts have tried to answer them.<sup>60</sup> But they do seem to me to provide strong enough reasons to look elsewhere, and to call into doubt the motivations for a reductive view. Beliefs and desires just don't seem, on the face of it, to do the same thing that intentions do, and it's not clear why we should try to force them to do so.<sup>61</sup>

## Intentions in Action

I have just presented, very briefly, some of the reasons that have been given for thinking that intention is a real thing, that it cannot be reduced to beliefs or desires. But how prevalent are such intentions? Are they always involved in intentional action? On the face of it, it may seem that they are not. This is because I have focused so far almost exclusively on intentions *for the future*. And this is because intention for the future plays a number of distinctive roles in virtue of which it may seem especially unlikely to be reducible to mere beliefs or desires. This is the key to Bratman's functionalist defense of intentions, and he points to unique functions they play like providing the opportunity to deliberate and commit to action ahead of time (when conditions for deliberation may be better), and facilitating intra- and interpersonal coordination on a larger scale than would otherwise be possible.

The problem with this focus, however, is that it may at the same time make it seem less likely that intentions have a role to play in all, or even most, intentional actions. After all, very many of the things that we do intentionally are simple and straightforward, requiring no significant

---

<sup>60</sup> For some attempts at reductive views, see (Davis, 1984), (Audi, 1986), (Ridge, 1998), (Sinhababu, 2013), (J. D. Velleman, 1989).

<sup>61</sup> This is the real foundation of Bratman's work on intention - if we accept a minimal functionalism, that part of what distinguishes mental states from one another is the different roles they play in our lives, then there seems to be a presumption in favor of thinking of intentions as distinctive states. I've focused here on philosophical reasons for thinking that intentions are distinct sorts of mental states, but there is arguably a case to be made based on empirical work as well - see (Holton, 2009, ch. 1).

deliberation or future planning. So there would be no real work, it seems, for intentions to play in such actions. And even if we insisted that there are what Mele calls “proximal intentions”, intentions to do things now rather than in the future, it would seem that the functional role of such intentions would have to be rather different from those played by intentions for the future, and therefore, given functionalist considerations, we should think of them as different types of states.<sup>62</sup>

But there does seem to be a continuity here which provides some plausibility to the claim that all intentional actions involve *an* intention. This continuity rests on the sense, so central to the above discussion, that to intend to do something is to be committed to doing it. I’ve already pointed out that to believe one ought to do something or to desire to do it, even very strongly, is not necessarily to be committed to doing it. Since to act is to be committed, it might seem that there is always a gap that must be filled in order to lead from any belief or desire to an action. And this might be thought to be the most primary role of intention.<sup>63</sup>

This thought is strengthened when we consider that nearly all intentional actions take some time. Consider what appears to be a very simple action: you are offered a plate of cookies, want one, and so reach out to grab one. In spite of the simplicity of this action, it is still temporally extended and things could go wrong in various ways - you could find that, due to some illusion caused by the lighting, the plate is a few inches lower than the spot you initially reached for or a few steps further away. Or you may find that, unbeknownst to you and the person offering the cookies, there is a completely transparent but narrow wall of glass between the two of you which you must walk around in order to get to the cookies. Supposing you had already embarked on getting a cookie, you would be expected in such cases to make the necessary adjustments. Unless you explicitly change your mind, there would be something puzzling about a failure to continue to take the apparent means necessary to your end. But this in turn seems to be true because, in beginning to act,

---

<sup>62</sup> For some variations on this basic worry, see (Bumpus, 2001) and (D. Velleman, 2007).

<sup>63</sup> This seems to be the primary function that intentions play in (Mele, 1992).

you moved from merely wanting a cookie to being committed to or settled on getting one. Thus, it seems that any intentional action comes with the same sorts of rational requirements and possibilities for guidance and coordination involved in intending for the future, even if in most cases of simple action, they are not usually apparent.

When these two thoughts are put together - that desires and beliefs are at least sometimes insufficient for action, and that when we do act we seem to be subject to rational constraints that cannot be explained in terms of belief or desire alone - then it becomes quite plausible to think that all intentional actions involve some distinctive state of intention. Such a postulate would explain both of these facts. It would also help us to understand something of the relationship between the concept of intention in the substantive sense - *an intention* - and that of intentional action; namely, that all intentional actions involve some distinctive mental "state" of intending.

I can't see any good argument against this position, and, at any rate, it is widespread and I will assume it in what follows. The important thing to note, however, is that accepting that there is such a thing as intention, and that we can talk of a mental state of intention wherever there is intentional action, is not yet to take a stand on the precise role that intentions play in the understanding of action, nor on the relationships between beliefs, desires, intentions, and actions. It is not, for example, to accept the theoretical claim that *what it is* for an action to be intentional is for it to be caused in the right way by a mental state of intention.

## **Intention and Agency**

So what *is* the relationship between intentions and intentional action? I have already granted that whenever someone acts intentionally, we can ascribe to them an intention, where an intention is a state of commitment that is irreducible to either belief or desire. But granting that does not fully settle any debate about what explanatory role intentions may play in intentional action.

To see this, compare the well-known debate between internalists and externalists about reasons. The "Humean", or internalist, side of the debate points out that desires, unlike beliefs, are intrinsically teleological -

they have a “world-to-mind direction of fit”, insofar as to desire that p is not to represent p as already true, but to be disposed to bring it about that p. Since actions are in a similar way teleological, are always an attempt to bring something about, the Humean argues that desires must always be involved and playing some role whenever there is an action.<sup>64</sup> And, indeed, it seems correct that whenever someone intentionally acts a certain way, there is a sense in which we can say he wanted to do what he did. So desire, just like intention, seems to be behind every intentional action. A further conclusion is then drawn from this. Since whenever we act, we necessarily have some desire to do that act, and since reasons must be the sorts of things that can motivate us, our analysis of what it is to have and to act for reasons must be in terms of desire. Reason itself is confined to helping us satisfy our desires as efficiently as possible.

On its face, the argument is appealing. One way for a “rationalist” (i.e., one who thinks there is an essential role for reason to play in motivation) to respond would be to deny the first premise, to say that the Humean is working with a bad - or at least incomplete - understanding of desire. This has been my argument in the previous three chapters. But another, more subtle response has also been given. The rationalist can grant that all motivated action depends on or involves desire in some broad sense, but can then point out that there are different kinds of desire. Some of our desires we just find ourselves having, they “strike” or “assail” us. But other of our desires we seem to have for reasons. We may come to desire to do something *because* we judge that we ought to. The Humean argument does not on its own rule this out. But in such a case, it seems that, while the desire may be a necessary condition for our being motivated, the real work, what gives the action the distinctive character that it has, is its relation to our judgment. So it seems that reason could have some important role to play in motivation after all.<sup>65</sup>

---

<sup>64</sup> For a clear presentation of this argument, see (Smith, 1987).

<sup>65</sup> This is roughly the tack taken by (Nagel, 1970) with his distinction between ‘motivated’ and ‘unmotivated’ desires, and defended at greater length in (Wallace, 2006a).

The point of this quick and obviously simplified detour into the debate about practical reason is that I think an analogous thing should be said about intention's relation to intentional action. Although it may be the case that we can always attribute an intention to someone acting intentionally, as the psychological component of the commitment involved in acting, it is a mistake to think that the having of the intention is enough on its own to fully account for an intentional action's character as involving a distinctive sort of attributability, or exercise of agency.

To begin to see why, note that sometimes forming an intention seems to be itself something that we do intentionally. I can *decide*, for example, that I am going to take this weekend off from writing and watch several movies instead. When I have made this decision, I have *done* something, and it feels this way to me, as if I've exercised my agency in some way. You can ask me why I've done it, using the sense of "why?" which asks for reasons, the sense which Anscombe singled out as distinctive of intentional action. And you can hold me responsible for the decision, blame me for it (as my advisors might do). Of course, the reasons for my intending will be the same reasons for which I (if successful) will have done the action. But you can ask the why question and blame me even before I've acted, and even if I never follow through. So the question and the blame must be directed towards the formation of the intention rather than the action. The best explanation for these facts is that coming to have the intention is itself something I've done intentionally.

According to one common understanding of the relationship between intentional action and intention, every intentional action has some intention or other standing behind it and explaining it. So if, as a result of deliberation, I come to intend not to write this weekend, then there must be some other, prior, intention which explains the intentionality of my forming this commitment. It's hard to see what that intention could be in this case, but suppose we could find one. We would then have a regress looming, since we could ask of this other intention, whether it too was formed intentionally. If the answer is yes, then there must be some other intention which helps secure the intentionality of that one, about which we could ask the same question. Obviously, the chain must come to an end at some point.



If the chain does come to an end, then at least some of our intentions will not themselves be formed intentionally. They will be things that we simply find ourselves having in certain circumstances. This possibility is recognized and even forms an important part of many views. So, for example, Mele:

Notice that if forming an intention is itself an action, it does not follow from the fact that no prior intention is *formed* in a particular case that the agent *had* no prior intention in that case. Some intentions may be *passively* acquired, as is the case with many of our beliefs, desires, thoughts, and so on (Mele, 1992, p. 184).

This seems like a reasonable move to make, but I think it raises an important problem for the sort of view under discussion.<sup>66</sup> On such a view, the intentionality of an intentional action is still secured by its relation to *an* intention. But now we can justifiably wonder about why such a relation should be so important, about how intention is able to play that role. The issue is this. An intentional action is one with respect to which we are particularly active, which can be attributed to us in a unique and distinctive sort of way. Importantly, we can be held *responsible* for it. But if the fundamental feature of an intentional action is that it stands in a certain relationship to a mental state which we may simply find ourselves with, that we “acquire passively”, then this distinctive attributability seems mysterious. How can an action’s relation to a state, with respect to which we are passive, help to explain the sense in which we are distinctively active with respect to it (i.e., the action)? Any sense of agency seems to be lost.

Let me review the dialectic of this section so far. The question we are asking is this: granting that all intentional actions involve a psychological “state” of intending, what is the explanatory connection between these two? In particular, can pointing to a relationship between intentional actions and psychological states of intention fully explain why all intentional actions, as such, are distinctive exercises of agency, involving a

---

<sup>66</sup> Though I think the move is reasonable, I don’t think the way it is stated is entirely without fault, for reasons that will become clear below. Basically, I think it is a mistake to consider all that is not acquired as a result of an action “passively acquired”.

distinctive sort of attributability? I think the answer is “no”, and the argument for this conclusion takes the form of a dilemma. If we think that all states of intention are deliberately formed by something like an act of will, then we can perhaps capture the idea that they are attributable to us in a special way. Unfortunately, we are led to an infinite regress by taking this route, since such intention-formations are themselves intentional in the relevant sense and would need further acts of will. The only way out of the regress is to say that some of our intentions are “passively” acquired. But then the fact that intentional actions involve intentions can’t do the work of explaining why such actions are distinctive exercises of agency, since these intentions themselves can seem like things that just happen to us. So we have to look elsewhere to explain at least one important feature of intentional actions, namely, our distinctive agency and responsibility with respect to them.

The best response I can see to this question is one that connects intentions to the agent’s reasons for action. At least one thing that makes us distinctive *as agents* is our ability to respond to and reflect on reasons in various ways. If an intention is something that is responsive to an agent’s reflection on, or distinctive ways of responding to, reasons, then it might seem like a fitting mental state to stand behind and help explain intentional action.

Though this answer seems to me right, it also seems to undermine the pull towards the way of understanding intention’s role that I’ve just been discussing. Though we should not want to deny that intentions are responsive to reasons in the relevant sense, they are not the only states that are so responsive. In fact, they seem to be only indirectly responsive to reasons, by being responsive to our judgments and desires. We can, for example, have standing beliefs about what matters or form judgments about what we have reason to do, and these would be responsive to reasons in at least as straightforward a way as intentions are.<sup>67</sup> Desires

---

<sup>67</sup> It may be wondered why there need be an appeal to mental states at all here - why not just say that actions are themselves directly responsive to reasons? In some sense, I think this is the right thing to say - the fundamental thing about actions is that they are explained in the right way by reasons, and it is sufficient to understand an action as such

also, at least on the understanding of them that I have been at pains to defend over the past three chapters, are ways of being responsive to reasons.

We can now return to the allusion I made earlier in this section to debates about practical reason. I pointed there to the anti-Humean suggestion that even if all action involves desire, that wouldn't mean that reason had no role to play in motivating action. For it is consistent with the ubiquity and necessity of desire that reasoning about what to do could produce a judgment that some particular action is required, and that this judgment itself could then produce the desire required for the action. We might even think that this last transition is done in compliance with a rational principle that says that one ought to desire to do what one judges one ought to do. In such a case, although desire is present, it seems like reason is actually doing all the work in motivating the action. Given the considerations above, I think we should say the same thing about intention. Although intention is always present in intentional action, it is the beliefs and desires that do all the work of explaining the distinctive type of agency such actions involve, at least in cases where there is "passive" intention acquisition. They do this work by being the two primary and direct ways that agents are responsive to reasons, since what makes an action intentional, *qua* a genuine exercise of agency, is that it is done on the basis of some perceived, or otherwise represented, reason.<sup>68</sup>

---

by knowing what counted in favor of it from the agent's point of view. In my approach, beliefs and desires are not important because they *constitute* the reasons which cause actions, but rather by being part of the psychological background that helps us understand how agents come to be able to respond to reasons at all. They can also provide further illumination about the character of the action, for example, in cases where there is conflict between the agent's judgment and desire. (Thanks to Pete Murray for pressing me to be more clear about this.)

<sup>68</sup> For further defense of this connection between activity/agency and responsiveness to reasons (though with a particular emphasis on *judgment*), see Raz (1997) and Smith (2005).

## Decision and the Will

But what about cases where we want to say that the intention acquisition is not “passive”, but is brought about by an act of the agent? I can’t see any good reason to deny that this actually happens. On the contrary, it seems to be an important way that we exercise our agency.

It is also, perhaps, the most phenomenologically salient. When we think of what it is to act intentionally, to take responsibility, to exercise our freedom, and the like, we don’t typically think of our routine and habitual actions, but of those over which we struggle, being torn in opposite directions.<sup>69</sup> In such cases we are aware of having distinct options available to us and having to choose or decide what to do. This ability to decide is different from the capacity to desire or to judge. It is a distinctive capacity which I think deserves the name *will*.<sup>70</sup>

As should by now be clear, I think there is a very close connection between the will, decision, and intention. In particular, I think we ought to say that the will is the capacity to decide, and the relevant capacity to decide is the capacity to deliberately form an intention.<sup>71</sup> Note that I have moved from speaking of “active”, rather than “passive”, acquisition of

---

<sup>69</sup> This may also be a source of distortion in our thinking about action, as it makes intentional action look far more deliberative and mindful than it often is.

<sup>70</sup> In using “will” in this sense, as a capacity to actively choose, I am following, among others, recent work by Thomas Pink (1995), R. Jay Wallace (2006), and Gary Watson (2004). Of course, the will is sometimes understood in other ways (see especially (O’Shaughnessy, 1980)). It is also sometimes used to pick out the much broader capacity to act intentionally, or the different narrow capacity to resist temptation (i.e., as in “will-power”).

<sup>71</sup> This notion of will is almost exactly the same as Pink’s: “To make a decision to act is not, like trying, to initiate bodily movement, but rather to form a persisting psychological state - a state of intention in which we are left motivated to act as decided. Decisions are forms of agency which explain intentions...” (Pink, 1995, p. 3). Though Mele’s view is more complicated, and he does not connect the will primarily with deciding as I do, he does make this connection between decision and intention: “...in *deciding* to A one forms an intention to A. Deciding is one mode of intention acquisition - an active, or intention-forming, mode.” (Mele, 1992, p. 231).

intention to “deliberate” rather than “non-deliberate” formation of intention. There is room for confusion here, so it is worth taking some time to be clear on what I mean by this.

In speaking of the will as a capacity for the “deliberate” formation of an intention, I mean that it functions in the context of *deliberation*. However, there are different types of deliberation. Here it is helpful to remember Pamela Hieronymi’s distinction, discussed in the first chapter, between attempts to answer different types of question. In particular, we distinguished there between the question of what *one ought to do* and the question of what *to do*. I disagreed with her claim that the attempt to answer the second question is the domain of practical reason while the first is merely theoretical, since any attempt to *reason* about the second question will collapse into an attempt to answer the first. Nevertheless, these are distinct questions and sometimes we settle the first without being settled on the second. This happens primarily in cases of strong conflicting desire. The will comes into play in answering the second question, and it answers it not by reasoning, for all the reasoning may have already been done, but by *choice*. To say that exercising one’s will involves the deliberate formation of an intention, then, is to say that it involves consciously settling the question of what to do. There is often a distinctive phenomenology to the making of such choices, and, as I will discuss below, the phenomenology in cases of choosing to act in line with one’s judgments over one’s desires is often one of a mental *effort*.

This way of thinking about the will is distinctive in two main ways. The first, indicated above, is with respect to where it “locates” the will. In a recent paper discussing volition, which we can think of as the “output” of the will, Jing Zhu (2004) discusses three “gaps” that arise in the process of acting that various philosophers have tried to fill by appeal to the will. First there is the gap between the reasons for action and the decision to act. Second is the gap between the decision to act and the initiation of the act. And third is the gap between the initiation of the act and its completion.

As discussed above, I think there is actually a fourth gap that is possible, between a decision about what one *ought* to do and a decision about what *to do*. My claim is that it is only to fill this fourth gap that the

will is needed. The first gap is filled by practical reasoning. The second and third can be taken care of by intention itself, so long as we are thinking of intention as the complex state of being committed to a course of action rather than a self-standing mental item that may or may not be connected with behavior.

Importantly, it is part of my view that even the gap filled by the will does not always appear. Sometimes practical reflection which concludes in a judgment about what we ought to do itself settles the question of what to do, and sometimes the question never consciously arises - we act more immediately out of desire. In these cases, the will in my narrow sense is not exercised. This is the second distinctive feature of my approach, which distinguishes it from views like that defended by Gary Watson in his important discussion in "The Work of the Will" (2004).

Watson explicitly connects the will, as I do, with decision, and though he also recognizes that we "ordinarily restrict the term to contexts in which there is prior uncertainty and an attempt to make up one's mind by deliberation", he doesn't want to restrict the exercise of the will to such cases. He gives an example involving his seeing a person in need of help and immediately dropping what he is doing to do so, "the thing to do [being] clear at once". Though this "is not strictly a case of deciding what to do", he claims it involves the same "active phenomenon that also occurs in explicit deciding, namely, adopting and forming an intention...and this is itself an instance of agency" (p.125).

The implication seems to be that all intentional action involves this prior act of agency which consists in deciding, and thus that all intentional action involves the will, at least if one grants that all intentional action involves an intention.<sup>72</sup> But this is puzzling, and for several reasons. First,

---

<sup>72</sup> Actually, whether Watson intends this implication is not entirely clear. Halfway through the next paragraph, Watson says that he is "not claiming (or denying) that all instances of intentional bodily movement are preceded by intention formation. Nor [is he] claiming (or denying) that all intention acquisition is active". I'm not sure how leaving these issues open remains an option given his example, unless he thinks it is possible to act intentionally without the formation of any intention whatsoever, an option I tried briefly to argue against above. He does say that he assumes that "there are ways of acquiring intentions besides *adopting* them", but I can't see what that would involve

it doesn't seem true to the phenomenology of agency. Although I admitted above that the most salient cases of agency involve deliberately making up one's mind, surely these are in the end only a small percentage of cases. Often we act intentionally out of habit, or simple desire, or based on a judgment about what we ought to do, without there being any gap to be filled between such judgments or desires and the action by some additional act of the will. Indeed, in Watson's own example, it seems that all that is needed to explain the action is that "the thing to do was clear at once", and any talk of "adopting and forming an intention" comes across as forced. One might be tempted to think that if actions are explained by intentions, then our actions can only be instances of agency insofar as we also exercise our agency with respect to the intentions (Watson, as I do, supposes that deciding is itself an exercise of agency). But this is the way of the regress discussed above, and Watson himself admits that focus on the status of theoretical judgment should convince us that "the boundaries between the active and the passive are not marked by the will" (p.153).

Second, thinking that the will is always exercised in intentional action also conflicts with his own understanding of "the work of the will". Watson distinguishes between "internal" and "external" accounts of the will. On an internalist account, willing is always done "under the guise of the good", whereas on an externalist account, this is not the case - agents can intentionally act in ways that go against their understanding of what it would be best to do. The will would obviously seem to play a more important role in an externalist theory, since on such a view a judgment about what I ought to do is not on its own sufficient to explain my action. I sometimes decide to go against such judgments and the need to allow for this type of, what Watson calls "counter-normative", agency is a primary draw for externalist accounts. But Watson insists that there is work for the will to do even on an internalist view, both because we can face significant uncertainty or conflict between our ends, and because the adoption of any particular end often leaves open the constitutive or instrumental means

---

given the apparent lack of any explicit "activity" of adoption in his example and his own later account of the active/passive distinction.

towards its realization. However, none of these considerations justifies thinking that the will is used in all intentional action. Sometimes reason *does* completely settle the issue of what to do for us, and sometimes there are also no doubts about the required means, or the action is very simple so that no additional planning is needed. In these cases, there seems no work for the will to do on Watson's own accounting of it. And this matches the phenomenology mentioned above.

## What does the Will have to do with Willpower?

In recent work<sup>73</sup>, Richard Holton has drawn attention to the possibility that *akrasia*, that is, action against one's better judgment, ought to be distinguished from weakness of will.<sup>74</sup> Once we think of intention as distinct from and not requiring judgment, as I have done here, there seem to be two different phenomena which we might colloquially think of as involving a failure of self-control. One is where we *judge* that we ought to act some way, but act another way instead, i.e., *akrasia* as typically conceived. But another is where we *form the intention* to act some way in the future, but then in the face of temptation act another way instead. Holton thinks that although philosophers have traditionally been focused on the former, the latter is both more common and more deserving of the label "weakness of will".

Whether or not we agree with those last two claims, there does seem to be a distinctive phenomenon here, and it raises interesting questions for the account of the will I've given so far. For Holton, willpower is a distinctive faculty which allows us to stick to our intentions *once we've already formed them*.<sup>75</sup> I have said that the will is the capacity to *form* intentions in the first place. Since in these cases the intention is already formed, it might seem that there is nothing for the will to do. And then it

---

<sup>73</sup> (Holton 2009)

<sup>74</sup> Note that so far, I have been using the terms interchangeably.

<sup>75</sup> More precisely, he thinks willpower is a capacity to resist revising certain types of intentions, which he calls resolutions, which are aimed at staying strong in the face of temptation. These details aren't important for my purposes.



will seem as if the will doesn't have much to do with willpower. But that seems odd.

Temptation, according to Holton, works primarily by way of a "judgment shift" - the more focused we are on some tempting alternative, the more likely we are to increase our weighting of the value of that alternative, and this causes us to call into question our earlier intention not to give in.<sup>76</sup> To put it into the language of decision given above, temptation leads us to re-open the question of what to do. Willpower, then, has to do with avoiding this re-opening:

It is the mental effort of maintaining one's resolutions; that is, of refusing to revise them. And my suggestion here is that one achieves this primarily by refusing to *reconsider* one's resolutions. On this picture, then, the effort involved in employing willpower is the effort involved in refusing to reconsider one's resolutions (Holton, 2009, p. 121).

How does one do this? While he points out that the best methods may be found in those, like St Ignatius of Loyola, who were professionally concerned with fighting temptation, he does point to two common strategies found in experiments on children by Walter Mischel and his colleagues (1996). Children who were successful at resisting the temptation to get a smaller reward now rather than a bigger one later tended to (a) reiterate their intentions to themselves (what Holton calls "rehearsal") and/or (b) distract themselves by focusing on other things rather than the temptation.

While this seems like a plausible understanding of temptation and willpower, it does not tell us much about the psychological capacities involved. Holton talks about willpower as a distinctive *faculty*, one which would presumably be defined as the faculty by which one effortfully refuses to reconsider one's resolutions. But this seems like an awkwardly negative way of saying what willpower involves. Instead, I think we can think of willpower precisely as the power of the will in the sense I have described it.

First, look at the distraction examples. In these cases, distracting oneself is something difficult to do precisely because our strong desires

---

<sup>76</sup> He relies here heavily on experiments by (Karniol & Miller, 1983).

lead us to focus on the value of what we are tempted to do. Shifting our focus requires an active effort, we must *decide* to do something that goes against what our desires incline us to. But this effortful capacity to decide what to do that goes beyond our judgements and desires just is one primary aspect of the will on my view.

This is even more clear in the cases of rehearsal. Holton says that rehearsal is something that falls short of reconsideration - one reminds oneself of the commitments one already has without calling them into question. But this seems to me not to capture the phenomenology of the effort involved. By the time one is face to face with temptation and must actively resist it, the volitional commitment is no longer stable. The question of what one ought to do may still be closed, as well as the question of what one *had* intended to do. But what one *will* do is precisely being called into question. So what is needed is not just a reminder of what one was previously intending to do, but a *re-intending*. If I had intended to quit smoking, then what I need to do when battling with temptation is at its heaviest is to intend, to settle for myself, that I will not smoke.

We can see, then, how willpower is related to the will. Willpower is the capacity to consciously form intentions required to act in line with one's previous intentions (or a certain subset of them) when tempting desires make it difficult to do so. Since the will just is the capacity to perform acts of conscious intention-formation, willpower is a particular type of activity of the will.

I tried to say above why I think this understanding of willpower fits the cases, as well as how it relates to an independently motivated conception of the will. An extra benefit of it is that it applies to *akrasia* as well. On Holton's understanding, and given his strong distinction between *akrasia* and weakness of will, it's hard to see how willpower could have anything to do with *akrasia* where there is no previous intention. But if willpower is just the exercise of the will against a tempting desire, then it is the solution to both *akrasia* and weakness of will; it is just that in the first case the tempting desires prevent us from forming the right intention in the first place, whereas in the second they work to undermine an intention already made. In both cases, what we

need to do is simply (though not easily) to *decide* to act in accordance with our judgment or resolution.

## Putting it All Together

I began this chapter by putting the basic question in moral psychology as: Which aspects of our psychology are most central for our understanding of intentional action, and how should we understand them, their roles, and their relations to one another? What I have called the tripartite view is my answer to that question, and I have been building up that view over the past five chapters. Now that I have said something in more detail about the will and the concept of intention, the basic exposition of the tripartite view is complete, and it will be worthwhile at this point to try to summarize how it answers that question.

At the most general level, the tripartite view is just the picture that seems to be suggested by the common-sense considerations discussed above: Sometimes we act because we desire to, sometimes we act based on our normative judgments, and sometimes our desires and judgments don't fully settle for us the question of what we will do, and so we must consciously decide, i.e., exercise our will.

This approach fits well with what was said about reasons and the work of the will above. All action, as such and as opposed to mere movement, is a motivated responsiveness to reasons. In some cases, though, like those picked out by Watson above, reflection on the reasons is not enough to settle what to do. We may judge the reasons to be equal or incommensurable, for example. And so we must choose.

But, more fundamentally, the tripartite view draws a distinction between two *ways* of responding to reasons - reason and desire. This contrast can be motivated by appeal to a plausible hypothesis about the development and primary function of the will itself. Other animals, although they are perhaps incapable of the same type of fully intentional action as adult humans, surely do *act*, and it is plausible to think that they do so by being equipped by nature to desire those things which will give them pleasure and are necessary to sustain their lives; thus, in desiring, they are responding to things that they, in at least those senses, have reason to do. Our own capacity for desire is something which we share

with the other animals. Like animals, we have desires which motivate us to respond in various ways to reasons we have to act.

But our capacity for reason is unique. Unlike the other animals, we can also deliberate explicitly about what to do and form judgments about the reasons we have to act, and these judgments can also motivate us. However, these two capacities, of reason and desire, maintain a relative independence and with this the possibility of conflict arises. We may find ourselves very strongly desiring something, and continuing to desire it, in spite of our judgment that we ought not to. The primary function of the will, on my understanding of it, is to solve these conflicts, or, more generally, to allow us to settle on a course of action when underdetermination or conflicting desire opens a gap between our deliberation and a settled intention. Given this function, however, the will is still in a way constrained by our beliefs and desires. We must always act for some reason, and when we decide what to do, the action will still be explained, in a normative sense, by the reasons made present to us through belief or desire. This is why, even when we deliberately choose which course of action to take, we don't need to mention that choice to make the action intelligible to others - we will typically just point to whichever reason(s) we ended up responding to. However, the choice itself is an additional psychological element that may have been lacking in other circumstances and which may even be ethically relevant. For example, this will often mark the difference between the virtuous and the merely continent agent: both correctly judge what they ought to do, but in one that judgment is sufficient for action, whereas in the other there is a struggle which requires the will.<sup>77</sup>

---

<sup>77</sup> Things are actually a bit more complicated here, insofar as the virtuous person will also desire to do the right thing, and so we might not think the judgment on its own is sufficient. One response would be to say that the judgment is what explains the desire, and that in general the mark of a virtuous person is that he always desires what he judges best *because* he judges it best. I think it is better to say that the virtuous agent is unified in a way the rest of us are not, so that a clear distinction cannot be drawn between his ethical judgments and his desires. For a somewhat similar view (without all of the moral psychological background of my account), see (McDowell 1975).

I said in my introduction that according to the tripartite view, there are three distinct “sources of motivation” for action, and I can now say a bit more about this by relating that claim to the earlier discussion here about intention. I said that each intentional action involves a mental state of commitment that we can call intention. The tripartite view, then, says that there are three distinct ways that we can come to have these intentions. In some cases, desire leads straight to intention; in others, judgment or belief does. In still others, the intention is deliberately formed by an act of choice. This deliberately formed intention is what I think we should call a volition. This parallels the distinction often made between belief and judgment - a judgment is a type of belief, one consciously held as a result of deliberation. In the same way, a volition is a particular type of intention, one consciously formed by an act of will.<sup>78</sup>

This concludes the basic exposition of the tripartite view. In the following chapter, I will briefly show how the view can be applied to another important topic in moral psychology, that of identification. This will allow me an opportunity both to review the basic shape of the account as well as to demonstrate its general appeal one last time.

---

<sup>78</sup> Volitions are often connected with intentions, but I’ve only seen the analogy to judgments hinted at by Wilfrid Sellars. He compares volitions to perceptual states, but then says that “Volitions are...but one variety of occurrent intention (state), as perceptual takings are but one variety of occurrent belief (state)” (Sellars 1976, p.53). I think he is right to think of volitions as a type of occurrent intention, but given their connection to deliberation, the most analogous occurrent belief state seems to me to be judgment.

## Chapter 6: Agency, Identification, and Character

### Introduction

Over the past five chapters, I have tried to explicate and defend a distinctive tripartite moral psychology. According to the approach I have been advocating, rational creatures like us have two distinct fundamental ways of responding to reasons for action: through direct desire, and through reflective judgment. Because of the conflicts that can arise from having these distinct sources of motivation, I've also argued that we need a third, the will, which allows us to decide to act a certain way when we find that the question of what we will do is not fully settled by our desires or by our practical reasoning.

In this, final, chapter, I will discuss one sort of worry that might arise for the view. The worry is that, given the motivational complexity posited, and the possibilities for psychological conflict that follow, we may lose our grip on where to locate *the agent*. What we may seem left with is a pair of clashing psychological forces and the possibility of an occasional arbitrary intervention rather than a unified agent.

I hope that my presentation of intentional action in general as a settling of the question of what to do on the basis of reasons can forestall the idea that what we are left with is anything like what Wallace (1999) calls the "hydraulic model" of action, where our actions are determined simply by vectors of force to which we are internally subjected. Still, it will be worthwhile to say a bit more about the conception of the acting self implied by the view and try to allay such worries. In short, the key is to stress that an agent's actions are always directed towards the good, and that we can talk of an agent's *character* as their more or less coherent total outlook on the good.

To speak about these issues is to be embroiled in overlapping philosophical debates about identification, alienation, autonomy, and the self. Obviously, I cannot hope to do justice to the literature that has arisen surrounding these concepts, especially in a short final chapter. What I do hope that I can do is sketch the shape that a view like mine might take in

addressing these issues, in a way that lets me both review the basic contours of the approach as well as re-emphasize what makes it appealing.

## The Problem of Identification

Although, as I mentioned above, we might understand the problem I'll be discussing as one about autonomy or the self, I find it easiest to start with talk about *identification*. The problem of identification is about how to understand the ways in which we can identify or fail to identify agents with certain of their motives and actions.

In a famous passage from his Epistle to the Romans, St Paul complains: "I do not understand what I do. For what I want to do I do not do, but what I hate I do." He goes on to say that, insofar as he does what he does not approve, it is in some sense not really him acting, but the sin within him. The seeming paradox is how his actions could both be and not be done by him. In one sense, he is the agent of all his actions and can be rightly held responsible for them. But in another, his own actions and motivations strike him as alien, as not true expressions of his self, as not fully attributable to him. I think we have all felt this way at various times. The question is how to understand these senses of identification and alienation.

To make the question more clear, I think it is important to spend some time focusing on illustrative cases. The philosopher who has arguably done the most to bring the problem of identification into the mainstream is Harry Frankfurt. Perhaps his most famous illustration involves distinguishing three possible types of drug addict.<sup>79</sup> On the one hand, we have the willing addict, the one who is addicted to his drug and likes it that way, and wouldn't have it any other way. Then we may have an addict who is a 'wanton', who has no higher order attitudes towards his addiction (or towards any of his other desires), but simply does whatever it is he desires most to do at any given moment. Finally, we have the unwilling addict who "hates his addiction and always struggles

---

<sup>79</sup> In "Freedom of the Will and the Concept of a Person", reprinted in (Frankfurt, 1988).

desperately, although to no avail, against its thrust. He tries everything that he thinks might enable him to overcome his desires for the drug. But these desires are too powerful for him to withstand, and invariably, in the end, they conquer him" (p.17). The willing addict, according to Frankfurt, identifies with his addiction (and therefore, his drug use), whereas the unwilling addict is alienated from his addiction. The wanton does neither and according to Frankfurt hardly counts as a person at all, since part of what it is to be a person is to take up second-order attitudes towards our own desires.

Or, consider another example, this time from Velleman, which I will quote at length:

Suppose that I have a long-anticipated meeting with an old friend for the purpose of resolving some minor difference; but that as we talk, his offhand comments provoke me to raise my voice in progressively sharper replies, until we part in anger. Later reflection leads me to realize that accumulated grievances had crystallized in my mind, during the weeks before our meeting, into a resolution to sever our friendship over the matter at hand, and that this resolution is what gave the hurtful edge to my remarks. In short, I may conclude that desires of mine caused a decision, which in turn caused the corresponding behavior; and I may acknowledge that these mental states were thereby exerting their normal motivational force, unabated by any strange perturbation or compulsion. But do I necessarily think that I made the decision or that I executed it?...Indeed, viewing the decision as directly motivated by my desires, and my behavior as directly governed by the decision, is precisely what leads to the thought that as my words became more shrill, it was my resentment speaking, not me.<sup>80</sup>

These two cases help us to understand in a more particularized way what it might mean to be alienated from or fail to identify with one's actions. But they are extreme in a certain way. What I mean is this: in both of these cases, the alienation is so extreme that we are not sure what to say about moral responsibility, if not inclined to outright deny it. The unwilling addict, insofar as he is said to struggle against his addiction

---

<sup>80</sup> "What Happens When Someone Acts", reprinted in (D. Velleman, 2000, p. 126).



with all of his strength, seems not to be responsible for the actions which flow from his addiction, and Frankfurt endorses this conclusion. It is less clear what to say about Velleman's case, but the more literally we take the final claim, that it is the resentment speaking rather than the agent, the more inclined we are to deny responsibility. Though identification is not the main concern in the context of the example, Velleman does seem to want to claim that his angry behavior in the example does not even count as an action in some robust sense.

Not all cases are this extreme. Ordinary cases of akrasia also seem to involve a lack of identification. We might find ourselves procrastinating, although we in some sense reject the motivation to do so. Or we may find ourselves reaching for another cookie in spite of our knowledge that it is bad for us and will make us feel bad. In these cases there is also a sense of alienation, a sense in which the agent is dissociated from his action. It is also more clear I think in these cases that the agent truly acts, and in a way that he or she can be held responsible for. Yet it might seem that to act in this way, such that one can be held responsible, depends on a notion of identification. That is, it seems that to assign moral praise or blame to an action requires the ability to attribute it to an agent in a certain unique way.

So it seems now that we have at least two different types of identification - that which is required for agency at all (or at least for robust agency, the type that one can be held responsible for), and that which is required for, we might say, wholehearted or unconflicted actions, and which is missing in akratic action. What relationship is there between these ways of speaking about identification? Are they unconnected? Several philosophers who have thought identification central to understanding agency in general have not sufficiently distinguished them, which leaves room for objections to their views. It is easy to focus on wholeheartedness as an ideal in giving an account of identification. But such an account will not be promising as a general approach to agency

since it seems unlikely that all genuinely robust actions will be wholehearted in the sense given.<sup>81</sup>

I think it is more promising to think that there is just one type of identification at issue, but one which can come in various degrees. This accounts for our different intuitions about the cases above. In the case of ordinary akrasia, we are confident that the agent is responsible, and though there is a clear sense in which he is not identified with his action (insofar as he thinks he ought not do it), there is another sense in which he is identified with his action (insofar as he does it freely). In the case of Frankfurt's unwilling addict, we are more likely to think he is purely dissociated from his actions, that he cannot be held responsible for them, but should rather be pitied. Velleman's case seems to fall somewhere in between and, depending on the details of the case, and which details we think most important, our judgments about responsibility could go either way. In general, I am inclined to think the person in a case like that which Velleman describes is responsible for his action, and does identify with it to some extent. Thus, there seems to be a close connection between identification and moral responsibility, and both seem capable of coming in degrees. A satisfactory view should be able to account for this, and I will return to this issue in a later section.

There is one further aspect of identification that I think has been left unclear in many discussions, but which will be important in evaluating different theories. In most cases, we talk about identification as something the agent does or fails to do. But we also tend to connect identification with the notion of the self, as when an agent says about an action from which he is alienated that it doesn't represent who he *really* is. This latter use seems to be of a concept that we can discuss from a more objective, third-personal standpoint. In other words, it seems like the sort of thing an agent could be wrong about. Frankfurt says that in identifying with an action, the agent *takes responsibility* for it. But on the face of it, it seems it should be possible for an agent to be responsible without *taking*

---

<sup>81</sup> For discussion of such views and arguments against them along these lines, see e.g., (Arpaly, 2003) and (C. Miller, 2013).

responsibility. And it should be possible for an action to represent who an agent really is even if they would deny that fact. This will be made more clear in the cases presented as counterexamples to the normative judgment view of identification, but it is worth keeping in mind from now.

Now that I've tried to make it a bit more clear what the problem of identification is supposed to be, as well as what we should be looking for in a satisfactory account, it's time to turn to a couple representative attempts. As I said above, the goal here is not to be exhaustive, but to give a sense of the problems involved and how the tripartite view I've defended may help to solve them.

## **Frankfurt's Desire-Based View**

Frankfurt first began discussing identification in his "Freedom of the Will and the Concept of a Person", and in that and several subsequent papers he explains identification in terms of second-order desires, or, more specifically, second-order volitions. Humans, like other animals, are subject to desires which can motivate us to act in various ways - these are first-order desires. When one of these first-order desires actually does motivate us, then such a desire is, according to Frankfurt, identical with our will. However, unlike other animals (and this is essential to our personhood), we are also able to "step back from" and reflect on our first-order desires. We are able to form second-order desires, desires about our own desires. One type of second-order desire is a desire simply to have another desire. Another type of second-order desire is a desire that some other desire be effective in causing action; that is, that it be one's will. Frankfurt calls such a desire a second-order volition.

It is these second-order volitions which Frankfurt thinks essential to identification (and personhood and free will). Initially, Frankfurt claimed that forming a second-order volition is sufficient for identification, that forming such a volition is to take a decisive stand on one's first-order desires. The view applied to actions, then, is this: An agent S identifies with his action A iff A is motivated by a first-order desire about which S has a second-order desire that it be effective.

This is a natural and attractive account. It does a good job of capturing the fact that identification is a problem that is unique to reflective creatures like us. It ties together several important and related concepts such as reflection, identification, personhood, freedom of the will. And it captures an intuitive distinction between first- and second-order desires, and why such a distinction might be important for the issue of identification: whereas our first-order desires are things that in some sense happen to us, which arrive unbidden and are beyond our direct control, second-order desires, precisely insofar as they involve the reflectivity that is partly definitive of personhood, seem to be good candidates for being definitive of the stance of the person.

But there are serious problems as well. They all revolve around a sense of arbitrariness that comes from identifying the agent's fundamental standpoint with his second-order desires. One way of seeing this arbitrariness is to note that the same ability which allows persons to form second-order desires also allows them to form third- and fourth-order desires. Why should the second order be definitive? Suppose Bob desires a drink of alcohol, but then forms a second-order desire that it not be effective in motivating him to get a drink. Now suppose he forms a third order desire that he stop having this second order desire and that it not inhibit his first order desire. It certainly isn't obvious in a case like this, which seems perfectly possible, that we should identify the agent with his second-order desire rather than his third (or his first).

This is more clear if we build into the case that the desire for a drink of alcohol is innocent (i.e., that it is not a desire which might be otherwise morally problematic, such as part of a desire to get drunk, etc.) and that the second-order desire arises because of his strict religious upbringing by teetotaling Methodists, a religious upbringing which he consciously rejects. So described, we have reasons for rejecting any identification of Bob with his second-order desire. His second-order desire doesn't represent how he really feels, it is just a leftover influence of a worldview he thought he had moved on from. We get this result partly because it seems like second-order desires can also be the sorts of things over which we don't have control. Again, it seems like merely being at a certain level

in a potentially infinite hierarchy will not be able to capture what is important about identification.

In response to this sort of worry, Frankfurt slightly amended his view. Though second-order volitions are still central, he now also appeals to a notion of satisfaction. Satisfaction is thought of as a certain type of inner harmony. In the example above of the man with a tee-totaling upbringing who wants a drink, it is plausible to say that, though he forms a second-order desire, it is not one with which he is satisfied, insofar as he rejects the worldview which is responsible for that desire and thus has further conflicting attitudes towards it. This satisfaction, however, need not be actively imposed: “[b]eing genuinely satisfied is not a matter of choosing to leave things as they are or of making some judgment or decision concerning the desirability of change. It is a matter of simply having no interest in making changes” (Frankfurt, 1992, p. 13). This condition is not met in the example because Bob is unhappy with his second-order desire and has a disposition to change it.

But satisfaction thought of in this way, as simply a lack of any disposition to change, seems no more likely to work. Frankfurt elsewhere admits that such harmony could be brought about by exhaustion. But as Laura Ekstrom asks, “Why think that A is what I “really want” if I am only satisfied with the desire for A because I am too tired to think any harder about the matter?” (Ekstrom, 2005, p. 50 n.9) We might even think that such satisfaction could be akratic in a certain sense. Frankfurt insists that satisfaction not be thought of as a sort of normative endorsement. But that makes it seem possible that an agent could, in a fit of frustration, ‘accept’ some second-order desire all the while believing he ought not. This seems to be the opposite of identification.

The concept of satisfaction does seem important, but it doesn’t seem to be best analyzed in Frankfurt’s dispositional sense. In most cases where an agent is not satisfied with some choice or desire, it seems that this is mainly because, as in the example of Bob, there is some decisive conflicting normative consideration. But if normative judgments play such an important role, why not privilege them when it comes to identification? At the very least, one could object that the addition of a requirement of satisfaction to the account is *ad hoc*, especially given the apparent

importance of normative considerations in the most common cases of dissatisfaction. I worry, in fact, whether, properly understood, satisfaction is just another name for much of what we want to capture in an account of identification.<sup>82</sup>

## Judgment-Based Views

A judgment-based approach to identification claims that an agent only identifies with those actions which he judges he has reason to do. Such a view seems on its seems quite attractive. Like Frankfurt's hierarchical approach, it captures the fact that identification is only a problem for rational creatures like us, who can reflect on their desires and reasons for action. Normative judgments seem to be capable of playing a role in the formation of the cross-temporal connections which on a popular view partly constitute personal identity through time.<sup>83</sup> Furthermore, it doesn't give the wrong answer in the case of Bob, but rather captures why Bob doesn't identify with his tee-totaling tendencies. On a more basic level, it seems to account very well for the failure of identification that arises in

---

<sup>82</sup> For the sake of space and structure of presentation, I must skip over Bratman's account of identification, worked out over several papers included in (Bratman, 2007), though I will return to Bratman in a different context below. Like Frankfurt, Bratman's account is hierarchical, but rather than focusing on desires he focuses on the sorts of planning states like intention that he has spent so much time developing. He summarizes his view like this: "For an agent to identify with a certain first-order desire is, in a basic case, for that agent to have a self-governing policy...that says to treat that desire and/or what it is for as a justifying consideration in motivationally effective practical reasoning. Such identification involves a higher-order policy that is not merely about one's motivation but is also about one's practical reasoning" (Bratman, 2007, p. 6). In short, I think this approach cannot account for the example of Bob the alienated teetotaler above. Bratman faces a dilemma in his view: either the policy he mentions must be explicitly formed or not. If not, then it seems to me we can say that Bob has a policy to discount his desires for drink, but he does not identify with this policy. If the policy must be explicitly formed, then the view seem overly intellectualized, and we will turn out very rarely to identify with our desires.

<sup>83</sup> A main theme in Bratman's appeal to cross-temporal planning states like intention. See previous footnote.

cases of weakness of will - on a common-sense interpretation, these cases simply consist in acting against one's judgment about what one has reason to do.

But there are also reasons to think that judgments about what we have reason to do have a unique authority, a unique claim to represent the agent's take on things. This is because it is at least partly our ability to ask questions about what we have reason to do which gives rise to the possibility of alienation or identification with our motives, and normative judgments just are the answers to these questions. As R. Jay Wallace forcefully puts it,

the question we put to ourselves when we step back in thought from our attitudes is not in the first instance about those attitudes themselves, but about their propositional objects. Wanting another drink, we are in a state that disposes us to think that the immediate pleasures of wine or whiskey would be good reasons for having some now. When we reflect on the desire, we suspend our provisional commitment to this normative claim, and put it up for assessment. This is the situation that raises the question of whether or not we are identified with the desire that we reflect on. An attitude will have authority to decide this question, then, just in case it is suited by its nature to resolve the state of suspended commitment that reflection has induced. But normative beliefs or judgments uniquely satisfy this condition...Their authority to speak for us stems from the fact that they represent our own answers to the very question that originally generates the problem of identification<sup>84</sup>.

This view is promising, and it captures something of importance that is left out of any view which focuses solely on desire. An agent's normative judgments do have a role in identification and cannot be ignored. But neither, I argue, can they be overemphasized or thought uniquely sufficient for identification. If they are, then the view once again becomes subject to counterexamples. A particularly interesting example is that which I appealed to in the opening chapter: Huckleberry Finn from the famous Mark Twain story.<sup>85</sup> At a crucial moment in the story, Huck has

---

<sup>84</sup> R. J. Wallace (2014), p. 124

<sup>85</sup> This example is one of the centerpieces of argument in (Arpaly, 2003).

the chance to turn in his companion, Jim, who is a runaway slave. His best judgment is that he should. He thinks long and hard about it and is sure that helping Jim escape is morally wrong, in addition to being against the law, and that doing so will ensure that he goes to Hell. Nevertheless, he cannot get himself to turn Jim in. Instead, he decides to act against morality (he never explicitly rescinds his judgment) and instead to help Jim reach freedom.

The incident seems to illustrate an interesting and in some ways typical case of *akrasia*, insofar as Huck freely acts against his better judgment. We have been treating *akrasia* as something which involves a failure of identification, and the normative judgment view of identification would tell us to identify Huck more strongly with his judgment than with his actions. As Arpaly points out, this fits in with Kant's views about moral worth. Although Huck ends up doing the thing we think is right, since his judgment was faulty, insofar as he acts against his principles and according rather to his feelings, Kant would claim that Huck's actions lack any true moral worth.

But this is the opposite of the conclusion that Twain wants us to draw, and that readers do in fact draw.<sup>86</sup> On the contrary, we think that Huck is morally praiseworthy for his action, that he is in many ways better than his peers and relatives despite sharing their explicit moral beliefs. Rather than identify him with the racist views he professes, we identify him with a deep moral perceptiveness that we attribute to him, even if it is only present for him at the level of emotions and desires. This seems to be a clear counterexample to the normative judgment view.<sup>87</sup>

---

<sup>86</sup> For more argument on this point, see (Arpaly, 2003, pp. 9–10).

<sup>87</sup> It also seems to be a counterexample to the previous two views. Both of those views, insofar as they are hierarchical, require some sort of higher level endorsement of desires. It is not at all clear that in acting Huck must have either a second-order desire about his desire to free Jim (why can't he simply act on the first order desire?) or that he must institute some sort of general policy of helping to free slaves (or any more narrow policy).



## A Character-Based Approach

The conclusion of the two previous sections seems to be this. Neither judgments nor desires, even hierarchically arranged desires, are necessary to establish identification in any given case. Sometimes, perhaps most often, we identify an agent with their judgments over their desires. But other times, notably in the Huck Finn sort of case, we identify an agent with their desires over their judgments. This makes sense, given the moral psychology developed earlier in the dissertation, since I've argued that judgment and desire are two distinct ways that agents can respond to reasons. But it leaves open the question about identification. What makes it the case that identification with one rather than the other is called for in any particular case? My suggestion here is the following, which I hope has the ring of common sense about it: an agent is identified with his or her action (or desire) to the extent that it expresses or is consistent with his or her character.

For this to make sense, I need to say something a little bit more about character. Basically, a person's character can be thought of as his overall conception of the good. Given the moral psychology I've been developing, which sees judgment and desires as distinctive ways of responding to value, a straightforward way of thinking of character, then, is as being made up of our background desires and normative judgments and our dispositions to form and act on such desires or judgments. So, for example, someone who is very honest, for whom honesty is a character trait, will tend to judge that telling the truth is important, to desire to tell the truth in most or all circumstances, and to actually tell the truth most or all of the time.

One problem for the view seems to arise rather quickly: people rarely, if ever, have fully settled characters. We are all complexes of contradicting desires and judgments, influenced in so many ways by our environments and upbringings. But this seems to imply that we will never be able to give clear answers to questions about identification. In one sense, I think this worry is misguided. In another, it points to something which is a positive aspect of the view, something that counts in its favor, and not a negative one.

It is misguided because I think that in common speech, claims about identification primarily arise in cases where someone is acting against character; that is, when we suppose that they do have a reasonably settled character, but a character which happens not to be properly expressed in a particular instance. We say, for example, "That is so very unlike him" or "You mustn't judge Johnny by his actions today, he is beside himself" or "I'm sorry, I don't know what came over me". Another type of common case is when we want to single out an action as being particularly illustrative of the character or personality of the agent, as when we say "Oh, that is so Mary" or "What did you expect, when you're working with Stuart?" The initial philosophical task, then, is not to be able to give a verdict about every case, but to specify what it is we are saying when we say that someone is or is not identified with a desire or action. And often, when we say those things, the context is one where there *is* a generally settled character, at least settled enough for a context of contrast.

On the other hand, it is true that our characters are not usually fully consistent or stable, and that this will often make it difficult to decide clearly one way or the other whether an agent is to be fully identified with his action. But this is as it should be, and the fact that the character view predicts this is something that counts strongly in its favor. The instability of our characters makes the most sense of the fact, mentioned in the beginning of the chapter, that identification is something that apparently comes in degrees. The typical akratic is dissociated or alienated from his actions insofar as he sees himself as acting against what he has most reason to do. But we do tend to identify him with his actions to the extent that we say that they are free, that he can be held responsible for them. The character view explains this: a person's character is his overall conception of the good, as this is expressed in both his tendencies to judge and desire. The akratic person (in the most typical case) has conflicting normative beliefs and desires, and this explains why he can be both identified and not identified with his action.

Still it should be possible to say more than simply that a person with a less-than-fully-formed character both is and is not identified with his actions. For example, in the case of Huck Finn, insofar as Huck exhibits a type of weakness of will, he shows that his character is not fully formed

and has conflicting elements. But in his case, we go beyond saying merely this to say that his action of freeing Jim shows who he really is, that it better expresses his character than his abhorrent explicit moral beliefs. How are we to understand these claims?

I think the right way to do so is to think of a person's character as a sort of web, in a way analogous to that in which coherentists in epistemology commonly speak of our body of beliefs being a type of web.<sup>88</sup> The key point here is not about epistemic justification but about interconnectedness and about location with respect to the 'center'. When we identify Huck with his praiseworthy action, we are not denying that his racist beliefs are genuinely a part of his character, of who he is as a person. We are rather claiming that they are not central; that, though they have been with him his whole life, on account of his upbringing, they now stand near the periphery of his web of values and value-dispositions. We are expressing our confidence they are on the way out, becoming less pronounced, and that they play a less important and prominent role in motivating his actions.

Of course, this is all just metaphor, and it only serves to draw the attention more clearly to the question of what makes some judgment or desire more central. This is a difficult question. Two things initially seem clear: what is most central to a person's character can only be seen over time, and actions (over time) are privileged criteria. For example, in a typical case of *akrasia*, I think that we identify the agent with his judgment rather than his weak-willed action because we think that this is an isolated case. But the less the case truly seems unique, the less likely we will be to identify the agent with the judgment rather than the action. For example, if I am constantly talking of how important I think moderation is in food and drink and then you see me drunk and stuffing myself with food, you might believe me when I say that my beliefs really represent my views, but that I was simply weak-willed today, perhaps because of a difficult week. But after seeing me in the same state several more times and finding from someone that it is a nearly nightly occurrence, you will

---

<sup>88</sup> See, e.g., (Quine & Ulian, 1970).

be more likely to conclude I am actually a glutton who is either being dishonest to others or to myself. Even if my beliefs are genuine, you will be unlikely to simply identify me with them at that point. Instead, you will simply think of my character as tragically corrupted, drawn in two different directions.

Another distinguishing feature will be overall coherence with the other beliefs and desires that make up the person's character. In some sense, this will be related to the point above, since failures of coherence will be most clearly on display in a mismatch between the agent's judgments and desires. But there may also be mismatches that are detectable within the agent's value system even at the level of explicit belief. This is obvious in a case like that of Bob the former teetotaling Methodist above. Knowing that he completely rejects nearly all of the distinctive values of his upbringing, he still finds himself feeling guilty or apprehensive at the thought of drinking alcohol. Even if he found himself tending to judge on these occasions that he should not drink, I think we would not identify him with such judgments, as he himself does not in later moments of cool reflection. And this is due to the way that such a judgment fails to fit in with his other judgments about what is valuable and with his general rejection of the religious views that provided their original foundation.

Both of these considerations seem to play a role in our judgment about Huck Finn. In the first case, as we say, "actions speak louder than words", and though Huck claims racist beliefs, his action of helping Jim escape tells against those views. I said above that a single action is not enough, at least for attributions involving more than mere intentionality. So I think it is important that in the novel we are not only presented with that decisive action, but a whole history of the relationship between the two which shows a continuous and growing, if still implicit, sensitivity to the humanity and moral status of Jim by Huck. In addition, I think it is not uncommon in our day to think that the sort of racist beliefs common in 18th and 19th century America (and beyond) could only have been maintained through a certain sort of self-deception and inconsistency. True, for some, sufficiently nasty, folks, we judge that they were consistently racist and identify them with such beliefs without caveat. But for many others, like Huck, we tend to judge that the racist beliefs were an

unfortunate result of upbringing and culture that fit uneasily with the rest of their moral outlooks. I will not try to argue that we are correct in such judgments. But the point is that such an argument, if it could be made, would have to be made using precisely the sorts of considerations given (i.e., consistency with other value judgments and actions throughout time) and which seem to me to imply an understanding of identification that fits most comfortably within the character view.

Now, in the above discussion I have almost fallen back to speaking of identification as if it were an all-or-nothing thing, something I denied earlier in the chapter. So a bit more nuance and complexity must be added. As I said above, I think there are different levels of identification, where perhaps the lowest level is that required for so much as seeing an action as intentional, and the highest is that required for *wholeheartedness*. Huck's case, as I've just explained it, seems to involve something in between. He is conflicted, and less than fully virtuous, but we perhaps identify him more with his actions and desires with respect to Jim than his explicit judgments. How does the character view make sense of these differences?

At the lowest level, all that is needed for bare intentionality, or for being held responsible, is that one acts for a reason. Although in some sense such actions may be "out of character", in another sense all actions, insofar as they intentional, demonstrate something important about one's character. Take a rare akratic action, for example: Jonathan, a normally abstemious person, gives in to peer pressure and drinks too much at a party. Such an action may be out of character, for example, insofar as Jonathan usually resists such temptations. Still, in being tempted, he saw some consideration or other as being a reason to act as he did, and he acted on this reason. A person with a different character may not have seen himself as having any reasons to drink too much, or would not have experienced such reasons as a temptation. So there is still a sense in which his action counts as an exemplification of his character, and this is why I think we can identify him with his action, at least at the most minimal level needed to call it intentional and hold him responsible. If over time Jonathan began to drink more and more at parties, and his views about the proper role of alcohol in social events began to change, we might

identify him with his future bouts of over-drinking more strongly. In one sense, this would just be because his actions on such occasions would be less “out of character” in the sense that they would involve him in less internal conflict. But there is also an additional, temporal, element involved, like that I appealed to above in discussing Huck Finn. We can become confident that his character is developing into one where such drinking is not at all out of place, and identify him with the “direction” his character is going. In most cases, as I said, character is not fully formed, and so context will always have to affect our judgments of identification. Huck Finn’s actions stand out as especially admirable given his social context and his own upbringing. At the same time, he is not fully virtuous, and his character would be even better if he did not also judge that Jim is rightfully someone else’s property.

I hope to have shown, then, that the character view fits with common sense, illuminates some aspects of our discourse about identification left out by other views, and doesn’t seem subject to the same counterexamples. Insofar as it depends on a moral psychology that recognizes the importance of normative conceptions in general and judgments in particular, it shares some of considerations in favor of the judgement-based views. Yet, like the desire-based and policy-based views, it appeals to something, an agent’s character, that is hierarchically structured. Further, insofar as our concept of a person’s character is intimately connected with the concept of a unity through time, it seems to capture the main advantage that Bratman sees in his own view,<sup>89</sup> that is, the connection between identification and personal identity, and the cross-temporal bonds that help to constitute the latter. To see something as flowing from an agent’s character is precisely to see it both as expressing his conception of what reasons there are to act, and to see it as being related to the development through time of the agent’s conception of the good or what he has reason to do. For all of these reasons, the character view seems to me the best account of identification.

---

<sup>89</sup> Not fully discussed here. See footnote above.

## The Need for a Distinctive Concept of Choice

So far, my discussion has focused on the relative roles of normative judgment and desire in identification. It might have come as a surprise, given the moral psychology I've been at pains to develop in previous chapters, that my account of identification came without discussion of the will. Is there a distinctive role for the will to play in an account of identification? Now unsurprisingly, my answer is "yes". But as with the explanation of action more generally, it is a secondary and subsidiary role.

As with action more generally, the will is needed to solve a certain sort of problem that can arise for agents like us. We can call this problem the problem of underdetermination. Bratman appeals to his self-governing policies to solve just this sort of problem, though he specifically singles out what he calls "underdetermination (of the contours of one's life) by value judgment".<sup>90</sup> The problem can perhaps be put most clearly this way: Part of a mature perspective on things involves the recognition that there are many types of goods and many types of equally good lives. Yet the question of what type of life to live must be decided. For example, one can recognize that becoming an artist and becoming a teacher are both valuable, and incommensurably so, and yet still decide to be an artist. But such a decision cannot, by hypothesis, be made with respect to one's evaluations, since both are conceived of as equally (or incommensurably) valuable, and so one's identification with them must be in virtue of something else, like Bratman's higher-order intention, which I would understand as involving at least an initial "act of will".

This is not exactly the same as the problem that I initially introduced the will to solve earlier in this dissertation, namely the problem of conflict between normative judgment and desire. I'll say more about this below, but for now, I think we can see how solving both of these problems can be conceived of as the work of the will. Both involve a gap that arises between completed practical reasoning and an intention to act some particular way, and both require a capacity to be motivated that is distinct

---

<sup>90</sup> See "Autonomy and Hierarchy" in his (Bratman, 2007).

from either one's desires or practical judgments. This capacity is the capacity to *decide*.

Now, there is room for debate about how pervasive these cases of underdetermination by value judgment that Bratman discusses are. It may also be plausible to say that in most such cases, what happens is that, although one judges that being a teacher and being an artist are *in general* equally or incommensurably valuable, one can still consistently judge that being an artist is most valuable *for me*, i.e., given my talents and inclinations. Such an appeal to talents and inclinations fits in quite well with my view, since the appeal seems to be precisely to part of what might be colloquially called one's character. Furthermore, while without this caveat these cases seem to present a serious objection to a judgment-based view according to which an agent is only identified with those actions she judges she has *most* reason to perform, it doesn't seem quite so applicable to the character-view, according to which one need only conceive of one's action as good in some way, without having to judge that it is most good.

But the worry still applies, at least in some cases - those involving what I'll call existential commitments. Think, for example, of Sartre's famous example of a man who is torn between going to fight for the French resistance movement or staying home to care for his ailing mother. He thinks that each of these is of the utmost importance, that he must decide to do one or the other in order to live meaningfully, and that he must decide wholeheartedly and totally commit himself to whichever he chooses. The issue here is not just one of how action at all is possible given underdetermination with respect to value judgment, though that is important, but rather that, whatever decision is made, the man will be totally identified with it. The choice made is not referred to something antecedent, like character or judgment, to see whether it represents the man's 'true self' - rather, the choice itself will in part constitute what the self is and will become. The level of identification with the decision will, in other words, go far beyond what would be predicted on the basis of the character view by itself.

All we need to understand such cases is that the intention that is the result of the agent's choice involves a certain content. The decision one makes to pursue a career in art, say, or join the French Resistance, if it is to



be an existential commitment and support a strong claim of identification, cannot be just a bare intention to act a certain way on some one particular occasion. It must involve something more general and far-reaching. Bratman says that such a commitment involves an intention to treat certain sorts of considerations as reason-giving. Though this seems right, I think we may just as accurately speak in a less specialized way by saying that such a choice involves in some sense constituting one's identity. This allows us to relate the discussion back to character, since one's identity seem on the face of it to be in large part constituted by one's character, that is, one's tendencies and dispositions to act or value. Thus, on my view, one's choice to become an artist just is the formation of an intention to act in the ways and develop the dispositions that are characteristic of an artist, to make those characteristics part of one's character. The connection of this intention with character shows its connection with the account of identification developed in the previous sections.

But really this is not so different from what all choices (in the strong sense of acts of will) involve. In every case where the will is needed, on my view, there is at least some level of underdetermination. In first arguing for the need for the tripartite view, I focused on a conflict between reason and desire. Such cases involve what we can call an underdetermination of character. As I mentioned in the previous section, none of us has fully determined characters, and so, unlike the virtuous person, we frequently find ourselves being drawn towards actions that we judge are bad. In these mundane cases, the point of the will is to allow us the possibility of exhibiting strength of will, of not just being driven by our desires or being lucky if our judgment is strong enough to carry the day. But the mundaneness of these cases can obscure an important fact - that in exhibiting or failing to exhibit strength of will in them, we are also responsible for shaping our character going forward. Each time we fail to resist what we see as a temptation, it becomes harder to do so in the future (and the likelihood is increased that in the future we will fail even to see it as a temptation); and the only way we can change our negative habits is by continually resisting them when the necessity arises (though of course better still is to try to avoid them entirely).

Thus the primary role of the will with respect to identification on my view is that it allows us some control over our character, over the moral self that we become.<sup>91</sup> This role is subsidiary to our desires and judgments because in many cases our characters, as partly constituted by tendencies to those desires and judgments, will already settle for us the question of what to do in any given case. Our wills do not allow us to create ourselves entirely anew, all at once, and our desires and values depend to a large extent on the initial upbringing that was almost entirely beyond our control. Rather, when we come to those cases where we are internally divided or where the way forward is unclear, our wills allow us to make the choices that can play a role in bringing about those small changes in disposition which over time can have a drastic effect on the types of people we become. With the possible exception of those existential choices mentioned above, we stand to our characters like those in Neurath's boat:

We are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support. In this way, by using the old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction.<sup>92</sup>

## Conclusion

The goal in this chapter, as mentioned above, has not been to give the last word about identification and related issues. But I hope to have shown that the tripartite psychology I've been defending has something plausible to say about the problem and something to contribute to the

---

<sup>91</sup> Though I cannot fully develop the suggestion here, it seems to me that the will might also have a role in explaining why we sometimes fail to hold addicts responsible or identify them with their actions, even though such actions seem in one sense fully "in character". The problem is that they have lost (to a significant extent) the use of their wills with respect to their addictive actions. Not only do they often have conflicting judgments about the worthwhileness of their addictive actions, they have an abnormally difficult time resisting temptation and thus exercising control over their characters over time with respect to the addiction.

<sup>92</sup> Quoted in (Cartwright, Cat, Fleck, & Uebel, 1996, p. 191).

debate. Indeed, I hope that what initially seemed to raise a worry for the view can now be seen to count in its favor. Not only can the tripartite view give an account of the various ways an agent can be said to be identified or alienated from their actions and desires, but something like the tripartite view seems to be required to make sense of the complexities involved.

In this respect, this final chapter mirrors the first. There, I argued that making sense of weakness of will, and in particular how weak-willed actions can be intentional, required something like the tripartite view. Then I spent several chapters filling out this tripartite view, arguing for conceptions of desire and the will that were independently plausible and theoretically powerful. Now I have tried to show that putting the parts together again can do work in addressing other important problems in moral psychology. These three layers of argument combine, I hope, to make the tripartite view an attractive and compelling general approach to a philosophical understanding of the psychology of rational action.

## References

- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders.
- Anscombe, G. (1963). *Intention* (2nd ed.). Oxford: Blackwell.
- Arpaly, N. (2003). *Unprincipled virtue*. Oxford University Press.
- Audi, R. (1986). Intending, intentional action, and desire. In J. Marks (Ed.), *The ways of desire*. Precedent Publishing, Inc.
- Audi, R. (1993). *Action, intention, and reason*. Cornell University Press.
- Baier, A. (1986). The ambiguous limits of desire. In J. Marks (Ed.), *The ways of desire*. Precedent Publishing, Inc.
- Baker, D. (2014). The abductive case for humeanism over quasi-perceptual theories of desire. *Journal of Ethics and Social Policy*, 8(2).
- Blackburn, S. (2010). The majesty of reason. *Philosophy*, 85(1), 5–27.
- Blackburn, S. (1998). *Ruling passions*. Oxford: Clarendon Press.
- Boghossian, P. (2003). The normativity of content. *Philosophical Issues*, 13(1), 31–45.
- Boyle, M., & Lavin, D. (2010). Goodness and desire. In S. Tenenbaum (Ed.), *Desire, practical reason, and the good*. Oxford University Press.
- Bratman, M. (1979). Practical reasoning and weakness of the will. *Nous*.
- Bratman, M. (1987). *Intention, plans, and practical reason*. Harvard University Press.
- Bratman, M. (2007). *Structures of agency*. Oxford University Press.
- Brewer, T. (2006). Three dogmas of desire. In T. Chappell (Ed.), *Values and virtues: Aristotelianism in contemporary ethics* (pp. 257–284). Oxford University Press.
- Brewer, T. (2009). *The retrieval of ethics*. Oxford University Press.
- Bumpus, A. (2001). Actors without intentions: The double phenomena view. *Philosophical Studies*, 103, 177–199.
- Callard, A. (2008). *An incomparabilist account of akrasia* (PhD thesis). University of California, Berkeley.
- Cartwright, N., Cat, J., Fleck, L., & Uebel, T. (1996). *Otto neurath: Philosophy between science and politics*. Cambridge University Press.

- Copp, D., & Sobel, D. (2002). Desires, motives, and reasons: Scanlon's rationalistic moral psychology. *Social Theory and Practice*, 28(2), 243–276.
- Dancy, J. (2000). *Practical reality*. Oxford University Press.
- Darwall, S. (2006). *The second-person standpoint: Morality, respect, and accountability*. The MIT Press.
- Davidson, D. (1980). *Essays on actions and events*. Oxford University Press.
- Davis, W. (1984). A causal theory of intending. *American Philosophical Quarterly*, 21(1), 43–54.
- Ekstrom, L. W. (2005). Alienation, autonomy, and the self. *Midwest Studies in Philosophy*, 45–67.
- Frankfurt, H. (1988). *The importance of what we care about*. Cambridge University Press.
- Frankfurt, H. (1992). The faintest passion. *Proceedings and Addresses of the American Philosophical Association*, 66(3), 5–16.
- Garrard, E., & McNaughton, D. (1998). Mapping moral motivation. *Ethical Theory and Moral Practice*, 1(1), 45–59.
- Ginet, C. (1990). *On action*. Cambridge University Press.
- Ginsborg, H. (2006). Reasons for belief. *Philosophy and Phenomenological Research*, 72(2), 286–318.
- Hawkins, J. (2008). Desiring the bad under the guise of the good. *The Philosophical Quarterly*, 58(231), 244–264.
- Hieronymi, P. (2009). The will as reason. *Philosophical Perspectives*, 23.
- Holton, R. (2009). *Willing, wanting, waiting*. Oxford University Press.
- Johnston, M. (2001). The authority of affect. *Philosophy and Phenomenological Research*.
- Karniol, R., & Miller, D. T. (1983). Why not wait?: A cognitive model of self-imposed delay termination. *Journal of Personality and Social Psychology*, 45(4), 935–942.
- Korsgaard, C. (1986). Skepticism about practical reason. *Journal of Philosophy*, 83, 5–25.
- Korsgaard, C. (2004). Fellow creatures: Kantian ethics and our duties to animals. In
- Lenman, J. (2011). Reasons for action: Justification vs. explanation. *Stanford Encyclopedia of Philosophy*.

- Lewis, D. (1988). Desire as belief. *Mind*, 97, 323–32.
- McCann, H. (1974). Volition and basic action. *Philosophical Review*, 83(4), 451–473.
- McDowell, J. (1980). The role of eudamonia in aristotle’s ethics. In A. O. Rorty (Ed.), *Essays on aristotle’s ethics*. University of California Press.
- McDowell, J. (2009). *Having the world in view*. Harvard University Press.
- Mele, A. (1992). *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.
- Mele, A. (1997). Introduction. In A. Mele (Ed.), *The philosophy of action*. Oxford University Press.
- Miller, C. (2013). Identifying with our desires. *Theoria*, 79(2), 127–154.
- Mischel, W. (1996). From good intentions to willpower. In P. Gollwitzer & J. Bargh (Eds.), *The psychology of action* (pp. 192–218). Guilford Press.
- Nagel, T. (1970). *The possibility of altruism*. Princeton University Press.
- Oddie, G. (2005). *Value, reality, and desire*. Oxford University Press.
- O’Shaughnessy, B. (1980). *The will* (Vols. 1, 2). Cambridge University Press.
- Pears, D. (1982). Motivated irrationality. *Proceedings of the Aristotelian Society*.
- Pearson, G. (2012). *Aristotle on desire*. Cambridge University Press.
- Pink, T. (1995). *The psychology of freedom*. Cambridge University Press.
- Quine, W., & Ulian, J. (1970). *The web of belief*. New York: Random House.
- Quinn, W. (1993). Putting rationality in its place. In R. Frey & C. W. Morris (Eds.), *Value, welfare, and morality*. Cambridge University Press.
- Raz, J. (1997). The active and the passive. *Aristotelian Society Supplementary Volume*, 71, 211–228.
- Raz, J. (1999). Incommensurability and agency. In *Engaging reason: On the theory of value and action* (pp. 46–66). Oxford University Press.
- Raz, J. (2010). On the guise of the good. In S. Tenenbaum (Ed.), *Desire, practical reason, and the good*. Oxford University Press.
- Ridge, M. (1998). Humean intentions. *American Philosophical Quarterly*.
- Scanlon, T. (1998). *What we owe to each other*. Harvard University Press.
- Schafer, K. (2013). Perception and the rational force of desire. *Journal of Philosophy*, 110.
- Schapiro, T. (2009). The nature of inclination. *Ethics*, 119, 229–256.

- Schapiro, T. (2014). What are theories of desire theories of? *Analytic Philosophy*, 55(2), 131–150.
- Schroeder, M. (2007). *Slaves of the passions*. Oxford University Press.
- Schroeder, M. (2008). How does the good appear to us? *Social Theory and Practice*, 34(1), 119–30.
- Schueler, G. (1995). *Desire: Its role in practical reason and the explanation of action*. The MIT Press.
- Setiya, K. (2004). Hume on practical reason. *Philosophical Perspectives*, 18, 365–89.
- Setiya, K. (2007). *Reasons without rationalism*. Princeton University Press.
- Sinhababu, N. (2013). The desire-belief account of intention explains everything. *Nous*.
- Smith, A. (2005). Responsibility for attitudes: activity and passivity in mental life. *Ethics*, 115(2), 236–271.
- Smith, M. (1987). The Humean theory of motivation. *Mind*, 96, 36–61.
- Smith, M. (1994). *The moral problem*. Cambridge University Press.
- Smith, M. (2013). A constitutivist theory of reasons: Its promise and parts. *Law, Ethics, and Philosophy*, 1.
- Stampe, D. (1987). The authority of desire. *The Philosophical Review*, 96(3), 335–381.
- Steglich-Petersen, A. (2006). No norm needed: On the aim of belief. *The Philosophical Quarterly*, 56(225), 499–516.
- Stroud, B. (2013). *Engagement and metaphysical dissatisfaction*. Oxford University Press.
- Tenenbaum, S. (2007). *Appearances of the good*. Cambridge University Press.
- Tenenbaum, S. (2008). Appearing good: A reply to Schroeder. *Social Theory and Practice*.
- Velleman, D. (2000). *The possibility of practical reason*. Ann Arbor, MI: University of Michigan Library.
- Velleman, D. (2007). What good is a will? In A. Leist (Ed.), *Action in context*. Walter de Gruyter.
- Velleman, J. D. (1989). *Practical reflection*. Princeton University Press.
- Velleman, J. D. (1992). The guise of the good. *Nous*, 26, 3–26.
- Walker, A. (1989). The problem of weakness of will. *Nous*.
- Wallace, R.J. (2006). *Normativity and the will*. Oxford University Press.

- Wallace, R. J. (1999). Addiction as a defect of the will: Some philosophical remarks. *Law and Philosophy*, 18(6), 621–54.
- Wallace, R. J. (2006a). How to argue about practical reason. In *Normativity and the will*. Oxford University Press.
- Wallace, R. J. (2006b). Three conceptions of rational agency. In *Normativity and the will*. Oxford University Press.
- Wallace, R. J. (2014). Reasons, policies, and the real self: Bratman on identification. In M. Vargas & G. Yaffe (Eds.), *Rational and social agency: The philosophy of michael bratman*. Oxford University Press.
- Watson, G. (2004). The work of the will. In *Agency and answerability*. Oxford University Press.
- Williams, B. (1979). Internal and external reasons. In R. Harrison (Ed.), *Rational action* (pp. 101–113). Cambridge University Press.
- Williams, B. (1981). *Moral luck*. Cambridge University Press.
- Zhu, J. (2004). Understanding volition. *Philosophical Psychology*, 17(2).