

# **AN ESSAY ON THE DESIRE-BASED REASONS MODEL**

By

Attila Tanyi

A Doctoral Dissertation

Submitted to the Central European University  
in partial fulfillment of the requirements for the Degree of  
DOCTOR OF PHILOSOPHY

Central European University  
Department of Political Science  
November 2006

© 2006

ATTILA TANYI

All Rights Reserved.

## ABSTRACT

This dissertation aims to contribute to the discussion about the viability of what is sometimes labeled as the classical theory of practical reason: the Desire-Based Reasons Model (the Model). The line of argumentation employed is negative in character. Its aim is to not to construct a novel theory of practical reason, but to examine and criticize the Model from different angles. To do so, we need first a detailed presentation of the Model; this is the task of Chapter I. Since the Model offers us an account of normative reasons, the chapter focuses on the clarification of this notion. The strategy employed is comparative: I discern the notion by contrasting it with the notion of motivating reason. The framework thus arrived at helps me to distinguish three versions of the Model against which I argue in proceeding chapters.

Chapter II is the first step on that road. It attacks the second and third version of the Model through their naturalist underpinnings. My aim is to show that the Model understood in this way is unable to account for the normativity of reason-claims. To this end, I employ a recent argument by Derek Parfit that points to a problem with the naturalist account of normativity. Parfit's claim is this: naturalism trivializes the agent's practical argument and therefore abolishes the normativity of its conclusion. Although Parfit intends his objection to refute naturalism *per se*, my analysis shows that naturalists might be able to avoid his criticism in case they can vindicate the reduction proposed. However, by developing an argument borrowed from Connie Rosati, I show that this is exactly what advocates of the Model are unable to do.

Chapter III takes up another line of argument against the second and third versions of the Model. The approach I consider questions the idea that the reason-relation must contain reference to the agent's desires. There are several ways to do this, but I focus on the attempt that in my view promises the most: the idea of reason-based desires. On this view, since desires are based on reasons (first premise), which they transmit but to which they cannot add (second premise), they cannot themselves provide reasons for action. In the chapter I defend both premises against potential counter-examples. Furthermore, in the course of doing so, I also consider and reject the so far neglected first version of the Model.

Chapter IV turns back to the second and third version of the Model and investigates their motivational defense. The defense infers the Model from two premises: the Internalism Requirement (IR) and the Humean Theory of Motivation (HTM). In the chapter I attack the latter by focusing on its three corollary theses. These are: desires must (a) have real psychological existence and be present when action takes place, I call this the Existence Criterion (EC); must (b) constitute, together with a suitable instrumental belief, the agent's motivating reason, which I label the Motivational Criterion (MC); and (c) must be independently intelligible from beliefs, which gets the title of the Intelligibility Criterion (IC). In the course of discussing the path that leads to my preferred solution, I argue that the EC makes sense as a requirement, whereas rejection of the IC would take us too far from the scope and elements of the HTM. Analysis of further objections, however, shows that the MC is not met because the role it attributes to desires makes it impossible for them to serve as motivators. A version of Jonathan Dancy's pure cognitivism is true: it is beliefs about the object of the desire together with corresponding normative beliefs that constitute the agent's motivating reason. I call the resulting theory the Cognitivist Theory of Motivation (CTM) and devote the remainder of the chapter to its elaboration and defense.

## ACKNOWLEDGEMENTS

I have accumulated many debts in the course of writing this thesis. My greatest thanks go to my supervisor at CEU, Professor János Kis who has patiently provided me with support and inspiration through the years of my doctoral research. I would also like to express my gratitude to Professor Krister Bykvist, fellow of Jesus College who supervised my research at the University of Oxford in the academic year 2002/2003. His friendly and encouraging guidance helped further clarify the positions defended in my thesis. My gratitude is even greater given that he was willing to read and discuss my work after I left Oxford, even though it was no longer his duty to do so.

The thesis would not have been possible without the excellent academic environment of the Political Science and Philosophy Departments of the Central European University, at whose doctoral seminars I have presented my work. I would like to thank all the participants in these seminars. I specifically thank Loránd Ambrus-Lakatos, Gergely Bognár, Serhiy Pukas and András Miklós for their comments and the stimulating discussion we had on the thesis as well as on my thinking in general. In its different versions almost all parts of the dissertation benefited from the comments of several people. Although the list I provide here is certainly not exhaustive, I hope to make best use of my memory. Jonathan Dancy and Joseph Raz read several drafts of Chapter I, while Ferenc Huoranszki and Anders Strand helped me focus my argument in the chapter. My discussion with John Skorupski at a summer school in the Netherlands about his work and related matters also influenced my thinking in this chapter. Some of the comments also appear in Appendix I, which originally formed part of the chapter. Chapter II was presented at several conferences and seminars, most notably at the ECAP 5 conference in Lisbon, Portugal and the 'Ethical Naturalism' conference in the University of Durham. I thank audiences at both places: some of the comments and objections appear in my discussion. An important part of Chapter III concerning the role of pleasure would not have been possible without the generous help of Leonard D. Katz who took the effort of trying to explain his position to me. I take this occasion to thank him for his patient correspondence about these matters. Chapter IV also bears the influence of several people. Fred Schueler read an early draft, whereas later versions were commented on by Greg Bognár, Eric Brown, Daniel Friedrich and Edmund Henden. The chapter was also presented at the Political Science Department's Annual Doctoral Conference. I thank Laura Andronache who served as my commentator on this occasion. Finally, Appendix II owes a lot to my correspondence with William FitzPatrick whose work on Korsgaard's constructivist theory influenced heavily my discussion.

My special thanks go to all those institutions that funded my research in the past seven years. First and foremost, I thank the Central European University for all the financial and logistical support. I also thank the United Kingdom's Foreign and Commonwealth Office for providing the Chevening Scholarship for my research at Oxford University, the Research Council of Norway that in cooperation with the Hungarian Scholarship Board funded my stay at the University of Oslo, and the EUSSIRF grant that helped me use the facilities of the European University Institute in Florence. I also want to express my gratitude to the coordinators of my program, Éva and Kriszta who provided me with help whenever I needed it. Finally, I would like to thank my parents for their assistance and encouragement throughout the writing of this thesis.

ABSTRACT .....	3
ACKNOWLEDGEMENTS .....	4
INTRODUCTION.....	7
CHAPTER I: REASONS AND THE MODEL .....	12
<b>I. Motivating versus normative reasons</b> .....	12
1.1.1. Two types of motivating reason: first difference .....	13
1.1.2. Two types of motivating reason: second difference.....	17
1.1.3. Motivating versus normative reasons.....	20
<b>II. Normative reasons</b> .....	23
1.2.1. The structure of normative reason.....	23
1.2.2. Mental states as normative reasons: desires .....	25
1.2.2. Mental states as normative reasons: beliefs .....	31
<b>III. The Model</b> .....	40
1.3.1. First version.....	41
1.3.2. Second version .....	43
1.3.3. Third version .....	51
<b>IV. Refuting the Model</b> .....	54
CHAPTER II: NATURALISM AND TRIVIALITY .....	56
<b>I. Naturalism and the Model</b> .....	56
2.1.1. Ethical naturalism.....	57
2.1.2. Naturalism and the Model .....	61
<b>II. The objection: triviality</b> .....	62
2.2.1. The objection outlined.....	62
2.2.2. First response: irreductionist naturalism .....	66
2.2.3. Second response: non-descriptive naturalism .....	70
2.2.4. Third response: vindicatory reductionism.....	73
2.2.5. Summary .....	76
<b>III. The condition: tolerable revisionism</b> .....	76
2.3.1. Tolerable revisionism defined.....	76
2.3.2. Rosati’s objection.....	79
2.3.3. The ideal advisor account as the best idealizing approach.....	84
2.3.4. Attempts to meet Rosati’s objection .....	87
2.3.5. Attempts to avoid Rosati’s objection .....	92
2.3.6. Summary .....	96
<b>IV. Summary and what follows</b> .....	97
CHAPTER III: REASON AND DESIRE.....	98
<b>I. Reason-based desires: the first premise</b> .....	98
3.1.1. Issues not discussed.....	99
3.1.2. First counterexample: affective desires.....	102
3.1.3. Second counterexample: hedonic desires.....	110
3.1.4. Third counterexample: urges.....	120
<b>II. Reason-based desires: the second premise</b> .....	126

3.2.1. The primary application .....	127
3.2.2. The secondary application.....	133
<b>IV. What's left: the theory of motivation.....</b>	<b>139</b>
<b>CHAPTER IV: A COGNITIVIST THEORY OF MOTIVATION .....</b>	<b>140</b>
<b>I. The motivational defense of the Model.....</b>	<b>140</b>
4.1.1. The Internalism Requirement.....	141
4.1.2. The Humean Theory of Motivation .....	145
4.1.3. The argument.....	148
<b>II. Refuting the HTM .....</b>	<b>150</b>
4.2.1. Nagel's pure ascription theory .....	150
4.2.2. Motivated desires .....	155
4.2.3. Besires and desire entailing beliefs .....	160
4.2.4. The Cognitivist Theory of Motivation .....	165
<b>III. Defending the CTM.....</b>	<b>174</b>
4.3.1. The CTM and previous objections .....	174
4.3.2. The CTM and new objections .....	177
<b>IV. Summary.....</b>	<b>187</b>
<b>CONCLUSION .....</b>	<b>189</b>
<b>APPENDIX I: REASONS – FACTS OR PROPOSITIONS? .....</b>	<b>192</b>
<b>I. Reasons: facts or propositions? .....</b>	<b>192</b>
<b>II. Dancy's objections .....</b>	<b>193</b>
<b>APPENDIX II: KORSGAARD'S CIRCULARITY ARGUMENT .....</b>	<b>197</b>
<b>I. Korsgaard's argument.....</b>	<b>197</b>
<b>II. Extending the argument.....</b>	<b>198</b>
<b>III. Responses to the argument.....</b>	<b>203</b>
A.2.3.1. First response.....	203
A.2.3.2. Third response .....	206
A.2.3.3 Second response .....	208
<b>APPENDIX III: FAILED ATTEMPTS TO REFUTE THE MODEL.....</b>	<b>218</b>
<b>I. Five ways to reject the Model .....</b>	<b>218</b>
<b>II. Why the attempts fail .....</b>	<b>219</b>
A.3.2.1. First argument.....	219
A.3.2.2. Second argument .....	219
A.3.2.3. Third argument .....	220
A.3.2.4. Fourth argument .....	224
A.3.2.5. Fifth argument .....	226
<b>BIBLIOGRAPHY .....</b>	<b>228</b>

## INTRODUCTION

Reasons are everywhere. We encounter them whenever we deliberate about what to do, think or feel. This in itself masters enough legitimization for a study of reasons. But there are more particular motivations for such a study. Two of them especially inspired the writing of this thesis. In political philosophy, one central issue is the question of political legitimacy and its corresponding idea of political obligation. And one particularly influential account, a view that I also share, claims that legitimacy should be grounded in the sound reasons of the subjects of political authority. Hence our task is to see what sort of reasons these are, where they come from, and how they function in this context. The second inspiration comes from moral theory. There is a widely held objection to certain, typically Kantian and consequentialist theories. The charge is that these accounts demand too much of people: they require them to promote the good or to do what is right to a point where morality ceases to be a tenable project for ordinary human beings. There are several ways to counter this objection but one looks particularly attractive. If ethics is a practical subject - which, we should admit, is a view that not everyone accepts - we can examine whether morality provides us with reasons and if it does, what sort of reasons these are, where they come from, and how strong they are.

It is thus hard to overrate the importance of reasons; the question is how we set out to investigate them. In the literature there are two main ways to deal with reasons. One examines them in a very general manner, focusing mainly on their logical structure, metaphysics and epistemology. The other investigates particular theories of reason, that is to say, substantive views about what gives people reason to act, feel or think. This dissertation aims to combine the lessons of the former approach with the substance of the latter, while focusing on reasons for

action. The idea is to present and analyze one particular theory of practical reason, which is often labeled as the ‘classical theory’ and is taken to be the default position by many of its supporters. The theory, which I call the Desire-Based Reasons Model or simply the Model, claims that what we have reason to do is what would satisfy one or more of our desires. This brings in view a well-known antagonism in meta-ethics: that between desire-based and value-based theories of practical reason. The aim of this dissertation in this context is negative: it is to argue against the former without attempting to give a novel formulation of the latter.

The argument proceeds as follows. The aim of Chapter I is to give a detailed presentation of the Model. Since the Model offers us an account of normative reasons, the chapter focuses on the clarification of this notion. The strategy employed is comparative: I discern the notion by contrasting it with the notion of motivating reason. Accordingly, in the beginning of the chapter, I devote some time to clarifying our concept of motivating reason. I argue for three points. First, there is only one kind of motivating reason: that of the agent’s reasons. Second, the agent’s reasons have a hybrid, normative-explanatory character. Finally, the agent’s reasons are nevertheless separable from normative reasons. This last claim requires further elaboration leading me to investigate the notion of normative reason on its own. After presenting the general structure of normative reason, I discuss its ontology and argue that normative reasons are not mental states. The framework thus arrived at helps me to accomplish my ultimate task: to give a detailed description of the Model by distinguishing its three main versions.

The objective of subsequent chapters is to argue against the Model from different angles. Discussion in these chapters mainly concentrates on the Model’s second and third version, postponing the discussion of the first until Chapter III. In this manner, Chapter II attacks the Model through its naturalist underpinnings. My aim is to show that the Model understood in this way is unable to account for the normativity of reason-claims. For this purpose I first describe



how the Model is presented along naturalist lines. After this I investigate a recent argument by Derek Parfit that points to a problem with the naturalist account of normativity. Parfit's claim is this: naturalism trivializes the agent's practical argument and therefore abolishes the normativity of its conclusion. Although Parfit intends his objection to refute naturalism *per se*, my analysis shows that naturalists might be able to avoid his criticism in case they can vindicate the reduction proposed. This, however, leads us to further conditions whose fulfillment is necessary for the success of vindication. I deal with one such condition: the demand that naturalist reductions must be tolerably revisionist. Invoking a recent argument by Connie Rosati, I show that insofar as the Model is concerned, the condition is not met. Hence, though not a knock down argument as Parfit intended it to be, the triviality objection does pose a serious threat to the Model.

Chapter III takes up another line of argument against the Model. The approach I consider questions the idea that our understanding of reason must contain reference to the agent's desires. There are several ways to do this, but I focus on the attempt that promises the most: the idea of reason-based desires. On this view, since desires are based on reasons (first premise), which they transmit but to which they cannot add (second premise), they cannot themselves provide reasons for action. The first part of the chapter is devoted to the fleshing out and defense of the first premise. I consider three possible counter-examples – affective desires, hedonic desires, and urges – and argue that they can all be accounted for without compromising the premise. After this, in the second part of the chapter, I discuss the second premise of the idea. Here again there are potential counter-examples, most importantly cases when the desire acts as a tie-breaker in the conflict of reasons. In opposition I claim that desire can only serve this role by virtue of its motivational power but then it lacks the normative potential necessary for the objection to go through. Here, at the end of the section, I also consider the first version of the Model as a possible source for the denial of the second premise as well as a source in itself.

Chapter IV turns back to the second and third version of the Model and investigates their motivational defense. The defense infers the Model from two premises: the Internalism Requirement (IR) and the Humean Theory of Motivation (HTM). In the chapter I attack the latter by focusing on its three corollary theses. These are: desires must (a) have real psychological existence and be present when action takes place, I call this the Existence Criterion (EC); must (b) constitute, together with a suitable instrumental belief, the agent's motivating reason, which I label the Motivational Criterion (MC); and (c) must be independently intelligible from beliefs, which gets the title of the Intelligibility Criterion (IC). In the course of discussing the path that leads to my preferred solution, I argue that the EC makes sense as a requirement, whereas rejection of the IC would take us too far from the scope and elements of the HTM. Analysis of further objections, however, shows that the MC is not met because the role it attributes to desires makes it impossible for them to serve as motivators. A version of Jonathan Dancy's pure cognitivism is true: it is beliefs about the object of the desire together with corresponding normative beliefs that constitute the agent's motivating reason. I call the resulting theory the Cognitivist Theory of Motivation (CTM) and devote the remainder of the chapter to its elaboration and defense.

The dissertation also has three appendices in which I discuss issues that were left out of the chapters as not being essential for the arguments presented there. Yet, since they are important in their own right and tell us interesting new details about the Model, I judge them to be worthy of discussion. Appendix I takes up an issue that was set aside in Chapter I: the question whether reasons are states of affairs or propositions. In my brief treatment, I discuss two recent objections by Jonathan Dancy, which I think present the best case against the propositional reading. Due to the nature of my discussion, I don't claim to make a decisive case for the propositional account; at best I draw attention to the complexity of the issues involved. Appendix

II deals with Christine Korsgaard's influential circularity argument against the rationalist account of the instrumental principle. Here I first argue that the argument can be extended to the naturalist account of the Model, and then I examine several attempts to meet Korsgaard's charge so understood. Some of these attempts I claim to be promising, though often not properly worked out, others I argue to be failed attempts. Finally, Appendix III considers those arguments against the Model that, being unsuccessful, were left out of Chapter III. Since much of what I say here has already appeared in press, I am brief both in my exposition as well as in my discussion of the attempts.

## CHAPTER I: REASONS AND THE MODEL

This dissertation aims to contribute to the discussion about the viability of what is sometimes called the classical theory of practical reason: the Desire-Based Reasons Model (from now: the Model). The present chapter is the first step on that road. Since the Model gives us an account of normative practical reasons (from now: normative reason), it seems sensible to devote this chapter to the examination of the notion of normative reason. The resulting answers will be useful in two ways. On the one hand, they will help me to define my target: to give a properly circumscribed account of the Model. On the other hand, they will identify those elements that need to be attacked for a successful refutation of the Model.

### ***1. Motivating versus normative reasons***

Practical reason is inherently normative. It purports to guide human conduct by recommending courses of action. In ordinary discourse, however, we use the term ‘reason’ in many ways. Several of these applications employ the term in a non-normative way. Therefore, in this first section of this chapter, I spell out and analyze the different reason-sentences one finds in practical discourse. I argue for three points:

- (1) There is only one kind of motivating reason: that of the agent’s reason;
- (2) The agent’s reasons have a hybrid, normative-explanatory character;
- (3) Normative reasons are nevertheless separable from the agent’s reasons.

I proceed as follows. I first examine the two types of reason-sentences that refer to the explanatory application of 'reason'. Here I argue for points (1) and (2) above. Then I specify the third and fourth applications, which both use the term normatively and show their relation to the explanatory senses. This discussion will, however, only partially cover the third claim above; the need for further elaboration will lead us to the next section.

### **1.1.1. Two types of motivating reason: first difference**

The explanation of intentional action refers to motivating reasons. However, there are two different ways to explain action, and this suggests that there are two senses of motivating reason.

I have in mind the following explanatory sentences:

- (I) The reason why *A*  $\varphi$ -d was that *q*.
- (II) *A*'s reason for  $\varphi$ -ing was that *q*.

We can illustrate the difference with examples. In conversation we can often observe that the participants mimic each other's bodily postures. From this one can infer that the reason why one participant changed his posture was the slightly earlier similar change in posture by another participant. Yet, we wouldn't refer to this consideration as the mimicking *discussant's reason* for changing posture. (Darwall 1983, 29) Jonathan Dancy offers further examples. (Dancy 2000b, 5)

What explains why many people buy new, trendy clothes is the effect TV advertisements have on them, yet, this is not among *their* reasons to go shopping. Similarly, what explains why I didn't go to the party was that I am shy, yet, *my* reason for not going was probably that there would be lots of unfamiliar people there. What explains why someone responded so aggressively may be that he has taken some sort of medication, yet, this wasn't *his* reason to respond in that way. And

what explains why the teacher gave his student a better grade than she deserved is that he was influenced by her looks, yet, *his* reason to do so was probably that he pitied her.

No one really denies that there are these two ways of explaining action, but it is more doubtful that there are indeed two senses of motivating reason corresponding to them. This is not a merely verbal matter of how to use the term ‘reason’ in explaining action. One’s position on this question can influence one’s theory of motivation by effecting the metaphysics of motivating reason one ascribes to (see the arguments in Dancy 2000b, Chapter 5-6 and the debate between Dancy 2003b, Smith 2003, and Wallace 2003). At present, however, it is useful to deal with this problem for another reason: it can show us more about the connection between motivating and normative reasons. Let me begin with stating my view on the matter. I believe that it is at least doubtful whether reasons-why explanation legitimately uses the term ‘motivating reason’ in explaining action. More precisely, it is doubtful if they use the term legitimately apart from cases when they also pick out the agent’s reasons.<sup>1</sup> There are two general considerations that support this position. I don’t claim that they decide the issue; yet, they offer some ground for my preferred conclusion.

Here is the first idea. The circle of events that is open to reasons-why explanations can include events other than intentional actions, whereas explanations in terms of the agent’s reasons only deal with such actions. Three types of examples show this. I start with the most straightforward cases, and then proceed to the more complex, though I think still fairly uncontroversial examples. In the first group, we find cases where the event concerned is more like a *reaction* than an intentional action. Take Darwall’s example with the discussing parties. Here the discussant’s mimicry cannot count as a full-blown action since the agent is missing from

---

<sup>1</sup> I am not alone with this thought: both Jonathan Dancy and Stephen Darwall hold similar positions. See Dancy (2000b) and Darwall (1983) about this. There is considerable opposition though. See Smith (1994), pp. 92-3, 131-3; (2003), pp. 464-5; Parfit (1997), pp. 114n; Cuneo (2002), pp. 477-8; Wallace (2003), pp. 431-2.

it. Although it is the discussant who performs the movement, it is not under his control in any sense: it is a reaction, not an action. Reflex movements invite the same diagnosis. To take an example from David Velleman, if my child suddenly brushes the glass off the table and my hand shoots out to catch it, then my action did not involve me properly speaking. It was my body of course that moved, but I was not involved in any other way. (Velleman 2000c, 189-190) In explaining both events, however, it is perfectly legitimate to give a reasons-why explanation, whereas we find it false to speak of the discussant's reason or the father's reason. Hence, unless we want to deflate the notion entirely, we should conclude that reasons-why explanations don't give us another sense of motivating reason.

Cases of so-called deviant causal chains form the second group of examples. In Donald Davidson's case a climber knows that he could get rid of the weight and danger of holding another man on a rope by loosening his hold on it. However, the belief (by loosening the hold he could avoid danger) and desire (to avoid danger and get rid of weight) might so unnerve him as to cause him to loosen his hold without his ever choosing to do so. (Davidson 1980c, 79) In Roderick Chisholm's case a nephew accidentally runs over his uncle as an indirect result of nervousness produced by his desire (to inherit a fortune) and belief (that by killing his uncle he would inherit it). (Chisholm 1966, 19-20) Here again we have the same diagnosis. On the one hand in neither situation does it sound plausible to say that the agent was properly involved in the action, hence these events don't qualify as full-blown intentional actions. On the other hand in both cases we can give a reasons-why explanation, whereas we wouldn't explain these actions with reference to the climber's reason or the nephew's reason. Consequently we had better refer to these belief-desire pairs as causes and not as motivating reasons just as we did in the previous cases.

Cases of the third type involve the medication, the TV advertisement, and the teacher's case. Our thesis stands on shaky ground here. It seems that these events are instances of intentional action, hence the thesis doesn't follow. Yet, I think, this appearance is explained by the fact that in these cases there is another explanation invoking the agent's reasons. If we disregard this alternative explanation and just take the skeleton of the cases, that is, the part covered by the reasons-why explanation, the thesis follows. Let me illustrate this with an example. (Velleman 2000b, 126-7) In a meeting with an old friend the purpose of which is to resolve some minor difference, the agent finds himself raising his voice due to the offhand comments of his friend, until they part in anger. Reflecting upon it later, the agent realizes that accumulated grievances of the past led him to a resolution, weeks before their meeting, to sever their friendship over the matter at hand. But though it is true that the agent's desires, through his decision, caused his behavior, there is no reason to think that he made the decision or that he executed it: the agent is again not involved. I think the same analysis is applicable in our original examples and the alternative explanation only colors but doesn't significantly change the analysis. Hence our thesis follows.

One can dispute my claim in two ways. Perhaps in assessing the examples I was misled by the fact that these events are automatic: this is why I thought that the agent was not involved in the act. But intentional action can also be spontaneous; hence that the events concerned are spontaneous doesn't itself deprive them of intentionality. This argument is mistaken. The agent is missing from the discussed cases not because they are spontaneous. In cases of the third type, the concerned factors operate *subconsciously*: they lay deeply hidden in the agent's consciousness, are often surpassed and sometimes impossible to dig up. Or they have no connection to consciousness and exert a purely mechanical influence as in the medication case. The events the other examples involve, on the other hand, are unintentional because they are reactions and reflex



movements, not because they are automatic. In all cases spontaneity is more of a symptom than a cause; it is certainly not enough to deprive these events of their intentionality. Another objection would be to claim that reasons-why explanations are a species of teleological explanation: the bodily movement has a goal, though it is a goal not set by the intention of the agent. However, there is no guarantee that teleological explanations only have actions as their explanandum. To take Velleman's first example, my act of catching the falling glass clearly has a goal, yet it is not an intentional action, but only a reflex movement. Having a purpose is not enough to make an action intentional.<sup>2</sup>

### **1.1.2. Two types of motivating reason: second difference**

There is a second important difference between the agent's reasons and the considerations that appear in reasons-why explanations. Earlier I said that practical reason is inherently normative but that there are non-normative applications of it. Yet one might, I think, sensibly claim that the term 'reason', no matter where it is used, has some normative connotation and that even if in explanation we don't use normative reasons properly speaking, the considerations we refer to should still have at least the *appearance* of normativity about them. This is the only way, so to speak, to see the reasons, which explain the agent's action *as reasons* and not merely as causes. And, the argument concludes, while the agent's reasons do indeed have this appearance and thus permit this kind of explanation, the considerations picked out by reasons-why explanations do

---

<sup>2</sup> The relation between acts and purposes can be weak and strong. It is *weak* when the relation claimed is that every intentional action has a purpose: having a purpose is a necessary requirement for an event to qualify as intentional action. The relation is *strong* when in addition we claim that every event that has a purpose is an action: having a purpose is a necessary and sufficient requirement for an event to qualify as intentional action. In this dissertation I won't deny the weak claim, only the strong. But there are some who deny the weak claim as well. See Stocker (1981), pp. 750-758; though see Smith (2004a), pp. 161-163 for a counter-argument against Stocker and others. Also cf. Mele (2003), Chapter 2.

neither (unless they are identical with the agent's reasons). (cf. Nagel 1986, 142; Dancy 2000b, 97)

The claim is that the agent's reasons have a normative aspect to them. This appears to be true by default: the essential feature of the agent's reason is that it is regarded as *normative* by its possessor. (Baier 1958, 156; 1978, 711; Locke 1974, 170; Milligan 1974; Darwall 1983, 33; Gibbard 1990, 162; Grice 2001, 41) There are exceptions such as weakness of will or urges and cravings. But even in these cases we have something to say. When the agent acts on an urge or a craving, he normally disowns his reasons: he doesn't understand why he does what he does and cannot see the reasons as his own. Therefore it is hard to speak of *his* reasons at all. Weak-willed agents, on the other hand, often act on a consideration they take to be a reason; what is missing in their case is the *further* thought that the consideration is also the *best* reason to act on. When this is not the case, when the agent doesn't even hold the first view, we get very close to urges and cravings and the analysis they invite. Hence, we can say that agents almost always take their reasons to be normative. That is, the reasons referred to in sentences of the form (II) are hybrid, normative-explanatory reasons.<sup>3</sup>

Let me briefly offer two considerations that explain why people have the tendency to see their reasons as normative. Note that these considerations do not identify the *sources* of normativity; they only explain our *need* to find such a source. The first concerns the ability that only humans possess: reflection. Reflection enables people to decide that some consideration *be* a reason. They can see it as good reason; they try to collect such reasons, contrast them and make decisions on their basis. The emphasis on reflection need not bring with it an undue focus on

---

<sup>3</sup> We should be careful not to confuse this understanding of the agent's reasons with a different idea that I mention in footnote 20 below: that beliefs - typically false beliefs - can serve as excuses when explaining action and as such can be taken to be 'justifying' reasons. That role, as I emphasize it there, is not a normative role either; but it is not normative for a different reason: it does not have the appearance of normativity; it has no such 'purpose'.

conscious thinking though. True, reflection is most clearly present here. It turns our internal thought into a kind of dialogue with ourselves in which we object to and accede to arguments that we present to ourselves. (Gulliver 1973, 681; Hastie and Pennington 1988) But reflection appears in our apparently unreflective behavior too. (Scanlon 1998, 23-4) Our spontaneously formed attitudes are constrained by general standing judgments about the adequacy of reasons. If a person holds that some evidence is not good round to believe something he doesn't normally form that belief automatically; and the same is true of spontaneously formed intentions and actions. Finally, reflection also pervades our thinking after the attitude was formed or the action done. And the result is the same: we try to find justification; we try to see our reasons as normative. (Haidt 2001, 814; Nisbett and Wilson 1977)

A dialogue from Henry James' novel 'The Portrait of a Lady' best illustrates the second consideration. In the book we find the following conversation between the heroine, Elizabeth Archer and her friend Miss Stackpole:

"Well", said Miss Stackpole at last, "I have only one criticism to make. I don't see why you promised little Miss Osmond to go back."

"I am not sure that I see myself, now," Isabel replied. "But I did then."

"If you have forgotten your reason perhaps you won't return."

Isabel for a moment said nothing, then: "Perhaps I shall find another," she rejoined.

"You will certainly never find a good one."

"In default of a better, my having promised will do," Isabel suggested.

"Yes; that's why I hate it." (James 1881/1997, 520).

This is a discussion about reasons: one party pushes another to present reasons in favour of her action. Miss Stackpole criticizes and evaluates Isabel's idea that she shall go back to Italy. She does this by calling Isabel's attention to the fact that she has no good reason to travel back there. So, her 'argument' silently implies, why should she do it? But then Isabel finds the proper response. Perhaps her reason is not the best one (Miss Stackpole thinks it is not even a good one), but it is good enough for her. The message is clear. The social embeddedness of human action gives rise to a pressure to justify our actions to other participants of normative discussion. (Mead 1934, 335) While the previous consideration was internal, the present one is external, but the conclusion is the same: there is sufficient ground to think that people often take their reasons to be normative.

### **1.1.3. Motivating versus normative reasons**

I started the section with the claim that practical reason is inherently normative. Although the agent's reasons do have an air of normativity around them, they are not yet normative. This leaves room for considerations whose primary aim is to justify and not to explain action and which can properly be called normative reasons. They appear in two reasons-sentences:

- (III) There is a reason for  $A$  to  $\varphi$ , namely that  $q$ .
- (IV)  $A$  has a reason for  $\varphi$ -ing, namely that  $q$ .

It is these sentence types that appeal to an application of ‘reason’ whose purpose is exclusively normative. Although there may be differences between the two formulas, since they are both normative, here we can use them interchangeably.<sup>4</sup>

Not everyone swallows the distinction between the agent’s reasons and normative reasons so easily, however. Some claim that we can explain away this appearance once we clarify the ordinary use of the relevant terms. According to Dennis Stampe: “There is of course the fact that we say that a person “has no reason whatever” for an action when we may mean merely “no good reason”; a person who does something just because he wants to may, indeed, have no good reason, even though he does, to speak literally, have a reason. The statement that something is a reason thus has an associated force of commending it as a basis for action, etc.; but that is not part of the statement’s meaning, for such suggestions can readily be cancelled with consistency: “Don’t do it. Granted you have a reason to do it, but it’s not a good one.” (Stampe 1987, 345-6) As Rüdiger Bittner puts it, the alleged difference between the agent’s reasons and normative reasons is based on a ‘verbal disagreement’ about how to use the phrase ‘has a reason to act.’ (Bittner 2001, 122)

But I think most of us would have a different reaction to Stampe’s example. (cf. Schueler 1995, 67) We would say that the use of the phrase ‘has a reason to act’ actually suggests a normative reading. We would point out that since the agent clearly hasn’t done the action in question, we couldn’t be using an explanatory sense of ‘reason’. On this basis we would then go on to claim that the agent must have at least *some* reason to do the action, but *on the whole* he is

---

<sup>4</sup> In particular, it is widely accepted that formula (IV) articulates reasons that the agent can come to know simply by attention to *p* without any serious effort. The reasons formula (III) refers to, on the other hand, may be taken to be reasons regardless of what the agent is able to discover. For the distinction see Woods (1972), pp. 190; Audi (1986), pp. 513; Skorupski (2002), pp. 114-15. The distinction also resembles to Gibbard’s distinction between available and potential reasons in (1990), pp. 161-2 as well as to Richard Brandt’s distinction between subjective and objective rationality in Brandt (1979), pp. 72-3. But here much turns on their understanding of rationality and reasons, as I will explain later in the text. In any case, though I think this distinction is important, I side with Woods in thinking that it should not be given a merely verbal representation. This explains my choice not to mention it in the text.

wrong: the balance of reasons does not support his choice. And I think we would be tempted to give this sort of explanation exactly because we typically find the difference between the agent's reasons and normative reasons intuitively appealing. This is the old story again. Although the agent's reasons appear to be normative, they are not yet normative. Those who hold the opposite must claim that whenever one acts for a reason one also has a reason to act. But most of us find this claim absurd: people do many silly, inconsiderate, or compulsive things; surely we don't want to say that they nonetheless have a reason to act. In light of this apparent absurdity, we require more than a verbal explanation to persuade us of the truth of this claim; but neither Stampe nor Bittner provides any such thing.

Bittner, however, has another suggestion to make. He denies that normative reasons are inherently normative. If they aren't normative, there is obviously no real difference between them and the agent's reasons. "Contrary to initial appearances", he claims, "our ordinary ways of speaking and thinking of reasons to do something do not support that idea [that reasons are normative], either." (Bittner *ibid.* 135) But the consideration he refers to are indecisive. He says that one who suffers a great loss and asks 'What shall I do now?' does not mean by this 'What is it that I ought to do now?' "Such a person", he claims, "simply wants to be told a course of action for her to follow, and [...] a bare imperative or the indication of a reason for doing this or that would be a proper answer." (*Ibid.* 143) But what is a 'bare imperative' if not an act of guidance? And what is guidance if not a normative notion? At the same place, Bittner also denies that we blame people for failing to do what they have reason to do. He claims that often we don't blame people on this ground; and that only some of us blame people on this ground anyway. (*Ibid.* 143-4) But these points prove nothing. For reasons to be normative it is enough if only some of us blame people on grounds of reasons and do so in only some cases. How could Bittner deny this?

Other attempts in the literature, however, go deeper than these two. The idea is to show that normative reasons and the agent's reasons belong to the same *ontological* category. This claim takes two forms. The 'classical' version holds that the agent's reasons are mental states and that these mental states also serve as normative reasons. (Davidson 1980a; Stampe 1987; Hurley 2001; Gert 1998, 64-6) The 'radical' version denies this, arguing instead that the agent's reasons and normative reasons are the same ontologically because they are both states of affairs. (Dancy 2000b, Chapter 5-7; Bittner 2001, Chapters 4, 7) In my treatment I will only examine the first attempt. I do so for two reasons. The first is that the second version articulates a particular and highly contested theory of motivation, which I cannot evaluate here properly given that my primary subject is not motivation in this chapter. The other reason is that though the second version puts normative and motivating reasons in the same ontological category, it does not eliminate all the difference between them. Normative reasons necessarily do something what motivating reasons need not do (if they do, they become normative themselves): they favour action. Let us turn to the investigation of this idea now.

## ***II. Normative reasons***

### **1.2.1. The structure of normative reason**

In the literature normative practical reason (from now on 'reason', unless qualifier is needed) is typically pictured as a consideration that speaks (alternatively: counts or weighs) in favour of action. (Raz 1975, 186; Gibbard 1990, 163; Scanlon 1998, 17; Broome 2004, 37; Dancy 2004, Chapter 1) It is then further added that reason understood in this way is a *pro tanto* consideration because there is always a certain weight we can attach to it and which then determines its contribution to what the agent ought to do in the given situation. (Dancy *ibid.* 15; Broome *ibid.* 37) This weight need not be as precise as number but may be an entity of some vaguer sort, and

the function employed in aggregation need not be additive either. Reasons, for instance, can behave as organic unities ensuring that their aggregate weight is bigger than the total sum of their weights. But they can also behave as rats: though being on the same side, they can turn and fight against each other. (Dancy *ibid.*) Although, arguably, reasons can do other things than weigh against each other (even if we understand weighing as flexibly as above), the Model is clearly built on this particular picture of reason.<sup>5</sup> Hence from now on I will understand weighing as an essential characteristic of reasons.<sup>6</sup>

On this view reasons have a complex structure. There is first the thing for which we give reasons, i.e. the thing that is favoured: action, belief or emotion. We are here solely interested in practical reasons. Although in Chapter III we will deal with reasons for desires, even then our focus will be on those reasons that can convert into reasons for action. Then, second, there are the considerations that can be reasons, i.e. the things that do the favouring: mental states, states of affairs or propositions. In my discussion I will set aside the question whether reasons are propositions; I say more about this in Appendix I. Finally, we have the most important element of the structure: the favouring relation. It is this relation that connects the considerations that can be

---

<sup>5</sup> The alternative behavior of reasons is noted both by Raz and Scanlon. See Raz's notion of exclusionary reasons and of canceling in Raz (1975), pp. 27, 73-6; 185-6, and Scanlon's notion of silencing in Scanlon (1998), pp. 51-3. Also, Broome claims that what we have overall reason to do is not the same as what we have most reason to do. That is, to use his terminology, not every overall 'ought' can be given a weighing explanation. See Broome (1999), pp. 403; (2004), pp. 43, 45. Dancy (2004), Chapter 1, on the other hand, rejects the claim showing, to my mind persuasively, that overall oughts can always be explained in terms of *pro tanto* reasons. Finally, Skorupski (2006), pp. 28 and (ms) propose the notion of sufficient reason (perhaps Dancy's (2004), pp. 16, 43 notion of 'decisive reason' articulates the same idea), which again is supposed to be different from the weighing centered notion of *pro tanto* reason. But Skorupski also notes that in the case of practical reasons, the sufficient-reason relation can be accounted for in terms of the reason-relation.

<sup>6</sup> I don't think this claim needs much backing, but here is a very schematic explanation. The default explanation the Model offers of normative claims is a weighing explanation: reasons are based on desires whose strength determines their stringency and it is these reasons that we weigh against each other when determining how to act. I call this a 'default explanation' because I don't want to rule out the possibility – as some do, see e.g. Scanlon (1998), pp. 51-3 – that on the Model reasons can only behave in one way, no alternative behavior is allowed. Chuard and Friedrich's (ms) complex understanding of defeaters may, for instance, offer a solution to this problem, a solution that is compatible with the Model. And Hubin's (2003) version of the Model also claims to accommodate the function of silencing. But these are controversial ideas, whereas the claim that the Model is articulated in terms of *pro tanto* reasons is uncontroversial because it is based on a natural reading of the Model. Hence follows my choice in the text.



reasons to the action and thereby endows them with the status of a normative reason. It is also this element that determines the weight of a reason: depending on how strong the relation is, the consideration that purports to be a reason will have a varying contribution to the normative assessment of the case at hand. To put it formally, reasons have the following structure:  $\mathbf{R}(q, \varphi)$ , where  $q$  is the thing that purports to be a reason,  $\varphi$  is the action for which  $q$  purports to be a reason, and  $R$  is the favouring relation that connects  $q$  to  $\varphi$ -ing.<sup>7</sup>

### 1.2.2. Mental states as normative reasons: desires

Turn now to the question whether reasons are mental states. Our first candidate is desire. The ‘classical’ argument comes from Donald Davidson. It goes like this. The claim under investigation holds that motivating reasons are also normative reasons. If we accept the Humean Theory of Motivation (HTM), we should have no trouble with the first part of the claim: motivating reasons consist in a pair of desire and suitable instrumental belief and this is all what we need now. For the sake of discussion I assume that the HTM is true; I will express my doubts about the theory in Chapter IV. The question now is what we can do with the second part of the claim. Davidson’s idea is to ‘subjectivize’ the favouring relation by picturing it as an inferential relation among the *contents* of certain mental states including desire.<sup>8</sup> The first step is to

---

<sup>7</sup> In the literature one can find an important addition to this structure. Dancy (2004), Chapter 3 argues that besides considerations that favour action, there are also *enablers*: considerations in the absence of which another consideration would not favour the action. He also argues that there are *intensifiers*: considerations that strengthen the relation between the favouring consideration and action. For Dancy these further factors are neither part of the favouring relation, nor are they things that can be reasons. On another reading, however, these considerations together with the *pro tanto* reason constitute *one* single reason, one complex consideration that speaks in favour of action. They all form part of the thing that purports to be a reason: they constitute a ‘perfect’ (or ‘ultimate’, ‘complex’, ‘whole’) reason. See Broome (2004), pp. 34; Crisp (2000), pp. 37; Raz (1999b), pp. 228n; Bennett (1995), pp. 80. Although deciding this question has a bearing on the debate between particularists and universalists (see Dancy (2004) and McKeever and Ridge (2005) on this), my discussion needs no such settling. Hence I set the problem aside in the text.

<sup>8</sup> Admittedly, there is plenty of ambiguity about Davidson’s own position. For the possible readings and a treatment of at least some of them see Velleman (2000a), esp. pp. 104-5 as well as notes 15, 18, 20. In the text I discuss what appears to be the most widespread interpretation of Davidson’s views in the literature.

understand desire not as a favourable attitude toward the representation of some outcome, but rather as an attitude toward a favourable representation of some outcome. In other words, desires are propositional attitudes whose content is given by an evaluative proposition such as ‘having the object of this desire would be good’. (Davidson 1980d, 86)

Once we see desires in this way, a ‘subjectivist’ account of their normativity follows. Take Davidson’s example about improving the taste of the stew. (Ibid.) Ordinarily, we identify the content of the agent’s desire with the proposition “that I improve the taste of the stew”. But if we accept Davidson’s account of desires, we should rather identify the content of the desire with the proposition “that it would be good to improve the taste of the stew”. Then we add the belief of the agent with the propositional content “Adding the sage to the stew will improve the taste of the stew”. And on this basis we conclude “Adding the sage to the stew would be good”, which serves as the propositional content of a desire of the agent. Hence the agent’s transition from his desire to improve the taste of the stew to his desire to add sage to the stew is dictated, in the fashion of an inference, by a privileged logical relation between the contents of the attitudes involved. Desires favour action through their propositional content inferentially supporting the action-focused propositional content of a further desire. The favouring relation is inseparable from desires: whenever desires are present, the relation is present as well.

This position has the virtue of simplicity: it offers a candidate both for  $q$ , which it takes to be a desire (together with an instrumental belief), as well as  $R$ , which it pictures as a relation among the contents of mental states. It also bleeds from many wounds, however. Here is what I, following the lead of others, take to be the most obvious and most serious problem with Davidson’s position.<sup>9</sup> There are well-known cases when one desires something without making

---

<sup>9</sup> Further problems stem from the fact that if we accept Davidson’s idea, we get close to a picture of desire as a mental state very much akin to an evaluative belief with the propositional content ‘having this object would be good’.

Attila Tanyi, THESIS, Page 26

an evaluative judgement about it. (Smith 1994, 140-1; Stocker 1975) Take the classical example of compulsive behavior from Judith Rapoport. (Rapoport 1989) The boy who can't stop washing his hands certainly has a desire to wash his hands; yet, he doesn't see anything good about acting in this way. And there are many such examples in the literature on urges and cravings. In Warren Quinn's case a man has an urge to turn on the radio whenever he sees one, even though he finds nothing attractive about radios or the sound they make or the programs they play. (Quinn 1993, 236) And Gerry Watson, A. J. Ayer and Harry Frankfurt describe further cases when we cannot handle the agent's desire in the way that Davidson advises us to do. (Watson 1975, 210; Ayer 1954, 20; Frankfurt 1971, 87)

There are, however, attempts to complete the Davidsonian enterprise. The first comes from Stampe. He tries to remedy Davidson's failure by harboring normative force not in the propositional content of desire, but in the way that proposition is regarded when it is the content of desire rather than belief. The idea is this. The desire of the agent is taken not as an attitude toward the evaluative proposition that *p* would be good but as an attitude toward *p as good*. That is, in desiring something we don't express a value judgement about how things are in the world; instead, we aim to change the world as our desire requires, expressing this with the predicate "as good". (Ibid. 355) According to the so-called dispositional or functionalist theory, beliefs and desires differ from each other in their 'direction of fit'. One who has a belief is disposed to accept a proposition that is patterned after the world (downward direction of fit), whereas one who has a desire is disposed to accept a proposition as a pattern for the world to follow (upward direction of fit). Stampe's account holds that regarding something as good amounts to an attitude with an

---

good'. At least this is what Davidson's type-type identification of evaluative belief and desire seems to suggest. There are two problems with this proposal (for a possible third that I am not sure about, see Stampe 1987, 339, 375-380). The first is Stampe's problem that later I also mention in the text, the second is Velleman's problem that if the content of desire is given by an evaluative proposition, this implies that the capacity to desire requires the possession of evaluative concepts. But young children, for instance, do not possess such concepts, yet they do have desires. See Velleman (2000a), pp. 104.

upward direction of fit. And it is this direction of fit as expressed by the evaluative predicate “as good” through which desires favour action. The favouring relation is again inseparable from desires: whenever desires are present, the relation is present as well.

Stampe’s theory shares the simplicity of Davidson’s approach: we get a candidate both for  $q$  (desire) and  $R$  (direction of fit). Moreover, since Stampe does not identify the content of desire with an evaluative proposition, his account avoids the problems arising from such a move.<sup>10</sup> Even when someone desires something that he does not take to be good, it is still the case that he regards that thing as good, that is, as to be brought about. Where Stampe’s account stumbles is its claim that normative force can be explained in terms of desire’s direction of fit. We can see this if we invoke an analogy with beliefs. We owe this point to David Velleman. (Velleman 2000a, 110-119)

Reference to direction of fit, Velleman claims, might be enough to distinguish beliefs from desires, but it is not enough to distinguish them from other cognitive attitudes. Take the case of hypothesizing. This attitude has the same downward direction of fit as belief: one who is in this state is disposed to accept a proposition about certain hypothetical facts. Yet hypothesizing is also different from belief. Where it differs is its criterion of success or, to use Velleman’s term, its ‘constitutive aim’. When one hypothesizes, one is disposed to accept a proposition only for certain polemical purposes, i.e. for the sake of the argument. In contrast, having belief amounts to having a disposition to accept a proposition with the aim of finding out what really is true. Cognitive attitudes, Velleman sums up, are two-tier attitudes where the first-order attitude of accepting something as true is coupled with different second-order attitudes, which in turn

---

<sup>10</sup> It is not clear though how this reading of Stampe fares with the rest of what he says. For he takes desires to be perceptions and holds that it this perception that ultimately grounds the ‘authority’ of desire. See Stampe (1987), esp. sections 4-6. And though it is also true that the perception Stampe has in mind is a special kind of perception, he calls it ‘proprioception’, this move might nevertheless cause trouble for his theory by taking it dangerously close to a position like Davidson’s. In the text I do not take up these difficulties because we are interested in that bit of Stampe’s theory that might offer an answer to the problem that beset Davidson’s conception.

depend on what the constitutive aim of the given attitude is. And, he concludes, beliefs, unlike hypothesizing or fantasizing, can play a normative role exactly because they are attitudes whose constitutive aim is to track what really is true.

It is not obvious, however, why this tracking role would turn a mental state into a reason. As Stampe points out, such mental states would be best understood as reasons *per objectum*: they would not be reasons in their own right, but would only comprise reasons *through* their objects, namely through the facts they represent. (Stampe 1987, 337-8, 342-44) That is, were desires only to track evaluative facts in the world, it would be these facts and not the mental states that provide us with reasons to act. But even if we set aside this problem, this would still not help Stampe's case. For desires are clearly not mental states that have the constitutive aim of bringing about what really is good. No doubt, many conative states fail to meet this criterion. Think of wishes and hopes: though they have an upward direction of fit, they have no constitutive aim of tracking the good. Yet, desire does not come in the niche created by these failures. Observe that holding this view would take us very close to Davidson's account with all the counter-examples mentioned there. As admitted, these counter-examples fit the characterization of desires as attitudes with an upward direction of fit; yet, they are certainly not instances of a two-tier attitude that has the constitutive aim of tracking the good. Hence, if we accept Velleman's logic, it follows that desires cannot in the Stampean way be normative - and this is all what we had to prove here.<sup>11</sup>

But, perhaps, the analogy with beliefs can be carried further to produce a viable account of the idea that desires are reasons. Recently, Paul Hurley has appealed to the analogy in defense

---

<sup>11</sup> This, of course, brings us to the question whether desires differ from other conative attitudes, and if they do what that difference consists in. Velleman's candidate is *attainability*: desire differs from other conative attitudes, such as wishing or hoping, in that it is only directed at what is attainable, i.e. what it is possible to get or pursue. See Velleman (2000a), pp. 116-7; cf. Mele (2003), pp. 135-6 who makes a similar proposal. I don't want to evaluate Velleman's proposal here for I am perfectly happy with the negative point made in the text.

of a Davidsonian (but not Stampean) model.<sup>12</sup> (Hurley 2001, 11-5) In epistemology, Hurley points out, we make a distinction between ‘seeming’ reports and perceptual beliefs. The basis of this distinction is that while the latter involve commitments to something’s being the case, the former do not: they are *resisted* inclinations to express such commitments. Our experiences of the world flow by default into beliefs, but sometimes this process gets to a halt and ‘seemings’ report this: they signal the withholding of the commitment that flows by default into beliefs. The same happens, Hurley argues, in the practical sphere. Urges, cravings, compulsions are all *failed attempts*; they all signal the withholding of normative commitments that otherwise flow by default into desires. Hence the problem that beset both Davidson’s and Stampe’s theory is tackled without compromising the original Davidsonian idea.

I am, however, doubtful whether urges and cravings are indeed the practical analogues of seemings. Hurley describes perceptual seemings as being a sort of halfway house towards proper perceptual beliefs. He brings as his example the half-submerged stick that looks bent. Here we can indeed claim that the agent was inclined to give in to his perceptual illusion but then, in the light of counter veiling rational considerations, he resisted his inclination. There are, no doubt, such seemings in the practical sphere too. Many of our desires are like Hurley’s stick: they have an object that *appears* to be good but then the agent refuses to give in to the temptation they pose. (Scanlon 1998, 65) But urges and cravings are not like that. Their whole point is that the agent, no matter how hard he tries, cannot see anything good about the object of his desire. It is not that

---

<sup>12</sup> If one reads Hurley it might not immediately be clear to him that Hurley is concerned with proving a Davidsonian position. His article is essentially negative in its argumentation: its aim is to prove that a certain position does not hold. This position, he attributes it to Christine Korsgaard and Barbara Herman, claims that desires properly understood are pre-reflective, pre-conceptual and, most importantly, pre-rational states, which, being what they are, cannot be normative. Against this position, his claim I believe stands well; I say more about this in Appendix III. But Hurley also makes it clear that he intends the negative argument to turn into a positive one, i.e. into an argument for the position that desires are reasons. See what he says in Hurley (2001), pp. 11, 13, esp. note 37.

the object tempts him, appearing to him good first, but then he resists the temptation. It is that the object has never tempted him: it is not a failed attempt; it is no attempt at all.

### 1.2.2. Mental states as normative reasons: beliefs

If desires are not reasons, we still have beliefs to serve in this role. Here is the story.<sup>13</sup> It is often claimed that there are two kinds of ‘ought’: the ‘ought’ of rationality and the ‘ought’ of reason. (Parfit 1997, 99; 2001; Scanlon 1998, 22-32; Dancy 2000b, Chapter 3) The latter relates the agent’s attitudes to features of his situation, whereas the former relates the agent’s attitudes to other attitudes of his. The ‘ought’ of rationality, moreover, is an ‘ought’ of consistency; in our case it holds that the agent’s action must be consistent with his beliefs about what is the case and what he ought to do on this basis.<sup>14</sup> This, however, does not itself lead us to the claim that beliefs are reasons. For that we need the further idea that in reality there is only *one* kind of ‘ought’: the ‘ought’ of rationality. This, however, requires explanation, since we now have to account for the normativity of reason in terms of rationality. Depending on how we do this, we get different accounts of the reason-relation. There are two options. Either we equate the reason-relation with the rationality-relation, or we understand reason in terms of rationality. In the former case we make a proposal about *q*, taking it to be a belief and about *R*, picturing it as a consistency

---

<sup>13</sup> Two alternative stories are conceivable. The first uses Velleman’s idea that if beliefs are normative this is because they have the constitutive aim of tracking the truth. The second copies the original Davidsonian idea and applies it to the case of beliefs. In the former case beliefs figure in *q*, in the latter case they appear both in *q* as well as in *R*. Both approaches, however, stumble on the Stampean claim, referred to earlier in the text, that beliefs, exactly because of their tracking role, are not reasons *per se* but only reasons *per objectum*. And the Davidsonian line would also have trouble with Velleman’s objection mentioned in note 9 above.

<sup>14</sup> It might be strange that while I describe the ‘ought’ of rationality as a consistency requirement that holds among the agent’s attitudes, in the particular case what stands on the right-hand side of the relation is an action, not an attitude. But there is nothing mysterious presupposed here. When I speak of intention I have in mind what Searle (1983) calls ‘intention-in-action’, an intention, which is part of action: action is nothing else but intention plus bodily movement. One can compare this to the volitional theory of action as described in Chapter IV that regards volition as part of action: it is where action begins. See also Searle (2001) about this and related issues.

relation. In the latter case,  $q$  remains a fact, whereas beliefs appear in  $R$ , which is still understood as relating facts to action.

However, neither reading gives us what we want: that motivating and normative reasons are identical. Besides skepticism about the motivating potential of beliefs, more trouble comes from the normative side. The second reading rules out the idea that beliefs *are* reasons since it takes beliefs to play a role in the reason-relation only. The first reading, on the other hand, allows for the former role but not for the latter; hence it is in a situation similar to the ‘radical’ approach I mentioned at the end of the previous section. I think, moreover, that the first reading is wrong. If we accept it, we don’t really explain one relation in terms of the other; we simply eliminate the reason-relation. Yet, because giving beliefs any role in our account of normative reason does pull motivating and normative reasons closer to each other and because the framework in which this issue appears will be important for us later, I discuss this idea in some detail here.<sup>15</sup> The claim that there aren’t two kinds of ‘oughts’ because the ‘ought’ of reason is an illusion is questionable. In particular, there are two alternative positions. One holds that both relations exist and are independent of each other; the other argues that it is the rationality-relation, which is the illusion, not the other way round.

The first alternative, however, is implausible. (Dancy 2000b, 51, 61; Kolodny 2005, 512, 556) What we are interested in practical matters is what the agent should do. True, from the

---

<sup>15</sup> This ‘pulling’ effect can be that the explanation of the agent’s action takes the form of justification or, as it is often called, ‘rationalization’. See Dancy (2000b), 8-10. This suggests an alternative interpretation of Davidson’s position since he also understands action-explanation as rationalization. In his view the governing rule of this enterprise is the principle - often called the principle of charity - that agents behave in ways that they ought to behave given their beliefs, desires, judgments and the like. We thus identify the agent’s mental states in a normative structure: we assume that the agent believes what he ought to believe and desires what he ought to desire given his action. Hence the agent’s mental states are normative even if they are desires that involve no evaluative commitment or beliefs that are false or for some reason crazy. See e.g. Davidson (1980a), pp. 9; the argument enjoys considerable support, see Woods (1972), pp. 189; Blackburn (1998), pp. 51-9; Mele (2003), pp. 74-6. We should note, however, that Davidson goes beyond this point and moves closer to the account we attributed to him earlier when he demands that we see the agent as the “believer of truth and the lover of good” in addition to being coherent. See Davidson (1980e), pp. 222.



agent's point of view, there is only one answer to the question 'What should I do?', since from his point of view the two relations necessarily coincide (because, though he can see that there are these two relations, he cannot, presumably, distinguish them in practice, since he cannot distinguish in practice how the word is and how he takes it to be). But from *our* point of view, i.e. from the point of view of the observer there are two answers to the question 'What should he do?' depending on which relations we have in mind. The two obviously do not coincide since the agent might hold false, though rational beliefs, and he might also be ignorant of certain relevant features of his situation. But if there are indeed two answers to this question, we, the observers are in trouble since we have to decide, which answer to accept. Since, however, both relations are taken to be autonomous, there seems to be no way to make this decision. And this is an unacceptable result for the question clearly needs an answer: without an answer we cannot guide the agent's conduct, cannot recommend or advise him to do anything.

Therefore we should turn to the second alternative. There are two ways of making this claim. We might try to say that the rationality-relation is an illusion because it is simply not normative. Again, there are variations here. We can follow Dancy in claiming that the 'ought' of rationality is evaluative but not deontic: that when we make claims about the agent's rationality, we only assess his action but do not specify how he should have acted. (Dancy *ibid.* 53, 62) What we say is that he acted *well* because he did what he would have had reason to do had his (rational) beliefs been true. After all, we can explain, duties of rationality are nothing else but those potential duties of reason that we would have had had our (rational) beliefs been true. (*Ibid.* 51) Alternatively, we might try to say that the 'ought' of rationality employs an impoverished notion of normativity. It pictures normativity as perfect coherence, and it is questionable if mere coherence produces normativity. (Gibbard 1990, 156-7; Dreier 2002, 140; Stratton-Lake 2000, 122) For the normativity we appeal to here is like when we say that if the circuit is wired up then,

when the key is turned, there ought to be a spark. (Blackburn 1998, 56) But this ‘ought’ is the ‘ought’ of *expectation*: it is about what is *likely* to happen. And it is at least questionable whether such an ‘ought’ is normative at all.

The problem with both objections is that the ‘ought’ of rationality does seem to be normative, even if to a minimal extent and perhaps in addition to being evaluative (though it is not obvious whether Dancy is right about the kind of evaluation that is at issue here, see Kolodny 2005, 551-554 for different proposals). From the point of view of the observer, claims of rationality seem to be in the register of *advice*. (Ibid. 554) They take the form “*Given* what you believe, you *ought* to believe or do this and that”, and claims like this don’t seem to grade the addressee (which, in advance of the addressee’s response would be impossible anyway), but rather recommend courses of action to him. Nor does the agent experience the demands of rationality as non-normative: *of* the response that rationality requires of him, the agent thinks that he ought to give it. (Ibid. 555) This, again, is not a merely evaluative thought. Recall our previous point: within the agent’s own perspective the two relations coincide. Hence the agent’s experience of rational requirements is in important respects the same as his experience of the requirements of reason. Both make demands from within his perspective and both are unsuitable as grounds of evaluation, being applicable only in advance of his response.<sup>16</sup>

This suggests that we should turn to the second main approach to refute the idea that the ‘ought’ of rationality is the only ‘ought’ around. It accepts that rationality is normative but it wants to show that this normativity is derived from the normativity of reason. Again, there are two ways to do this. The first holds that rationality is normative because there is reason to be

---

<sup>16</sup> One might ask why it is not enough just to point out that this is so. After all, this shows that, within the agent’s perspective, consistency requirements of rationality function on a par with the substantive standards of reason, appearing just as normative as those are. But recall what is at issue here; it is not only whether rational requirements are normative, but also where that normativity comes from. If we simply accept the claim in the text, then we will at best endorse idea that there are two normative relations. And this is exactly what we don’t want to accept for the reason outlined previously.

rational. (Broome 1999, 2004, 2005; Dancy 2000b, Chapter 3; forthcoming b) The idea is to reinterpret the reason-relation so that it encompasses rational requirements by putting the complexes they govern on the right-hand side of the relation, while keeping facts on the left-hand side. The problem with this approach is that it can lead to incredible ‘bootstrapping’, i.e. to the creation of reasons that no one would like to accept as a reason.<sup>17</sup> (Bratman 1987, 24-7; Scanlon 2004; Broome 1999, 403; 2004, 30) Take as an example the claim “If you believe that  $p$ , you are rationally required to believe what  $p$  entails.” If we suppose that you have reason to be rational, it follows that if you believe that  $p$ , you have reason to believe what  $p$  entails. And since  $p$  entails  $p$ , it follows that if you believe that  $p$ , you have reason to believe that  $p$ . That is, the fact that  $p$  entails  $p$  constitutes a reason for you to believe that  $p$ , if you believe that  $p$ . But self-justifying beliefs are hardly acceptable for anyone.

There is, however, a well-known way out of this difficulty. We just have to notice that the consistency requirements rationality makes on us are different from the requirements of reason. The former have wide-scope, whereas the latter have narrow-scope, to use Broome’s terms. (Broome 1999, 405; 2004, 30; Dancy 2000b, 53-4, 62-3) That is, in the above example the correct formulation of the requirement is ‘You ought, if you believe that  $p$ , to believe what  $p$  entails’, and not ‘If you believe that  $p$ , you ought to believe what  $p$  entails.’ Hence the ‘ought’ we are concerned with in the case of rationality is an ‘ought’ on a combination, a disjunctive obligation that can accordingly be satisfied in two ways. The agent ought either to believe that  $p$  and then believe what  $p$  entails, or not believe what  $p$  entails but then give up his belief that  $p$ . It is these combinations of attitudes that stand on the right-hand side of the favouring relation. And,

---

<sup>17</sup> There are other problems with the idea. It is not obvious what justification, i.e. reasons we can find for the different principles of rationality. Nor does everyone accept it that we reason from the requirements of rationality, a condition many accept as needed for a consideration to qualify as a reason. See Kolodny (2005), sections 3 and 4 for these claims. Broome (2005) takes up the first issue but does not give a reassuring response even by his own lights. See also Dancy (forthcoming b) who shares Broome’s conclusion. In Appendix II I say more about the justification of the instrumental principle and its role in reasoning, but I do so in a context different from the present one.

the further claim goes, wide-scope requirements differ from narrow-scope requirements in another respect as well: they don't allow for detachment. (Broome *ibid.*; Dancy *ibid.*) That is, from a complex obligation like the one above one cannot detach a simple obligation to believe what  $p$  entails, even under the condition that the agent believes that  $p$ . As a result, we cannot say that the agent ought to believe what  $p$  entails if he believes that  $p$ . Hence no bootstrapping, no creation of unacceptable reasons follows.

This argument is far from being unanimously accepted, however. Rejection comes from two sides. On the one hand, one can deny that the complex obligation the argument refers to gives us the right form of rational requirements. On the other hand, one can accept that rational requirements have wide-scope, but deny that there is indeed a ban on detachment. Start with the latter claim. There are two ways to put this point. One is that the conclusion we arrive at through detaching is not so far from the conclusion we reach by denying detachment. There is not much difference between the claim that one ought, if one believes that  $p$ , to believe what  $p$  entails, and the claim that if one believes that  $p$ , one ought to believe what  $p$  entails. The alternative is to draw attention to the fact that there are cases when we do allow for detachment. To take Dancy's example, from the disjunctive obligation 'you ought, if unavoidably delayed, to ring up and warn your host if you can', we do detach the conclusion that 'you ought to ring up and warn your host if you can'. (Dancy *ibid.* 71) So the advocate of non-detachability must explain why we should allow for detachable obligation in one case and reject it in another.

Dancy gives the following response.<sup>18</sup> (Dancy *ibid.* 73-6) He says that we should not allow detachment in any case, not even in those cases where it sounds plausible, exactly because

---

<sup>18</sup> Broome has a different solution but it doesn't work. See Broome (1999), p. 404-5. The claim we want to get is that if the agent believes that  $p$ , he ought to believe what  $p$  entails. The 'ought' involved, Broome points out, cannot be an overall ought. It cannot be that when the agent believes that  $p$ , he should *always* believe what  $p$  entails: what if he ought not to believe that  $p$ ? Insofar as the overall ought is concerned we should not allow for detachment because it

the difference between detaching and non-detaching is not insignificant. If one accepts detachability, he claims, one also has to accept absurd inferences that come with it. Take the following inference pattern ('Ox' means 'x ought to ensure that' and 'Dx' stands for 'x performs action D'):  $Ox (p \rightarrow Dx)$ ; by contraposition we get  $Ox (-Dx \rightarrow \neg p)$ ; and by detaching we arrive at  $(-Dx \rightarrow Ox \neg p)$ . Let us apply this pattern to the claim, "You ought, if there is someone before you in trouble, to give a helping hand." We get, "If you do not give a helping hand, you ought to ensure that there is not someone before you in trouble", which is obviously absurd. It is questionable whether we should employ a principle of inference that is capable of taking us from truth to non-sense so easily. The difference, i.e. the cost of choosing between detachment and non-detachment is high; hence on the whole it seems best to reject detachment *tout court*, not just in certain cases when it leads to absurd results.

Turn now to the first objection. Niko Kolodny has recently produced an argument that aims to show that at least some rational requirements are narrow-scope. The argument has two steps. Kolodny first claims that at least some rational requirements are what he calls process-requirements as opposed to being state-requirements. (Ibid. 516-8) The difference is that whereas state-requirements only ban situations in which the agent has conflicting states, such as when one believes that *p* but does not believe what *p* entails, process requirements also say *how* one is to resolve the conflict. State-requirements, moreover, have wide-scope since there is more than one

---

can easily lead us from truth to falsehood. But this still doesn't rule out the claim that believing that *p* may give the agent *some* reason to believe what *p* entails. Perhaps he ought not to have this belief overall and thus ought not to believe what *p* entails overall; yet, he can still have *some* reason to have the belief, thus some reason to believe what *p* entails. See Dancy (2000b), pp. 73 for this point. True, the situation would be different if, as Broome (1999) and (2004) held, rational requirements would be *strict*. That is, if when one believes that *p* but does not believe what *p* entails, he would be not entirely as he ought to be; if he would be failing in some respect. But it is not obvious why rational requirements should be strict (and not slack, to use another term of Broome). In other words, it is not obvious why our reason to be rational cannot merely be *pro tanto* as opposed to overall. As Broome (2005), pp. 324-5 and (ms) now admits we might only have a *pro tanto* reason to be rational because, on certain views at least, rational requirements can conflict with other moral, prudential etc. requirements. This seems to be, see the above reference, Dancy's own position too. Note also that this reading does not rule out the bootstrapping objection since though the reason created would only be *pro tanto*, it might be that we don't want any reason at all. Kolodny (2005), pp. 539 note 25 makes this point.

way of satisfying them. (Ibid. 519) This is just the picture we are familiar with: one either believes that  $p$  and then also believes what  $p$  entails, or one does not believe what  $p$  entails but then gives up the belief that  $p$ . Process-requirements, on the other hand, may not be satisfied in more than one way for here it also counts how one resolves the conflict: whether one does it in a rational manner or not. Since for Kolodny rationally resolving a conflict consists in arguing from the content of an attitude to dropping another attitude, we get what he calls the ‘Reasoning Test’. (Ibid. 520-1) It states that process-requirements have wide scope only if one can resolve the conflict they govern by reasoning from the content of one attitude to the dropping of another attitude and *vice versa*.

In the second part of his argument Kolodny goes on to claim that certain rational requirements fail the Test. Take an example that is relevant in the present context, the claim ‘Rationality requires one to intend to  $\varphi$ , if one believes one ought to  $\varphi$ ’. This requirement, Kolodny claims, is a narrow-scope process requirement. The explanation he offers is this. (Ibid. 526-7) There are two potential ways of resolving the conflict-state the requirement governs: one either intends to  $\varphi$ , if one believes one ought to  $\varphi$ , or one does not intend to  $\varphi$ , but then gives up the belief that he ought to  $\varphi$ . However, since the requirement is read as a process-requirement, it also matters whether these two resolutions come about in a rational way. And, Kolodny points out, this is certainly not the case with the second way of resolving the conflict. For here one would have to reason from the content of one’s not intending to  $\varphi$  to the dropping of one’s belief that one ought to  $\varphi$ . But this is impossible: since not intending to  $\varphi$  is simply lacking an attitude, one cannot reason from its ‘content’ to anything. Of course, one can lose the belief that one ought to  $\varphi$ , but this way of resolving the conflict would not be rational. The requirement fails the Reasoning Test.

Kolodny has another argument to prove that other rational requirements are also narrow-scope because they fail the Test. (Ibid. 528-30) But for us there is no need here to go through this argument, nor to see Kolodny's treatment of possible responses to the arguments. For the claim under investigation, i.e. that the normativity of rationality derives from the normative of reasons, is completely general aiming to govern all rational relations among our attitudes. Hence if we can show that one rational requirement does not confirm to the thesis, the thesis is defeated. (Ibid. 540) And this is what we just did. If the above requirement fails the Reasoning Test, it is narrow-scope, and if it is narrow-scope, it invites the bootstrapping objection. The question now is what we can do with Kolodny's argument. One option, of course, is rejection. Some of Kolodny's claims about reasoning might be disputed but I agree with them. His claim that rational requirements are process requirements, however, appears much less plausible to me. It is not that there cannot be process requirements; the question is whether rational requirements are like that. And, as Broome in his response points out, rational requirements as formulated here mention only states and no processes. (Broome ms) So it is not obvious, though it is possible, why we should interpret rational requirements in this way.

Kolodny, however, provides reasons for the process reading. (Ibid. 517) The first is that process-requirements make state-requirements superfluous because by complying with the former, we also satisfy the latter: if one avoids or escapes conflict-states, then one won't be *in* them. But this is only partly true. For process-requirements also specify *how* one can leave conflict-states; they only admit certain ways of satisfying state-requirements. Hence they do not make state-requirements superfluous. Kolodny's second point questions the normativity of state-requirements on the ground that such requirements are not response-guiding: they only say that we should not *be* in a particular state. But this is not precise. For when state-requirements prohibit a particular combination of attitudes, this prohibition implies that the agent should leave

that state and this *is* response-guiding. What state-requirements are silent about is *how* one should leave that state; it is like a general process requirement to leave a conflict-state in any way what one likes. Kolodny admits this. (Ibid.) But then I don't understand his problem. Sure, he can say that rationality consists in specific ways of resolving conflicts. But this would just assume the truth of the position to be proved and thus would beg the question.

If someone is not happy with this reasoning, he could still go along with Kolodny's alternative picture that also derives the normativity of rationality from the normativity of reason. He calls it the Transparency Account. (Ibid. 557-8) The idea is to derive all rational requirements from the core requirements of "If one believes that one ought to have *A*, then one is rationally required to have *A*" and "If one believes that one lacks sufficient reason to have *A*, then one is rationally required not to have *A*." Once this is done, from the observer's point of view we can give advice by pointing out the agent himself thinks he has reason to have *A*: that from his point of view such a reason clearly exists. And since from the agent's point of view there is no difference between the two relations, the reasons, as it seems to the agent that he has, and the reasons that he really has are the same. Such a distinction can only be made from the observer's perspective; from the agent's perspective, on the other hand, the 'ought' of rationality is transparent. The question is whether we can indeed derive all rational requirements from the core requirements. Kolodny, however, does not prove that this is the case, only gestures in that direction. (Ibid. 559-560; Scanlon 1998, 25 and ms pursue the same idea and the latter, according to Kolodny, answers most of the problems)

### ***III. The Model***

The previous two sections gave us a picture of normative reason. We know the structure as given in the formula **R (q, φ)**, and also have some idea about its elements. We know that desires as



mental states cannot figure in  $q$  and  $R$  at the same time. This would be endorsing the Davidson/Stampe/Hurley line, which, I argued, is false. From our discussion of beliefs we also know, moreover, that that the reason-relation either helps us account for the rationality-relation or the reverse is true. It is in this framework that we have to locate the Model. There are three variations. The first version understands the rationality-relation as having desires as mental states on its left-hand side and actions on its right-hand side. It then claims that there is reason to be rational, i.e. it puts the rationality-relation on the right-hand side of  $R$  that is regarded as relating facts to action, and then accepts bootstrapping. The second version claims that desires as constituents of facts serve in the role of  $R$ : the fact that the agent's desire that  $p$  would be served by his  $\varphi$ -ing is what turns another fact, i.e. a feature of his situation, which is related to his desire into a reason for the agent to  $\varphi$ .<sup>19</sup> The third version starts from the same picture of the rationality-relation as does the first and then uses this requirement (and presumably others) to give us an account of  $R$ , which it understands as relating facts to action. Let me now flesh out these positions in more detail.

### 1.3.1. First version

The first position begins with the instrumental principle pictured as a consistency requirement on human action and desire. The principle says the following:

**(IP)** Rationality requires of you that, if you desire that  $p$ , and if you believe that your  $\varphi$ -ing is the most effective means to your desire that  $p$ , then you  $\varphi$ .

---

<sup>19</sup> The facts that stand on the left-hand side of the reason-relation can also be subjective facts, i.e. facts about the agent's mental states. Here are two examples from Dancy (2000b), pp. 55, 65 concerning facts about belief. If I think that everything is worthless and the best is to die, then my belief may give me and, presumably, others a reason to help me to receive psychiatric help. Similarly, if a mother believes that his son has died on the sea, we have reason to support her, to respect her feelings and so on. Beliefs as cited here are clearly features of the situation that favour action.

It then claims that there is reason to be instrumentally rational, that is, it puts the complex requirement the instrumental principle articulates on the right-hand side of the reason-relation. Then it accepts what seems inevitably to follow, namely bootstrapping. The reasoning is analogous to the one we encountered in the case of beliefs. Start from the claim that if you desire that  $p$  and believe that  $\varphi$ -ing is the best way to satisfy your desire that  $p$ , then rationality requires you to  $\varphi$ . If we suppose that you have reason to be rational, it follows that if you desire that  $p$  and believe that  $\varphi$ -ing is the best means to satisfy your desire that  $p$ , you have reason to  $\varphi$ . Fully spelled out, the conclusion is this: the fact that  $\varphi$ -ing subserves your desire that  $p$  constitutes a reason for you to  $\varphi$ , if you desire that  $p$  and believe that  $\varphi$ -ing is the best means to satisfy your desire that  $p$ .

According to Dancy, this is just the thesis that stands behind the Model. He calls it the Advice Point, and illustrates it with an example. (Dancy 2000b, 34) If my friend wants to insulate his house so that no sound whatsoever can penetrate it, I might find his desire silly, yet give him advice. I can say that though what he wants to do is stupid and he should not do it, *once* he has set his mind to do it, he should, i.e. he has reason to do it this way rather than the other. In Dancy's view the Model needs no more than this claim and this claim, to repeat, articulates nothing more than a consistency requirement on desires/ends and action/means, now given in terms of reasons. We should note, though, that the viability of this requirement comes from those as yet uncovered reasons that tell us to be instrumentally rational. I, however, won't look for these reasons but assume, throughout this dissertation, that they can be found.

### 1.3.2. Second version

On the second version of the Model desires, more precisely, facts about desires appear in the reason-relation. With some modifications, which I will explain in a moment, we can take over the following definition from Dancy (2000b, 28):

**(DBR)** If its being the case that  $q$  is a reason for  $A$  to  $\varphi$ , this is because, there is some  $e$  such that  $A$  actually or hypothetically desires that  $p$  and given that  $q$ ,  $\varphi$ -ing subserves the prospect of  $p$ 's being realized (or continuing to be realized).

The philosophical rationale for this version of the Model is the following. (Hubin 1999, 32-3; 2001, 459, 467) The account decomposes the Model into two parts: one claims that it is our desires that are the sources of our reasons; the other transmits these reasons to the means. That is, the instrumental principle has no normativity-creating role: it merely transmits the normativity created by desires. Desires are taken to be 'brute facts' that sit at the end of the normative chain: the normativity of actions *just is* their being properly related to the agent's desires.

This version is clearly different from Dancy's understanding of the Model. To a certain a point, though, he also shares the above picture. He says that on the Model prospects are not ends in themselves; they become ends when we adopt them as *our* ends, which we do, effectively, by desiring them. Hence, he concludes, "reasons stem from this creation of ends by desire." (Ibid.) So far so good: this looks very much like the first part outlined above. But it is not. For Dancy is quick to point out that this does not mean that by this act of creation we now have a reason to pursue that end when we did not have before. Instead, what we have reason to do, the reason that we did not have before is to do whatever subserves that end (which, of course, involves doing just

the thing desired). (Ibid. 32) This is where the present and the previous position part ways, and where Dancy's position begins to drift toward the latter. For it is exactly the picture that Dancy denies that underlies the Model on the present reading, whereas his understanding leads directly to the Advice Point. The claim that desires are 'brute facts' that ground normativity in the way described above amounts to the claim that we have reason to pursue an end just because we desire it. Whereas the claim that when we adopt an end by coming to desire some prospect we give us a reason to subserve that end echoes the consistency requirement that stands behind the Advice Point.

We can now turn to the explanation of formula **(DBR)**. Start with the clause 'actually or hypothetically desires that  $p$ '. The background here is this. Advocates of this version of the Model admit that not every desire of the agent can ground reasons for him; not all of them are brute facts, so to speak. Since some of the agent's desires may be crazy, compulsive or inconsiderate, there should be a way of selecting among them. This is what the clause refers to; it articulates the claim that the selection procedure can invoke idealization, or may try to do without it. If it uses idealization it appeals to counterfactual conditions that make up what we may call condition  $C$ . The conditions come in two groups. In the first we find *epistemological* requirements that concern the amount and type of information the agent should have when making a decision about what to desire. They can demand that the agent have all the relevant information (Williams 1981a, 103; Railton 1997, 142-3; 2003a, 54), or only those that are available to him (though, to my mind, no one in the literature has actually endorsed this position), or only those that the science of the day provides him with (Brandt 1979, 72-3) and so on.

The second group contains requirements about *correct reasoning*. The common core includes three things. First, there is the idea that the agent possesses unlimited cognitive capacities, i.e. he never commits logical mistakes or misrepresents facts or has failings of

imagination. (Lewis 1989, 121; Railton 1997, 142; 2003a, 54) Second, there is the exclusion of certain forms of deliberation that are arbitrary and/or alienated from the agent. Typical examples include brain surgery, hypnosis, moving rhetoric or religious indoctrination. (Williams 1981a, 108-9; Rosati 1996, 302; Sobel 2001a, 222-3) Third, there are constraints on the agent's emotional make-up such as the exclusion of turbulent affections, depression or ignorance. (Firth 1952; 340-1; Darwall 1992, 164-6; Parfit 1997, 101-2) Yet, even within these constraints deliberation can take many forms. (Williams 1981a, 104-5, 110; Lewis 1989, 121-3; Darwall 1992, 164-6; Smith 1994, 159) It is not exhausted by instrumental reasoning but involves exercises of imagination, time ordering, reasoning about what constitutes one's ends, systematic justification of desires and so on.

The appeal to hypothetical desires raises an issue that we should clarify before proceeding further. It concerns a recent suggestion by Dancy. (Dancy 2000b, 28) He argues that only actual desires of the agent can ground reasons for him. Therefore, he goes on, if we opt for an idealizing version of the Model we would have to say that the role of condition *C* is to show us which of our *actual* desires would be *retained* under suitably idealized conditions. Other non-actual desires that also pass the test of condition *C*, however, cannot ground reasons for the agent. But even if we accept Dancy's claim that only actual desires of the agent can play this role, we can still point out that reasons can be relative to the agent's actual desires without accepting his proposal.<sup>20</sup> (Williams 1981a; also Smith 1994, 165) Following Williams we can claim that the desires we

---

<sup>20</sup> It is not obvious why we should accept Dancy's claim. The reason, Dancy argues, comes from the underlying picture of the Model. This picture, recall, tells that it is our ends that provide us with reasons to act. Yet, on the Model no prospect is an end for us in its own right; it can only become an end for us if we form a desire for it. Reasons stem from these ends created by desire, therefore, a desire that we would only have in condition *C* but don't now have, cannot turn prospects into ends, thus ground reasons for us; hence the reference to actual desires. But it is hard to see why it is impossible to claim that some prospect gives me a reason now because in a certain idealized situation I would desire it. For, by hypothesis, the idealized situation is such that what comes out of it is authoritative for the agent: it is these and only these desires that he should be concerned with. Unlike Dancy, I myself don't clearly see what difference it makes whether the outcome already exists or is only hypothetical. I thank Greg Bognár for pushing me on this point.

would have in condition *C* are desires we would have after correct deliberation where our understanding of this notion involves both groups of requirements mentioned above. And since deliberation is controlled by the agent's actual desires, the desires we would have in condition *C* are still a function of our actual desires. As a result, we get just the sort of relativity to actual desires that the Model needs.

Let us now turn to those attempts that give actual desires of the agent an even more signification role. I know of two theories that take this line. The first comes from Donald Hubin. He first points out that of all the actual desires of the agent we should only focus on those that are *intrinsic*. The agent's desires have a hierarchical structure and intrinsic desires are those conative states that lie at the foundation of this structure, i.e. that are not themselves dependent on other desires of the agent. (Hubin 1996, 44) But not even intrinsic desires are reason-grounding, only a certain subset of them, which Hubin calls *intrinsic values*. We don't get much explanation of what these values are, only that they are "expressions of our selves", i.e. that they "define who we are and what our lives are about", and that they can be "dispositionally and functionally defined." (Hubin 1996, 47, 49-50; 2003, 327, 329) These values, moreover, do not exercise their reason-grounding potential in the 'normal' way; it is not case that the reasons are grounded in these values due to what they are, i.e. their intrinsic nature. Instead, intrinsic values ground reasons *relationally*: all those conative states of the agent can ground reasons for acting that are not in conflict with the agent's values. (Hubin 1999, 331-2) This is Hubin's final position, which he dubs 'Value-Screened Neo-Humeanism'.

We owe the second theory to Robert Noggle. Noggle's core idea is that reasons for action can be based on only those desires of the agent that are not in conflict with the agent's deepest

concerns.<sup>21</sup> Noggle understands these concerns as desires that constitute the agent's self and secure his psychological identity, i.e. integrity over time. (Noggle 1999, 314-6) So what we get here is a picture very similar to Hubin's view above. There is a hierarchical structure of desires - a 'web of desires', as Noggle calls it - with identity-constituting desires at the core surrounded by peripheral desires that only ground reasons if they are not in conflict with the core. (Ibid. 318-320) The difference between the two theories or, rather, the most explicit difference between them is that Noggle gives a role to idealization when it comes to peripheral desires. (Ibid. 321-2) He endorses a particular version of the hypothetical desire approach according to which privileged desires are those that an ideal observer would have. The observer is the agent after idealization, i.e. after having gone through a process characterized by the two groups of conditions mentioned previously. Noggle's point in this context is then this. The observer is in line with the agent to the extent that he has the same core desires as the agent does, while his other desires are the idealized peripheral desires of the agent. It is only the desires of the observer understood in this way that can ground reasons for the agent.

The other element of formula (**DBR**) that requires further elaboration is the clause 'this is because'. To give a proper account of this clause we have to get a closer look at the wider meta-ethical background of this version of the Model. The most straightforward interpretation of this version of the Model follows naturalist lines. That is, the normative property of some consideration's being a reason (i.e. its status of a reason as endowed upon it by the reason-relation) is identical with a natural property, namely the property of desire-satisfaction (i.e. the fact that the reason-relation holds between the given consideration and action). It is this

---

<sup>21</sup> He actually deals with well-being, but I take it that he would say the same about reasons. If not, then instead of speaking about Noggle's theory, I should say 'Noggle-inspired' theory.

connection of identity that the clause is supposed to illustrate.<sup>22</sup> There may be, however, alternative meta-ethical pictures available. These are two: non-cognitivism and non-naturalism.

Start with non-cognitivism. Roughly, non-cognitivists hold that normative judgments are expressions (or projections or something similar) of non-cognitive attitudes (i.e. desires), not of beliefs for there are no normative facts to be represented. But non-cognitivists don't seem to be interested in the reason-relation at all. True, Allan Gibbard whose norm-expressivism is the most developed non-cognitivist account of normativity denies this. On his view, normative judgments are expressions of the agent's acceptance of norms where norms are understood as imperatives and acceptance is seen as a motivational syndrome of avowal characteristic of normative discourse. (Gibbard 1990, Chapter 4) And, Gibbard claims, this theory gives us an account of the relation for whenever the agent says that  $p$  is a reason for him to  $\varphi$ , he expresses his acceptance of norms that say to treat  $p$  in favour of  $\varphi$ -ing. (Ibid. 163) The first reaction to this analysis is that it does not explain the relation but merely repeats it in characterising the attitude involved. (Scanlon 1998, 58; Dancy 2004, 58) Gibbard thinks that this claim is mistaken, there are in fact two relations here, he says. If we paraphrase the claim " $p$  is a reason for  $A$  to  $\varphi$ " to say " $p$  weighs in favour of  $A$ 's  $\varphi$ -ing", then we get an analysis, which explains the relation in psychological terms: with reference to the mental state involved in the agent's acceptance of norms. (Gibbard 2003, 189-90)

---

<sup>22</sup> There is an issue that I would like to mention here. Dancy has recently argued that the favouring relation should be distinguished from the relation he calls ought-making. See Dancy (2004), pp. 22-3, 33 and (forthcoming a). Both relations have the same thing on their left-hand side, namely states of affairs. But they differ in their other relata. The ought-making relation has the oughtness of the action on its right-hand side where 'oughtness' signifies a monadic normative property. The favouring relation, on the other hand, relates states of affairs to the action itself. The reason why I mention this is because I am puzzled as to the relevance of this distinction for the understanding of the Model presented in the text. Does Dancy want to say that when dealing with reasons we should not be concerned with normative facts? I would find this proposal awkward since I have always thought that naturalists explain normativity exactly by reference to such facts. But then what is the point behind making the distinction? So, while I would not deny the difference between the two relations (how could I? the difference between them is metaphysically plain), and I also maintain that my interest lies in the favouring relation, I still hold that in order to evaluate the naturalist reading of the Model, we must deal with normative facts. I just don't see how we can proceed otherwise.



But I think this response only shows that there is no real account offered here. Since acceptance is understood as a purely motivational syndrome, Gibbard's characterisation of the attitude in terms of favouring seems to me to be nothing else but a mere figure of speech: there is no relation there where Gibbard claims it to be. True, we were very fairly lenient in our treatment of favouring in the case of Davidson and Stampe. But here we get even less. Davidson's theory at least gave us favouring as a relation among the propositional contents of certain attitudes, and Stampe invoked direction of fit expressed in the evaluative predicate of desire. But Gibbard does no such thing. In a way, he also subjectivizes normativity since he takes it to be motivational force. But this is just the 'old' non-cognitivist story over again in which favouring plays no role at all. It is nevertheless a thesis worth examination; but to do so we should rather focus on the meaning of normative sentences than on the favouring relation. And this is not a task I will take up in this dissertation.<sup>23</sup>

We are thus left with a choice between naturalism and non-naturalism. But the non-naturalist alternative is clearly not available. That is, we cannot say that when some act satisfies the agent's desire another fact obtains, namely that the agent has reason to carry out this act. This is so for a simple reason. To give desire a reason-grounding role on the present reading, desire must give us an account of the reason-relation. And this demands a naturalist background. For desire to account for the relation, facts about the relation must be facts about desires. And this can only happen if there is no further normative element involved in the relation. Desires, remember, are brute facts, meaning that they require no such normative backing. Nor should we give such a backing. For what else could that backing be than claiming that we ought to have certain desires?

---

<sup>23</sup> Dancy, in personal communication, made a similar point about Gibbard's theory. He pointed out that what at best Gibbard can get is the notion of weighing, which is an activity and can be accounted for in the purely psychological terms of acceptance of norms (or, more recently, forming contingency plans). But what we are looking for here is a relation, which, if we want to translate into more activity-like terms, is something like awarding weight to a particular consideration when deciding to act. And this, Dancy says, is not a notion that Gibbard's analysis in any way captures.

From this it follows that though we can, in fact, we should select among desires on the present reading, we need not and should not do this by employing normative judgments. Hence follows the choice of naturalism we have begun this discussion with.

This is not to say, however, that non-naturalism cannot give a role to natural facts; in fact, it can attribute to them a significant role. And it is exactly this role that may be confusing. Let me explain. Non-naturalists can claim that the agent's desires help him track reasons for action without at the same time grounding them. (Sobel 2001a, 473; 2001b, 233; Enoch 2005, 764) The idea is the same as the one noted in passing during my discussion of Stampe's theory: that if desire tracks goodness in the world, then it is goodness what gives the agent reasons to act and not what tracks it. To apply this to the present case, if the natural fact of desire-satisfaction tracks the property of being a reason, it is not identical with it. In fact, we can generalize this idea. (Dancy 2000b, 15-20, 26-7) We can say that desires can function as conditions for the existence of reasons - because, say, they serve an epistemological function as above - without themselves grounding those reasons. This role of desires is no doubt important, but is not a role that would have a normative significance: desires do not in this way make propositions about the reason-relation true.

However, the clause 'this is because' may not be able to articulate this difference. True, it is better than the clause 'only if' because a merely necessary condition would not even guarantee that the two properties co-exist. But even the clause 'if, and only if' that signifies a necessary and sufficient condition would not be enough. Although, it would guarantee co-existence, even this much need not suffice for the purposes of the naturalist. For even necessarily co-extensive properties need not be identical: this only happens on a particular view of properties, which take a property to be a set of all its instances throughout the possible worlds. And this is not something everyone accepts. (cf. Lewis 1986, 55) One can hold that co-extension does not give us the right

dependency relation between the normative property and its reduction basis, i.e. the normative property obtains not because the natural property obtains. (Enoch *ibid.* 763-5) Obviously, this is something the present clause remedies. But it has no resources to handle a further problem. For one can claim that even if the relation is of the right kind, this doesn't prove that the normative property is identical with the given natural property. (Parfit 1997, 124; McNaughton and Rawling *ibid.* 31-2) One can just say that the obtaining normative property is still a different property; hence the naturalist does not get what he wants.

### **1.3.3. Third version**

Finally, let us turn to the third version of the Model. It gives us a mixture of the previous two positions. Just like the first position, this version also begins with the instrumental principle. But then it makes exactly the opposite move. Instead of deriving the normativity of rationality from the normativity of reason, it does the reverse. Just as in the case of belief, here too there are two options. We can equate the reason-relation with the rationality-relation, or we can understand the reason-relation in terms of the rationality-relation. And just as in the case of belief, here too I think that the first reading is wrong because it does not explain reason in terms of rationality but simply eliminates it. We should therefore turn to the second reading.

The question now is how to flesh out this strategy. There are two well-known ways to carry out this task. Kantian constructivism offers the first solution. Roughly, constructivists hold that there are procedures humans have to obey in order to bring normative facts into being, i.e. confer the status of a reason on features of their situation. (Korsgaard 1996b, 35-7; 2003, 117) These procedures, moreover, are governed by principles of rationality. Now, it is no doubt possible for constructivists to claim that the only relevant practical principle is the instrumental principle. But they need not agree to this and, what is more important, they do not typically do.

Today's constructivists are dominantly Kantians who claim that the relevant procedures are those the human will must conform to; more precisely, as I show it in Appendix II, these procedures constitute the human will. (Korsgaard 1996b, 36; 1997, 245) And being Kantians they are fiercely against the idea that there is only one such procedure, namely the one governed by the instrumental principle. (Korsgaard 1996b, 120-3; Korsgaard 1997) They emphasize that the phenomenon of commitment in willing, which is their version of the consistency relation typical of rationality, appears not only in the form of the instrumental principle.<sup>24</sup> And this is a problem because the Model demands exclusivity: it supposes that the instrumental principle is the only principle of practical rationality.

There is, however, another, perhaps even more prominent way to carry out this project. To present it, we have to adopt the idealizing account of the previous version of the Model. This reading says that what we have reason to do is what we would desire to do in condition *C*. This is a picture of the reason-relation along the lines mentioned in the previous section. There is no talk of consistency between desire and action; what the desire does is to relate some feature of the agent's situation to what he may do in that situation. Yet, and this is where the crucial bit comes, we can understand the idealized situation in terms of rationality. We can speak of a situation in which we ask what the agent would desire were he to be fully rational.<sup>25</sup> This brings with it certain changes in our understanding of condition *C*. As to the epistemological conditions, since we now speak of rationality, it is not facts that count as information but beliefs, true and false

---

<sup>24</sup> The other problem with constructivists is that we always have to pay attention to what they are constructivist about. There are only few who are constructivist over the entire range of normative concepts; in fact, I only know of one: Christine Korsgaard. Others are more 'picky' in their choice. Scanlon is a constructivist about moral principles, just as Rawls is about justice. But Scanlon is a non-naturalist about reason and goodness, whereas Rawls is a naturalist about these concepts. See Korsgaard (2003), 115-118; Scanlon (1998), Chapters 1-2; Rawls (1971), pp. 46.

<sup>25</sup> Among others Brandt (1979), Chapter 1; Smith (1994), Chapter 4; Williams (1981a) appear to endorse this view. There is plenty of ambiguity in the writings of these authors, though. Pertinent to the present case, it is not clear whether they make the proposal discussed in the text, or they just conflate (because, say, not aware of) the distinction between reason and rationality. We don't have to be concerned with this problem here, however. Sobel (2001b) gives a good discussion of this issue and Kolodny (2005), pp. 510 note 2 locates Smith's theory in the present framework.

alike.<sup>26</sup> As to the reasoning conditions, here we should be aware that it is conceivable, though, of course, needs to be argued for that consistency appears elsewhere in practical rationality. This is, recall, just what Kantian constructivists claim to be the case.

If we go along with this suggestion, we reduce reason to rationality. To speak more ‘meta-ethically’, facts about the reason-relation make reference to facts about the rationality-relation. The question is what sort of facts these are: we have to clarify the meta-ethical background of this version of the Model. Unsurprisingly perhaps, it can only be naturalist. Let me explain. We can set aside constructivism because, as noted, Kantians do not accept the Model. Non-cognitivists, as much as I can make out, do not endorse the idea of reducing reason to rationality. At least, this is what Gibbard’s strategy seems to suggest. After defining reason, he gives a definition of rationality on the basis of reasons. (Gibbard 1990, 163) And he doesn’t speak of normative facts anyway.

This leaves us with a choice between naturalism and non-naturalism. The former interprets facts about rationality as being identical with psychological facts of consistency. The latter claims that there are in reality *two* facts: psychological facts of consistency *and* irreducible facts of rationality. Now, it can be claimed that non-naturalists are not among the Model’s ‘fans’ because they can hold that there are practically irrational acts other than those that fail to meet the instrumental principle. But naturalists can make the same claim. They can also say that

---

<sup>26</sup> We should therefore be careful not to conflate the two versions of the Model when formulating the epistemic condition. Indeed, the situation is made worse by the fact that beliefs, even false beliefs can figure in an account of the reason-relation that does not make reference to rationality. One such way we have already seen (in note 19 above): the case when beliefs serve as facts on the left-hand side of the reason-relation. And there are other roles beliefs can play that are consistent with the reason-relation understood as the only real normative relation. False beliefs, for instance, often serve as excuse: they make the agent’s action pardonable by allowing the agent to point out that there were good grounds to believe what the agent believed. See Dancy (2000b), pp. 6-7, 106-107. But they play no normative role: they don’t qualify as a normative reason, though we may take them to justify the agent’s act in some sense. Beliefs can also function as constraints on reasons for action: a consideration can be a reason for the agent, i.e. appear on the left-hand side of the reason-relation only if he can come to believe it. This is a version of the ‘ought implies can’ principle. In this way beliefs may serve as epistemic filters selecting out those features of the situation that are, in some sense, epistemically accessible to the agent. See (Ibid.) 56-7, 65. But it is these features that function as normative reasons and not the agent’s belief that is targeted at them.

consistency appears elsewhere in human practice. The real reason why we need naturalism as the background for the Model in the present case is the same as before. Just as with the previous version, here too a non-naturalist account of condition *C* would mean invoking normative judgments to select among desires and this would surely mean the end of the Model.

Since we have here an account of the Model very similar to the previous version, let me highlight in what respect this is a new version and in what respect it repeats the previous version. In what it is different is its underlying philosophical rationale: the idea that the normativity of the reason-relation is derived from the normativity of the rationality-relation. This means that here we have normativity emanating from the instrumental principle and not from desire; so this is a consistency-based version of the Model, more than the first one where it was the supporting reasons that lend normativity to the principle. Yet, this version is also importantly similar to the previous version in that it also gives an account of the reason-relation in terms of desire. Moreover, it also does this in a naturalist framework. Hence everything I said about the tracking role of desire applies to this third version of the Model as well.

#### ***IV. Refuting the Model***

In this chapter I have worked out a framework for the upcoming discussion. We now have three versions of the Model at hand. Since this dissertation does not endorse the Model, in subsequent chapters I argue against these versions. I won't, however, refute the possible accounts of the Model one by one. In part, my reason is that this would make the dissertation extremely lengthy: an undesirable result in any case. But the real reason is that I see no need to follow such a cumbersome strategy. The first version of the Model that gives us Dancy's Advice Point will be discussed as part of an attempt to save the other versions of the Model from an objection I present in Chapter III. I will also make use of it in Appendix II and III. Until then, however, I set it aside.

This leaves us with the second and third version of the Model. Here I make the following move. We saw that though two versions are different in their underlying philosophical rationale, they are also similar in important respects. It is these similarities that I am going to make use of in my treatment. In the remaining chapters of this dissertation I present arguments that hold against both versions. Since this means that I am going to discuss the two accounts on the same occasion, I will speak of *the* Model throughout, except for the mentioned bit in Chapter III. My first argument offers a good example. Insofar as their naturalism is concerned, the two accounts can be handled simultaneously. This is the task that I turn now to.

## CHAPTER II: NATURALISM AND TRIVIALITY

In the previous chapter we have identified the different forms the Model takes in the literature. In the present chapter I begin examining these accounts. Until the next chapter I postpone discussion of the first version and focus instead on the second and third version. Both versions are naturalist and can therefore be handled on the same occasion, speaking of the Model in the single sense. This is what I am going to do now: I examine the naturalist side of the Model.

### ***I. Naturalism and the Model***

Naturalist advocates of the Model hold that normative judgments owe their normativity to the facts they refer to in the natural world: facts about desire-satisfaction. As it stands, this is a rather bold claim that is in need of further interpretation and clarification. We first need a suitable account of what ethical naturalism (from now: naturalism) is and how it is related to the Model. This will be the topic of the present section. Then, in the second section I will begin my analysis of Derek Parfit's recent objection to naturalism. Detailed exposition of his argument leads me to the claim that naturalism can only avoid Parfit's charge if it is able to respond to the challenge of what is often called tolerable revisionism. To see whether it can do so, will be the task of the third section. Here I examine an objection made recently by Connie Rosati and argue that at present there are no workable responses available to advocates of the Model.



### 2.1.1. Ethical naturalism

We do best in understanding contemporary naturalism by introducing a distinction between methodological and substantive naturalism. (Railton 1990, 155; 1993b, 315) On the former view, naturalism adopts an *a posteriori* explanatory approach to an area of human practice or discourse such as epistemology, semantics or ethics; on the latter view, naturalism proposes an interpretation of the concepts in some area of practice or discourse in terms of *natural* properties or relations. Since it is notoriously difficult to define what a natural property is, I will not attempt to give a precise definition.<sup>27</sup> Instead, I will act on the supposition that such an account can be given. This assumption is not only needed to get the argument going, but is also legitimate given the similar definitional problems other theories, particularly non-naturalist accounts have to face. Finally, since my aim in this chapter is to attack naturalism and not to defend it, making the assumption also makes my argument more charitable from the naturalist point of view.

Let me first turn to methodological naturalism. There are two things to note here. First, someone can be a methodological naturalist without being a substantive naturalist -think of Allan Gibbard's evolutionary grounding of norm-expressivism or Richard Hare's defense of prescriptivism on the basis of linguistic intuitions. Second, a substantive naturalist need not be a naturalist in the methodological sense. Certain naturalists use *a priori* conceptual analysis as their method. Three attempts deserve particular attention. David Lewis gives a direct naturalistic definition of value, while both Michael Smith and Frank Jackson employ a two-stage analysis. (Lewis 1989; Jackson 1998, Chapters 5-6; Smith 1994, Chapter 2) The idea is to see whether ethical concepts (conceptual stage) do indeed instantiate in the world (substantive stage). To use

---

<sup>27</sup> There are many attempts in the offing. A popular choice is to say that natural properties are those with which it is the business of the natural sciences, or of psychology to deal, or which can be completely defined in the terms of these. See Moore (1903), pp. 25; Wiggins (1993), pp. 303; Smith (1994), pp. 17; Copp (2003). But there are problems with this idea as well as with the attempts that try to remedy its shortcomings. For a good overview of different definitions and the difficulties they face see Ridge (2003).

Smith's example, if we want to know whether there are witches in the world, we first clarify our concept of a witch: we set up conditions a person has to fulfill in order to qualify as a witch. Then, armored with these conditions, we check if there is anyone in our world who fulfills these conditions, that is, if there are witches in our world. (Smith 1994, 64)

Let me elaborate on this a bit more. Take Smith's analysis of rightness. He claims that moral concepts are concepts of reason; therefore we first need a conceptual analysis of this term. (Smith 1994, 58) We do this by relying on the descriptions of the inferential and judgmental dispositions of those who have mastery of the term. (Ibid. 39) These platitudes give us a network of relations in which we can identify the meaning of the term. (Jackson 1998, 39) In the case of 'reason', Smith claims, they point to the notion of advice and result in the following definition: what one has reason to do is what one would desire to do if one were fully rational. (Smith *ibid.* 151) The next step is to connect this analysis to rightness. We can do this once we realize that what distinguishes moral reasons from normative reasons in general is their content and that the borders of this content is fixed by platitudes concerning the substance of morality. (Ibid. 183-4) With this at hand we can proceed to the substantive phase of the analysis. We can say that rightness is the feature that we would want acts to have if we were fully rational, where these wants have the appropriate content. (Ibid. 185) And it is possible that we find no such feature in the world in which case nothing is right to do.

Two remarks are in order. First, Smith and Jackson differ as to the reductiveness of the conceptual phase. Jackson's so-called network-analysis is given in reductive, natural terms, whereas Smith's summary-style analysis does not refrain from using normative, unreduced terms. (Jackson *ibid.* Chapter 5; Smith *ibid.* Chapter 2) However, Smith notes that the unreduced terms he uses have a natural reduction basis and if necessary, via another two-stage analysis, they too can be uncovered. (Ibid. 186) Second, Jackson supposes that we have identity relations

throughout the stages (this requires qualification, as I will explain later).<sup>28</sup> This is because he takes both stages to articulate necessary and sufficient conditions: *φ-ing* is right if and only if it fulfills certain conditions; it fulfills those conditions if and only if it has property F; hence *φ-ing* is right if and only if it has property F. But, remember the previous chapter, these material conditions signal identity only if one holds that necessarily co-extensive properties are identical. And this is possible only if one further supposes that a property is a set of all its instances throughout the possible worlds. While Jackson explicitly endorses this view, not everyone does so: it is particularly disputed in the field of reasons. (Jackson 1998, 124-6; cf. McNaughton and Rawling 2003, 30-34; Enoch 2005, 763-5)

Such complexities, however, need not trouble us at this point. Let us turn instead to the second type of naturalism: substantive naturalism. The two main approaches are analytical naturalism and non-analytical naturalism. The former has four characteristics. (Jackson 1998, 144-6) First, it rejects any form of methodological naturalism and opts for conceptual analysis with all the variations mentioned above. Second, it can be neutral on whether the analysis in the conceptual phase gives the meaning of the ethical term in the descriptive sense, i.e. the meaning of a term is given by the property that a competent speaker associates with it, or it only causally fixes its reference in Kripke's sense. Third, it allows that the discovery of the identity between ethical and natural properties in the substantive phase might take place *a posteriori*, i.e. empirically. Fourth, it can but need not hold that the identities thus discovered are themselves analytic and thus *a priori*. Finally, in all cases whenever meaning in the descriptive sense is concerned, two options are open. (Brink 1989, 152-3) Either the ethical term is synonymous with

---

<sup>28</sup> Smith is silent on the issue. This is one reason why it is unclear whether he actually gives us a naturalist account of reasons or instead he holds that there are independent normative truths, which our counterfactual desires track: a view that can but need not be naturalist and one that certainly need not support the Model. See Sobel (2001a), pp. 473 n. 19 and (2001b), pp. 234, n. 24 on ambiguities in Smith's position.

the non-ethical term in which case ethical properties are identical with natural properties; or the non-ethical term implies the meaning of the ethical term in which case ethical properties form a subset of natural properties.

Non-analytical naturalism comprises of two positions. One is the idea of reforming definitions. It is similar to analytical naturalism in that it appeals to an *a priori* analysis of ethical terms, but is different from analytical naturalism in that it doesn't intend to give an analysis of our *actual* use of the term, but as a substitute for it. (Brandt 1979, Chapter 1; Rawls 1971, 60-63§) The other position is non-analytical naturalism proper: it makes no claim about meaning, only about properties. It comes in two distinct forms. Either it is the view that ethical properties are reducible to natural properties and we can give synthetic identity statements about them. (Railton 1997, 2003) Or it is the claim that ethical properties are not reducible to natural properties, though they are nothing over and above natural properties. (Sturgeon 1985ab, 1986ab; Boyd 1988, Miller 1979, 1985; Brink 1989, Chapter 6) Non-analytical naturalism achieves this by denying two theses. One is the descriptive theory of meaning mentioned above; the other is the claim that all necessary truths are analytical. It instead settles for Kripke's causal theory of reference and holds that there are *a posteriori* necessities. As the comparison shows, it is this last claim that really distinguishes it from analytical naturalism.<sup>29</sup>

---

<sup>29</sup> We can compare this to the terminology of others. Smith's 'definitional naturalism' is what I call analytical naturalism where analysis is taken to be wholly reductive, while his 'metaphysical naturalism' corresponds to my non-analytical naturalism. See Smith (1994), pp. 26-7. Jackson's 'analytical descriptivism' covers those analytical naturalist views where analysis is given in purely reductive terms, his 'metaphysical naturalism' is my substantive naturalism that includes both analytical and non-analytical versions and his 'ontological naturalism' corresponds to my non-analytical naturalism. See Jackson (1998), pp. 146. Finally, what Darwall, Gibbard and Railton (1997), pp. 24-30 calls 'post-positivist non-reductionism' includes analytical views that allow for meaning implication as well non-analytical theories that opt for constitution; and what they call 'reductionism' includes analytical views that restrict discussion to synonymy as well as non-analytical accounts that favour identity.

### 2.1.2. Naturalism and the Model

With these distinctions in mind, let us now see how naturalists interpret the Model. To do so, we can adopt the definition from the previous chapter. Recall, it goes like this:

**(DBR)** If its being the case that  $q$  is a reason for  $A$  to  $\varphi$ , this is because, there is some  $p$  such that  $A$  actually or hypothetically desires  $p$  (in condition  $C$ ) and given that  $q$ ,  $\varphi$ -ing subserves the prospect of  $p$ 's being realized (or continuing to be realized).

On a naturalist reading, **(DBR)** designates two properties, one ethical (the property of being a reason), the other natural (the property of satisfying desire), and connects them in some way. This much I have already mentioned in the previous chapter. Now we know a bit more since we know (a bit) more about naturalism itself. That is, we know that the connection between the two properties can take many forms: it can be analytical or non-analytical, identity or constitution and different variations of these. The rest of the definition I leave unexplained here since I would be just repeating myself. I have said everything relevant at the end of the previous chapter, so anyone can turn to that explanation to refresh memories. Let us instead see what trouble, if any, this otherwise plausible naturalist backing might cause for the Model.<sup>30</sup>

---

<sup>30</sup> Why is it plausible? Because, unlike non-naturalism, it gives a straightforward account of the epistemology and metaphysics of reasons. Of course, there still remain problems, other than the one I will discuss; but these are problems that beset realist theories in general such as the well-known controversy about moral disagreement or the debate about moral explanation.

## ***II. The objection: triviality***

### **2.2.1. The objection outlined**

Here is the objection. Derek Parfit has recently attacked naturalism through its ‘substance’; he has claimed that the problem with naturalism is its insistence on reducing ethical properties to natural properties.<sup>31</sup> (Parfit 1997, 123-4) Applied to the analytical version of the Model, his objection takes the following form. The normative claim:

(1)  $Q$  is a reason for  $A$  to  $\varphi$

means

(2) There is some  $p$  such that  $A$  actually or hypothetically desires  $p$ , and, given that  $q$ ,  $\varphi$ -ing subserves the prospect of  $p$ 's being realized (or continuing to be realized).

If we accept (2) Parfit claims, we can no longer believe that we have a reason to do what satisfies our desires. We can only believe that to satisfy our desires is to satisfy them and this is a tautology that makes practical reason claims trivial and thus non-normative. We can call this, following Parfit, the triviality objection.

This might be obscure, so let me put this point in a different way. Take the example of Joe who wants to climb the Mount Everest. According to the Model, Joe's reasoning has the following form:

---

<sup>31</sup> Parfit notes that the objection first appears in a footnote in Sidgwick (1907), pp. 26n. Gibbard (1990), pp. 33 also makes use of it in his rejection of naturalism. A detailed presentation of the argument, one to which my discussion owes a lot, could be found in the earlier version of Parfit (ms), Chapter 2; it is, however, no longer included in the pre-ultimate draft of the manuscript.

- (3) Climbing the Mount Everest satisfies a desire of mine that I actually have or would have in some suitably defined condition  $C$

given that,

- (4) When applied to acts, the term ‘I have a reason to  $\varphi$ ’ means, “I desire  $p$  or would desire it in some suitably defined condition  $C$ , and  $\varphi$ -ing subserves the prospect of  $p$ ’s being realized (or continuing to be realized)”

I conclude,

- (5) I have a reason to climb the Mount Everest.

The triviality objection says the following. Joe’s conclusion in (5), though through a further premise (4), only restates his premise in (3). Can the truth of (4) help the naturalist out? It cannot. Since (4) is a definitional truth, it uses the very same concepts and designates the same properties as does (3). Therefore it makes the concept of reason redundant, a mere abbreviation for ‘satisfies my desire that I have or would have in some suitably defined condition  $C$ ’. But it just cannot matter to show how a certain term, in this case the term ‘reason’ is used. It simply does not make a difference if we show that we can refer to some longer term in a shorter, more convenient way: that we can use the word ‘reason’ for it. But unless these concerns receive an answer, Joe cannot properly believe that he has a reason to act if and only if acting satisfies certain of his actual desires or those that he would have in condition  $C$ ; he can only believe that to satisfy his desires is to satisfy those desires, full stop. And this, as it stands, abolishes the normativity of his belief:

he cannot believe that he *should* do what his desires command him to do. So he cannot believe, as a particular instance of it, that he *should* climb the Mount Everest.

So far we have only considered the analytical version of the Model. Is it not possible that non-analytical naturalism can somehow avoid triviality? After all, it is easy (technically, not philosophically) to formulate the Model as stating a synthetic truth, so it is important to see whether such a shift in substance can help. Parfit thinks it cannot. His point is this. (Ibid.) Take our previous example and give it a non-analytical turn. Instead of (4), the non-analytical premise would then be:

(6) As another way of reporting the natural fact or ascribe the property (3) refers to, I can say that there is a reason for me to climb, or that I should climb.

But again: (6) only says that (3) could be expressed in another way; otherwise it reports the same fact. That is, to say that climbing fulfills Joe's desire and to say that he should climb is to report the same fact. And, though now the terms (3) and (5) uses are different, they are not different in the relevant respect: they are both descriptive, expressing beliefs of the agent, which, by property identity, are about the same part of the natural world. Consequently, (5) adds nothing to Joe's original reasoning: from the perspective of the triviality objection it makes no difference whether triviality arises because (3) and (5) mean the same or because the two report the same fact. Hence the non-analytical form of the Model is just as much unable to account for the normativity of reason-claims as does the analytical version.

The triviality objection as presented here is different from well-known objections against naturalism. To begin with, it makes no claim about the impossibility of deriving an 'ought' from



an 'is'. This claim, traditionally known as the 'is/ought' thesis, has two interpretations, one logical (no normative statements can be derived from statements that involve no normative terms), the other motivational (normative statements necessarily motivate, hence cannot be derived from statements of facts). (Brink 1989, 145-9; Smith 1994, 190-3; Jackson 1998, 139-142) The triviality objection, however, requires no mention of either logical or motivational mistakes. Although in a sense it is also about how to bridge the gap between the normative and the non-normative, it need not postulate either of these failures to make its point. Furthermore, it is perhaps less obvious but is still clear enough that the triviality objection is different from Moore's open question argument. Very schematically, in Moore's view no naturalist definition of normative terms is adequate since such definitions always leave questions about what is right, rational etc. open and this would be impossible if they were to function as definitions. (Moore 1903, 10-21) The triviality objection, however, is not dependent upon claims about meaning. It has troubles with the reductionist element in naturalism and as such it is indifferent to whether that reduction takes an analytic or non-analytic form.

Furthermore, the triviality objection poses a real challenge to the naturalist. I emphasize this because one might wonder whether the naturalist really needs to look for a response to the triviality objection. More precisely, the naturalist might give a response, but a debunking one. He can admit that his understanding of normativity is trivial, but add that he has no problem with this view of normativity and, if it comes to that, he is willing to bite the bullet. The alternative, non-trivial view of normativity, he can point out, might be a valid one but to appeal to is question-begging. At the moment we just have two competing readings of normativity and an independent argument is needed to prove the truth (or falsity) of either.<sup>32</sup> I don't think this is so. The issue

---

<sup>32</sup> Many have put this response to me on behalf of the naturalist (regardless whether they are naturalist themselves). I remember Tony Booth and Kaj Strandberg, in particular.

here, rather, is where the onus of proof lies and I believe it lies on the naturalist side. This is because our ordinary understanding of normativity goes beyond the account the naturalist so happily embraces. I don't know how to prove this, but it seems to me that accepting the naturalist proposal would leave normative inquiry impoverished. (cf. Gibbard 1990, 33-4) We think – again, this is my impression - that when one reasons about what to do, one takes oneself to arrive at a genuine normative conclusion, not just a repetition of what was already involved in one's premises. It is this phenomenon the triviality objection articulates and that naturalists must account for.

### **2.2.2. First response: irreductionist naturalism**

I see three ways naturalists can accomplish this task. I first deal with the two that are unsuccessful leaving the third, more promising one to the end. The response that immediately comes to mind is that naturalists need not be reductionist but can also endorse some form of irreductionism. That is, they need not hold that the relation between properties is that of identity; instead, they can say that ethical properties are irreducible properties, though they are nothing over and above natural properties. In this case the triviality objection does not work. To take Joe's argument above, his conclusion reports an ethical fact that is not identical with the natural fact that his premise designates, but is still not itself an additional property. Consequently, Joe's reasoning is not trivial; hence its conclusion preserves normativity.

Let me set aside the issue whether the irreductionist move could avoid collapsing into reductionism.<sup>33</sup> There is another, equally serious problem lying here. To see it, we should turn to

---

<sup>33</sup> The dilemma hidden here is this. Irreductionism insists that it is still a form of naturalism. But if it is, then one can start to wonder whether it is really different from reductionism. After all, if inseparability is enough to save naturalism, what difference remains between the two approaches? Along general lines, both Parfit (ms: earlier draft), Chapter 2 and Darwall, Gibbard and Railton (1997), pp. 27 poses this question. In my discussion I set aside this

Attila Tanyi, THESIS, Page 66

metaphysics. Here we first find that the irreductionist approach is anything but self-evident. Irreductionists normally appeal to supervenience in picturing the relation between ethical and natural properties. However, this is an idea that forms part of many positions other than irreductionist naturalism and these positions view it differently. (Darwall, Gibbard and Railton 1997, 27) Reductionist naturalists think that the ethical supervenes on the non-ethical because the two are identical; non-naturalists claim that the ethical supervenes on the non-ethical because the ethical is non-natural; finally, non-cognitivists deny that supervenience would occur on the level of properties at all and instead claim that it holds between concepts. So the irreductionist reading of supervenience certainly needs further explanation.

The only attempt I know of to explain this strong relation between properties in irreductionist terms is David Brink's notion of constitutional supervenience. (Brink 1989, 157-8) He gives two reasons to support his proposal. Both are versions of the multiple realization argument and appeal to an analogy with other fields. The first is introduced like this: "For example, a table is constituted by, but not identical with, a particular arrangement of microphysical particles, since the table could survive certain changes in its particles or their arrangement. Similarly, moral properties are constituted by, but not identical with, natural properties if, though actually realized by natural properties, moral properties can be or could have been realized by properties not studied by the natural or social sciences." (Brink 1989, 158) The question is whether this example does the work Brink wants it to do. I think it does not. The reason is the following.

Brink's example employs an understanding of property that appears to be metaphysically flawed. (McNaughton and Rawling 2003, 39-40) He mentions a physical object and compares it

---

reaction because it doesn't answer the crucial ontological challenge: that even if entities do not exist independently, they are still separate ontologically – and this is enough for the irreductionist strategy to get off the ground. I thank János Kis for pressing me on this point.

to a property. It is, moreover, important for Brink that the analogy is between these two entities and not between two properties, i.e. the property of being a table and the property of being right.<sup>34</sup> For it is hard to see how he can get the kind of relation he is postulating between microphysical particles and properties: these are just two different kinds of things. If, however, he employs the analogy in its original form, he puts himself in trouble. This is because, metaphysically speaking, objects ‘behave’ differently than properties. Let me explain. Take the three rival metaphysical theories. On nominalism, there are *no* such things as properties: only concrete particulars exist that can only be in one place at any given time, and only one of them can be in the same place at the same time. Hence this ‘understanding’ of property would not help Brink to establish his analogy. On realism, properties do have an independent existence but they are taken to be abstract universals: they can be at many places at the same time, and many of them can be in the same place at the same time. This means that one can move from world to world and identify *instances* of properties just as one can identify tables (think of, for instance, Jackson’s view of property). But one cannot identify an *entire* property throughout the possible worlds: properties are trans-world entities and thus are unsuitable for Brink’s purposes.

Can we find resort in the third of the theories, trope theory? I doubt it. On this view there are such things as properties (*qua* realism), but they are particulars and not universals (*qua* nominalism). More precisely, a property is taken to be a class or set of exactly similar or resembling tropes where individual tropes are understood as *instances* of the given property. For a substance to have or to instantiate a property is for one of its tropes to exactly resemble all of the tropes that comprise that property (or for the set of tropes that is the substance to overlap the set of tropes that is the property). (McDonald 1998, 35-40) And this shows that trope theory cannot provide hope for Brink either. For on this view properties are again trans-world entities.

---

<sup>34</sup> Both Gianfranco Pellegrino and Anders Strand asked me why the alternative reading is not possible.  
Attila Tanyi, THESIS, Page 68

One can identify *instances* of properties from world to world just as one can identify tables, but one cannot identify a property *per se* (an entire property, as I called it above). Hence Brink's analogy with the table remains unfounded.

Nor does the second analogy help his case; in fact, it makes things worse. It appeals to the one/many relationship that, according to some functionalists in the philosophy of mind, exists between mental states and physical systems. Brink's claim is that we find the same situation in morality. He says: "For example, both the property of injustice and particular instances of injustice, in whatever social and economic conditions they are actually realized, could have been realized by a variety of somewhat different configurations of social and economic properties and property instances. Moral properties could have been realized by an indefinite and perhaps infinite sets of natural properties." (Ibid.) Jackson makes a similar proposal. (Jackson 1998, 141) Revising his original idea that we have identity relations throughout the two-stage process, he now says that his conceptual analysis tells nothing about the metaphysics of rightness. Drawing on the analogy with the distinction between role and resultant property in the functionalist theory of mind, he points out that the property of rightness need not be identical with the property that we find in the world in the substantive phase of the analysis. For it might be that the property of rightness is the *second-order* property of having that property.

However, once one takes over the multiple realization argument from the philosophy of mind, one should not be silent about what comes with it. In particular, one standard objection against irreductionist functionalism seems to have an analogue in ethics as well: Jaegwon Kim's causal exclusion argument. With significant simplification, Kim's claim is that if the first-order physical property ('realizer property') can be a sufficient cause of a physical event (what is called the 'physical closure' principle), and we suppose that the mental supervenes on the physical in the irreductionist way, then there appears to be no causal work left for the second-order mental

property ('role property'). (Kim 1999, 37, 53) It seems that the role causality plays in the mental causation debate is taken over by truth in the corresponding ethical discussion.<sup>35</sup> For Jackson clearly supposes that both first-order properties as well as second-order properties can play the role of truth-maker in the conceptual phase. (Jackson *ibid.*) But if this is so, then we seem to have no need to complicate the picture by introducing the second-order property: there is just no truth-making work left for it to do. Its role as a truth-maker is entirely preempted by the first-order property. The analogy with philosophy of mind is a dangerous consideration to appeal to.<sup>36</sup>

### **2.2.3. Second response: non-descriptive naturalism**

I conclude that irreductionist naturalism must give way to reductionist naturalism and then the triviality objection is allowed to do its job: we are back where we started. But an advocate of naturalism might still find my treatment unfair. He might claim that I am forgetting something crucial about the position: that it is a claim about concepts, not just about properties. And this is a striking omission for naturalists need not hold - what I took them to be holding - that ethical concepts are exclusively descriptive. Perhaps some form of non-descriptive naturalism is true, and then the concepts that appear in the agent's reasoning will be different in kind thereby avoiding the trap of triviality.

There are three problems with this proposal but they are all aspects of the same basic issue. Let us take a closer view at the structure of non-descriptivist naturalism. It must navigate between two equally unacceptable positions: non-cognitivism on the one hand and descriptivist

---

<sup>35</sup> I thank Anders Strand for calling my attention, even if indirectly and perhaps not deliberately, to the possible analogy between the two fields.

<sup>36</sup> To be fair, we must note that Jackson himself favours the reductionist reading. Here is what he says about rightness (referring to Smith 1994, 74-6): "We want rightness to be what makes an action right, not in the causal sense but in the sense of being what ought to be aimed at. [W]hat ought to motivate us, and what we should value and pursue, is not the moral status of our actions *per se*, but the goods that confer the moral status." See Jackson (1998), pp. 141. Hence, he concludes, we'd better opt for first-order properties instead of the second-order properties that possess them.

naturalism on the other. The way it does this is by combining the non-descriptivism of the former with the realism of the latter. The result is a hybrid view according to which the semantic role of ethical predicates is both to refer to robust ethical properties *and* to express some non-cognitive state of mind. There are well-known difficulties with constructing a defensible non-descriptivist side (basically all the relevant objections to non-cognitivism belong here, plus some specific claims concerning certain naturalist proposals), but there is another problem characteristic of this position only: how to combine the different aspects so that they fit properly together? Keeping them separate without establishing any connection between them seems a rather implausible solution. After all, we are talking about *one* mental state, so it is an obvious requirement to say something about how the sides involved in that state exist together.

However, I do not regard this as a daunting obstacle. There are proposed solutions in the literature both on the naturalist and the non-naturalist side (for a recent naturalist attempt see Copp 2001). My concern is driven more by two corresponding problems. First, the position as described here seems unstable. For it is just difficult to see what explanatory advantage non-descriptivist naturalism has over non-cognitivism and descriptivist naturalism. Accounting for motivational judgment internalism – the thesis that normative judgments necessarily motivate – is a good candidate, but non-cognitivism takes this hurdle by construction, whereas descriptivist naturalism has its own solutions (e.g. Smith 1994, 177-181; Jackson 1998, 157-160). At the same time, both rival accounts are simpler. Non-cognitivism can do without any metaphysical commitment, while not denying that ethical claims may refer to plain natural facts; and descriptivist naturalism can do without any non-descriptivist commitment, while not denying that ethical claims may have motivational aspects. The question therefore is what explanatory work is left for the additional features of non-descriptive naturalism to do, and the answers seems to be that none.

The third problem takes us back to the context of the triviality objection. Recall the way non-descriptive naturalism aims to overcome the objection. It says that while the agent's reasoning begins with purely descriptive premises, his conclusion will be an ethical claim that has a non-descriptive aspect as well. But it is unclear whether this response delivers, by the theory's own light, the result it is intended to. For the objection's driving thought is that naturalism cannot bridge the gap between the agent's non-normative premise and his normative conclusion because the latter simply repeats the former. Now it is true that the theory under discussion adds something to the agent's reasoning, namely that its conclusion is not purely descriptive. Yet, it is unclear whether thereby it also bridges the gap between the normative and the non-normative. After all, non-descriptive naturalism is still a realist theory, hence it must explain normativity with reference to a fact of the world, in this case the natural world. But if ethical judgments become ethical and thus normative by virtue of representing certain natural facts (this is their *primary* semantic role), the non-descriptive element of naturalism will not add anything *normatively significant* to the agent's reasoning. Therefore it is at least unclear whether non-descriptive naturalism faces up to the triviality objection in the first place.

I don't see how naturalists can deny this.<sup>37</sup> They just cannot take the alternative way because that would be endorsing something like Allan Gibbard's revised account of his non-cognitivism. (Gibbard 2003, Chapters 1, 2, 9) He too accepts non-analytic naturalism but gives it a non-cognitivist turn, holding that normativity is best captured by a non-cognitivist analysis (as he earlier put it, our normative terms carry an element of endorsement that only non-cognitivism can capture). This would be a way of bridging the gap the triviality objection uncovers; but this would also be a non-cognitivist and not a naturalist way to do this. Alternatively, the non-descriptivist naturalist might try to give us a hybrid account of normativity. But I just don't see

---

<sup>37</sup> This short discussion has profited from conversations with Daniel Elstein.  
Attila Tanyi, THESIS, Page 72



how this would go: how can one claim that normativity is descriptive and non-descriptive at the same time? How can he marry the two without giving up one of them? Finally, there is the possibility of claiming that the presence of the given natural property only signals the presence of normativity: normativity itself still follows a non-descriptivist analysis. However, this is nothing else but a sophisticated (or, in any case, a more complicated) version of a Gibbardian position and as such it invites the same assessment: it is not available to a ‘real’ naturalist.

#### **2.2.4. Third response: vindicatory reductionism**

There is, however, a third response that is still open to the naturalist, even if he advocates reduction and descriptivism. He has to hold that ethical properties, though reduced to the natural, are nevertheless genuine properties with an independent role to play in human practice and discourse.<sup>38</sup> Peter Railton calls this project vindicatory reductionism. He puts the idea in the following way: “[...] the naturalist who would vindicate the cognitive status of value judgments is not required to deny the possibility of reduction, for some reductions are vindications – they provide us with reason to think the reduced phenomena are genuine.” (Railton 1990, 161) And, at another place, he says: “Some reductions explain away the reduced phenomenon, but others simply explain it – and thereby show it to be well-founded.” (Railton 1993b, 317)

To support his case Railton brings examples. The reduction of water to H<sub>2</sub>O or the reduction of salt to NaCl, he says, reinforces rather than impugns our sense that there really is water or salt. In short, it is vindicatory not eliminative. Similarly, to take an example from navigation, the reduction of seaworthiness to a set of physically realized dispositional properties of vessels, does not eliminate talk of seaworthiness; instead, it vindicates our use of the term. (Railton 1990, 166) We can find counterexamples too, cases when the reduction did (or would, if

---

<sup>38</sup> It was Krister Bykvist whose remarks prompted me to consider this response.

proposed) eliminate the reduced term: phlogiston, caloric fluid, vital force, polywater, the non-divine reduction of the sacred are all like that. Hence, Railton points out, whether a reduction is vindicative or eliminative will depend on the specific character of what is being reduced and what the reduction basis looks like.” (Ibid. 161)

This is still vague. What makes a reduction in one case eliminative and vindicative in the other? The key term is ‘vindicative information’. The idea is that vindicative reductions provide us with crucial information about the notion reduced by placing it in the world in an unproblematic way. (Railton 1993b, 318) Take the reduction of water to H<sub>2</sub>O. In knowing that water is water, what we knew was that water is the colorless liquid that flows in rivers, falls from clouds as rain, etc. But in coming to know that water is H<sub>2</sub>O, we were told that water is a substance whose molecules consist of two hydrogen atoms bonded to one oxygen atom. This is an important piece of information that explains why water exists and takes the form as it does. Hence, though the facts reported are the same, it is important that they can be reported in some other way: for such possibility conveys vindicative information. As opposed to this, eliminative reductions either do not produce any information or, if they do, it is such that it proves the reduced notion to be redundant or non-existent.

An important element is still missing. Railton himself is a non-analytical naturalist and this may suggest that vindicatory reductionism is open only to non-analytical views. It might be claimed that what matters is reality, not how reality can be described. It is unimportant whether we can say the same thing, or express the same concept, by using different words. But it may have great importance whether, when we say two different things, or use two different concepts, we are referring to the same property, or reporting the same fact. Truths about properties are

truths about reality.<sup>39</sup> I don't think this is so. Let me put the idea of vindicatory reductionism in a different way. Recall our claim that Joe's practical reasoning is trivial because his conclusion adds nothing to his premise. Now we see that, on the non-analytical reading at least, this may not be so. For on this reading Joe's premise and his conclusion employ different concepts and this means that their content is given by different propositions (for this view of the identity of propositions see Schick 1991, 72-8). What we have here then is a case of co-reportive propositions, which, we just saw, can be significant if the information it produces is vindicative not eliminative.

However, and this seems to be behind the point above, no such thing happens in the analytical case since here we have the same concepts, thus no co-reporting propositions are present. This is certainly so, but this need not rule out the possibility of producing vindicative information. What contemporary analytical naturalists claim is that by using conceptual analysis we can get a clearer view on concepts we are otherwise familiar with. This is because many facts about these concepts are hidden in the unobvious, opaque conceptual relations of what Jackson calls 'mature folk morality': the morality that we end up with after debate and critical reflection. (Jackson 1998, 151) As a result, though analytical views lack the sort of metaphysical background – in the form of separate but co-reporting propositions – that non-analytical accounts share, they may still be vindicative. This happens when they come up with an analysis of a normative concept, which is such that it brings extra information in the way mentioned above. The failure of vindication does not follow from the mere fact of analyticity.<sup>40</sup>

---

<sup>39</sup> Originally this claim appeared in Chapter 2 of the earlier draft of Parfit (ms). But, as has been noted previously, this chapter is no longer included in Parfit's manuscript.

<sup>40</sup> János Kis has pointed out to me that one might invoke Rawls' distinction between concept and conception to show that the analytical naturalist position is false. The idea would be to argue that the informative part of the analytical project does nothing else but to take us from the concept of reason to a conception of reason through interpretation and not through analysis. However, the distinction between concept and conception is the basis for the *constructivist* understanding of normative issues: the concept names the (practical) problem and the conception proposes the

Attila Tanyi, THESIS, Page 75

### **2.2.5. Summary**

The possibility of informative reduction provides metaphysical grounds for vindication. It shows that it can matter whether some fact can be reported in a different way or that the term some fact refers to has a synonym or implies another in meaning. But this is not enough: we just don't know what the clause 'informative in a vindicatory way' refers to. Conditions are needed that can serve as constraints on the possibility of vindication. In the next section I will discuss one such criterion: tolerable revisionism. This condition, it should be noted, poses a challenge in its own right; yet, its connection to Parfit's charge makes it even more illuminating.<sup>41</sup>

### ***III. The condition: tolerable revisionism***

#### **2.3.1. Tolerable revisionism defined**

Revisionism is best understood as an indirect form of G. E. Moore's open question argument (OQA). I have already referred to Moore's argument briefly; here it is in a more detailed form. (Moore 1903, 10-21; Ayer 1936, 103-106; Brink 1989, 152-3, 162-3) Since for any natural property F it is possible to think that x has F while at the same time thinking that x is not right

---

solution. See Korsgaard (2003), pp. 118. While naturalists may share the idea of problem solving, as good realists they also insist that we have normative concepts primarily because there is an ethical reality they aim to describe. The question, therefore, is whether this kind of rejection of constructivism would also bring with it the rejection of the distinction between concept and conception. That this is an important issue is shown by Korsgaard's claim that constructivists, with the help of the distinction, can avoid triviality. They can say that satisfying our desires is what we have reason to do because thereby we would just mean that satisfying our desires is what solves the problem signified by our concept of reason, namely, acting. See Korsgaard (ms a), Lecture 2 and (ms b). I say more about the constructivist understanding of normativity and its (possible) connection to naturalism (and realism in general) in Appendix II. I, however, won't take up the task of answering the above question: that project must wait for another occasion.

<sup>41</sup> There is another condition: practical justification. See Railton (1990), pp. 173 and Railton (1993b), pp. 324. However it is not clear what exactly practical justification requires, nor how it can be turned into an argument against the naturalist reading of the Model. I suspect that a connection can be made to Korsgaard's regress argument against what she calls substantive realism as it appears in Korsgaard (1996b), pp. 38-9, (1997), pp. 240-3, (ms a), Lecture 2, pp. 13-15 and (2003), pp. 110-112. However, besides the unclear relation between the argument and the condition, there is also the question of how successful Korsgaard's argument is. On this see FitzPatrick (2005) and (ms). Due to these complexities, I find it better to set aside this condition here. I treat Korsgaard's argument in detail in Appendix II.

(good, rational etc.), it follows that, for no natural property F is it self-contradictory to claim that x has F but x is not right (good, rational etc.). But it would have to be self-contradictory, if it were possible to give a naturalistic definition of rightness (goodness, rationality, reason etc.). Therefore no synonymy or meaning implication can exist between ethical and non-ethical terms.

Many have argued that Moore's argument is flawed. As we saw, analytical naturalists claim that there are unobvious, opaque, non-transparent or equivocal conceptual relations. It is these relations, they claim, that explain why it is possible to ask questions: what Moore's argument at best shows, therefore, is that he worked with an inadequate notion of conceptual analysis. (Firth 1952, 330; Lewis 1989, 129-30; Smith 1994, 38; Jackson 1998, 151) Non-analytical naturalists, on the other hand, reject the classical theory of meaning that underlies the OQA. And since they make no statements about meaning (in the classical sense), only about identity or constitution, they have no problem with doubts about meaning in discovering these relations between properties. (Sturgeon 1985a, 25-6; Railton 1990, 156; Brink 2001; Copp 2000) Finally, those naturalists who provide reforming definitions have no problem with open questions because they aim to give us a substitute of a term not an analysis of its actual use. Consequently, the fact that we find certain questions pressing comes as no surprise to them. (Brandt 1979, 14-6, 126-8) The price that we have to pay for finding a proper substitute, they say, is exactly that we lose certain bits of meaning in the process. Again, Moore's argument has shown nothing that advocates of a proper naturalist - in this case a reforming naturalist theory - wouldn't be able to live with.

This last answer, however, provides ground for a reformulation of the OQA. Perhaps naturalist critics of Moore are right and a *direct* defeat of naturalism is impossible.<sup>42</sup> But this doesn't rule out that an *indirect* defeat would also be impossible. This is where revisionism enters the stage. The idea is this. Every naturalistic account of ethical terms leaves some questions open, but this is not because our conceptual analysis is unobvious or because we can only come up with synthetic statements. Instead, as advocates of reforming definitions have long been emphasizing, an account of ethical concepts, even if it is intended to be an analysis of the actual use of the term, is necessarily revisionist to some extent. It gives us a proposal that aims to capture everything about the given term, but there may be some functions of the pre-reductive term, which the proposal cannot account for. This is why open questions occur. From this perspective it does not matter whether the proposal in question offers a conceptual analysis, a synthetic identity statement or a reforming definition. They all want to fulfill the task of giving us an account as perfect as possible and they all fail if they leave us with questions concerning important function(s) of the given term.

This gives us a condition for vindication. For a naturalistic reduction to vindicate our use of the term 'reason', it must be tolerably revisionist: it must retain all the central pre-reductive functions of the given term. (Stevenson 1944, 36-39; 1963, 11; Brandt 1979, 14-6, 126-8; Gibbard 1990, 32; Railton 1990, 159; 1993a, 282; 1993b, 316) Although any analysis will strain its concept to some extent, some do so more than others. We need those where the strains do not include the central functions we associate with the concept. From this two further points follow. First, to condemn a reduction intolerably revisionist, it is not enough to observe that we can ask questions about it. We can ask such questions for any number of reasons, including that we have

---

<sup>42</sup> This too is a disputed claim. For a recent criticism of the analytical answer see Hatzimoysis (2002), pp. 11-13 and Gampel (1997). Horgan and Timmons (1990-1); (1992ab) argues against the idea of rejecting the classical theory of meaning, as does in a different, and to my mind, more convincing way Elstein (ms).

never thought about these things before. To be relevant, our questions must be accompanied by a diagnosis that shows the link between them and the function the attempted reduction allegedly misses. Second, since revisionism is a comparative enterprise, we may conclude that we should not discard a particular theory even if it cannot account for an important function. This can happen, for instance, if, after we have examined all the available attempts, we find none that can account for the particular function while being no better, or in fact worse than the attempt under examination. This is where ‘real’ revisionism may come into play: we can declare that if this is so, we should perhaps do without that bit of meaning (however important it is), and reform our present use of the term.

### **2.3.2. Rosati’s objection**

In what follows I will focus on one important function: the inherent recommending force of our reason-claims. We have inherited this problem from non-cognitivist critics of naturalism. (Stevenson 1944; 1963; Hare 1952, 83-91; 1963; Gibbard 1990, 11-2, 15-7, 19-22; Darwall, Gibbard and Railton 1997, 117-8) It was them who pointed out that the OQA does reveal something important about proposed naturalist accounts of ethical terms. When we ask questions about a naturalist account what we are driving at is the recommending force of the given account. We ask the question ‘But is it really good for me?’ or ‘But is it really a reason?’ because we doubt that we should do, approve of or recommend whatever has the naturalistic property identified with goodness, reason and so on. From this non-cognitivists concluded that such questions remain open because claims about ‘good’, ‘reason’ or ‘rational’ express a non-cognitive state of mind and it is logically open to us to desire anything.

The naturalists we are dealing here with, however, are ready with an answer. They claim that the success of a naturalist answer turns on the *content* of the belief expressed in our reason-

claims. One part of this answer we have already touched upon. It understands the non-cognitivist challenge as making a motivational claim: normative judgments are inherently recommending in the sense that they necessarily motivate and since beliefs regardless of their content cannot motivate, non-cognitivism must be true. When they are not tempted to reject the idea that there is a necessary connection between motivation and judgment, naturalists typically claim that normative judgments with suitable descriptive content are such that under normal conditions they give rise to a corresponding desire, which may or may not lead to action (to repeat, the best such argument is offered by Smith 1994, 177-180; 1997, 305-7; 2004a, 43-6). Here, however, I am not interested in this argument. What concerns me is not the connection between belief and desire but the supposedly suitable content of the former. There is something that makes this content suitable for recommendation other than that - forming the content of a belief - it can give rise to a desire under normal conditions. This is the second part of the naturalist answer and this is, I believe, where it fails.

Connie Rosati has argued that certain accounts of the agent's good cannot preserve the recommending force the term carries. We can extend Rosati's argument to the case of reasons because such an extension changes nothing in the argument itself (in fact, authors in this tradition typically hold that reasons are value-based where value is understood in terms of desires). (Rosati 1995a, 311-2; 1995b, 53-5) The account that Rosati has in mind is often called the ideal advisor theory because of its description of condition *C*, i.e. the condition we use for selecting desires that are reason-giving for the agent. The view characterizes the condition as one in which an idealized self ('the advisor') tells his non-idealized self (the actual agent) what he would want for him to want were he to find himself in the actual condition and circumstances of his. The question is how we describe the idealized self. The two most influential formulations come from Railton (on goodness) and Smith (on reason). Railton understands the ideal advisor as someone "who is fully



and vividly informed about himself and his circumstances, and entirely free of cognitive error and lapses of instrumental rationality.” (Railton 2003a, 54) Smith agrees but adds that the ideal advisor must also have a maximally coherent and unified desire set. (Smith 1994, 151-2; 1997, 301-2) Although the epistemic and cognitive features of the advisor can be further elaborated, for our purposes this much should suffice. Let us turn to Rosati’s objection now.

She introduces the problem with an example.<sup>43</sup> Sally is wondering whether she has reason to undergo therapy. Although she doesn’t much like the idea, she is also concerned with her life as it is now. On the one hand, she loves order; on the other hand, she finds herself too orderly and admires the spontaneity of her friend, Madelyn. Part of her problem concerns what she misses by being so orderly; yet, there is a more serious question hunting her: whether to remain the sort of person who is rigid and controlled or become more adventuresome like Madelyn. That is, in posing her question Sally wants to know whether she should remain as she is now or she should rather change. Therefore, she is not interested in how she would react under idealized conditions if those conditions only concern epistemic and cognitive requirements. For, were she to stay as she is now, her rigidity would certainly inform her choice and this is exactly what she wants to avoid. In sum, when she is asking for advice, she wants her advisor not only to possess full

---

<sup>43</sup> There are other objections in the literature. Some point out that the notion of a fully informed agent is not clear or coherent. See Velleman (1988), pp. 364-371; Sobel (1994), pp. 796-810; Rosati (1995a), pp. 314-325; for a possible response see Zimmerman (2003), pp. 386-8. Bognar (ms) argues that the ideal advisor model has no resources to account for risk taking that naturally comes with our prospective evaluative or normative judgments. Finally, in her more recent papers, Rosati argues that the model (and naturalist accounts in general) has troubles with accommodating the autonomous agent pictured as possessing certain autonomy-making motives. See Rosati (2000), pp. 792-811 (on Brandt only) and (2003), pp. 510-527 (on naturalists accounts in general). It is important to note that this new problem is different from the one I present in the text. Whereas an ideal of the person is something agents aspire to in the sense that they want to bring their motives in line with it, the new problem concerns the way they can do this. In other words, the new problem is about how to picture the reflective agent, the ‘I’ who makes the decision and executes it contrary to pressing first-order desires; whereas the old problem is about the content of that decision, about the pattern that should be put into force. See Rosati (2003), pp. 511-4, 519-20. The reason why we should be aware of this difference is twofold. First, as Rosati notes, the idea that the autonomous agent can best be understood in terms of autonomy-making motives is controversial. Second, as again Rosati admits, there are naturalist accounts of agency along these lines in the literature. She mentions Sumner (1996) and Smith (1997), pp. 306 as examples (though here we should note that it is not clear what ‘rational capacities’ mean for Smith: presumably they are not motives. Smith (2004a), pp. 53, 55 note 5, and (2004b) explicitly deny this. I say more about this in Chapter IV.)

information and perfect cognitive capacities, but also to have the character traits of someone like Madelyn.

We can put this problem in a different way. (Rosati 1995b, 60-1) The role of condition *C* is to decide which desires are normative for the agent; in other words, what the agent asks is which desire should guide him in his conduct. This is Sandy's question too: she wants to know whether a desire to take a therapy would survive the test of condition *C* or not. But it is a well-known and widely accepted idea that when one asks such questions, one engages in self-reflection: one reflects upon one's desires and decides which one to follow. And it is another often shared view that such self-reflection is also a form of self-invention that gives rise to an ideal of the person understood as possessing certain character traits. It is these ideals - Christine Korsgaard calls them *practical identities* - that are most important for us. They constitute "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking."<sup>44</sup> (Korsgaard 1996b, 101-2) Sandy's enquiry, then, only brings out what we should have expected anyway: that when the agent, using condition *C*, asks which desire should guide him, he in effect asks what *sort* of person to be and what *that* would person desire. Sally is looking for an ideal of the person or a practical identity that would frame her conduct; but, at least on the ideal advisor theory, she can find no such thing.

This sets the task for advocates of the Model. Unless they want to beg important questions about reasons, they must find a way to involve an ideal of the person in their theory. But this is a

---

<sup>44</sup> We should be careful when invoking Korsgaard's idea. Although the notion of practical identity is clearly a un-Kantian element in her theory, it is not without a Kantian background. In particular, whereas Korsgaard understands practical identity as a description associated with certain desires, motives, character traits of the agent, there is also, she argues, a principle looming in the background and it is this principle, she claims, which the agent ultimately identifies with. And though this element of identification is not something Kant would have liked, the way Korsgaard understands the principle is: it is a principle not determined by the agent's desires. Hence, when the agent decides which desire to act on, she stands over and above all her desires and confronts them as a passive material. See Korsgaard (1996b), pp. 100, 103-4, 143. And this is not a picture that naturalist would be happy to embrace, as I try to show in Appendix III.

more difficult task than it may at first appear to be. An obvious solution would be to employ substantive evaluative judgments: to say that the ideal advisor *should* possess certain character traits or motives and *should not* possess others. But this is not something advocates of the Model are willing to do. Railton, for instance, while admits that the process of becoming fully informed and cognitively perfect may perturb the agent's psychology, insists that no other changes are allowed: the idealization he proposes holds a person's personality constant, permitting only the changes brought about by the process itself. (Railton 1997, 160 15n; 2003a, 58, 60; also Brandt 1979, 13) This is not surprising. The chief point of the naturalist enterprise is to determine what counts as a reason for the agent by appealing to normatively neutral conditions. If instead we employ normative judgments, we settle those reasons by prior normative commitments. This would not only beg the question from a naturalist point of view, but would also mean selecting desires on the basis of normative judgments. And this is incompatible with the Model.

Note that this issue is different from Smith's claim that a naturalist account of reasons can include normative, unreduced terms. (Smith 1994, 52) What he says is that naturalist accounts can employ such terms insofar as we are able to square such talk with a broader naturalism. (Ibid. 57) And under 'broader naturalism' he means that those terms, were we to construct a two-stage analysis for them, would turn out to refer to natural properties, namely, psychological properties of the agent. Hence, though not defined naturally, these terms are nevertheless reducible to natural properties. (Ibid. 186) An ideal of the person invites the same claim. Presumably, such an ideal also refers to the agent's psychology; hence the terms we employ in describing it would also have a naturalist reduction basis. But whereas Smith's claim is about the terms we use to describe the ideal advisor, the present point concerns the way these terms *get into* the Model: whether they do so via prior normative judgments or not. In Smith's case, for instance, the unreduced terms get into his account via conceptual analysis. Hence, once we accept the content of his analysis, there

is no need to question this part of his account; but this would not be the case were we to go along with the above proposal.

### **2.3.3. The ideal advisor account as the best idealizing approach**

There are two ways to respond to Rosati's argument. The naturalist can either claim to have found a way to incorporate an ideal of the person in his theory, or argue that he can do without such an ideal. Start with the first response. From the previous chapter we know that there are two main formulations of the Model: one that relies in some way on the agent's actual desires, the other invokes hypothetical desires. Rosati's argument only targets a particular version of the second approach, thus it is sensible to think that other versions can do better than the one she examines. However, it is not without reason that among the several hypothetical desire accounts, she picks exactly the ideal advisor theory. She does so because the consensus in the literature seems to be that the theory offers us the best account of condition *C*. Here is why.

Take the two main alternatives: Brandt on value and Williams on reason. Brandt's account of condition *C* is based on his notion of cognitive psychotherapy, which he defines as "a process of confronting relevant information, by repeatedly representing it, in an ideally vivid way, and at an appropriate time." (Brandt 1979, 113) Under 'relevant information' he means the "beliefs which are a part of the 'scientific knowledge' of the day, or which are justified on the basis of publicly available evidence in accordance with the canons of inductive and deductive logic, or justified on the basis of evidence which could be obtained by procedures known to science." (Ibid. 111-2) And under 'therapy' he means an entirely cognitive process, exhausted by facing facts and logic. No alternative methods such as "influence by prestige of someone, use of evaluative language, extrinsic reward or punishment, or use of artificially induced feeling-states like relaxation" are allowed. (Ibid. 113)

However, both conditions are mistaken. It has been pointed out that restricting information in the way Brandt does, unjustifiably compromises the intuition behind his theory: that being better informed one is better suited to find what is good for him (or, to apply this to our case, what gives him reason to act). (Loeb 1995, 9-11) Nor is the ban on non-cognitive methods of therapy more defensible. Brandt admits that there are desires and aversions that his theory cannot exclude because they were inhabited in early childhood and are not completely insensitive to relevant facts. (Brandt *ibid.* 113) To take Allan Gibbard's example, the boy who neurotically wants to keep his hands germ-free would not give up his desire to wash his hands many times every hour, no matter how vividly he represents the facts as well as the costs of his desire to himself. (Gibbard 1990, 20) It might also be the case that cognitive psychotherapy would produce desires the agent finds alien, arbitrary or crazy. Turning to Gibbard again, we might develop an aversion to dine with others provided we vividly imagine what people's innards are like. (*Ibid.*) At the same time, alternative therapies might succeed helping the agent to get rid of the desire or might avoid producing it in the first place. (Velleman 1988, 357-64; for further examples in the same vein see Hubin 1996, 36-43)

An even more serious problem is that Brandt's exclusion of non-cognitive methods makes us unable to decide in certain cases. (Velleman 1988, 357-9) Some of the desires would, once acquired by non-cognitive means, withstand confrontation with facts and logic alone. So we would like to ask whether we should have those desires. Conversely, some of our desires that would not be removed by information and logic alone might not be restored by it either. These desires are not incompatible with cognitive psychotherapy but neither would their absence be. So we would like to ask whether we should have these desires. But Brandt's theory would prevent us from asking either question and this is a clear failure of his theory. At the same time, the ideal advisor theory has no trouble with correcting these failures. Since it distinguishes between the

agent's ideal self from his actual self, it need not limit the advisor's information to what is available in Brandt's sense, nor need it restrict forms of therapy the advisor can employ to confrontation of facts and logic alone. The advisor can choose among several types of methods, including non-cognitive ones, with the exception of those - think of the first section - that would be alienating or arbitrary from the agent's perspective. The ideal advisor model is thus clearly superior to the construction of cognitive psychotherapy.

Turn now to Williams' account of reasons. Williams characterizes condition *C* as one of ideal deliberation: he claims that whether an agent would have a desire is a matter of his having the desire after sound deliberation on the basis of his subjective motivational set. (Williams 1981a, 101; 1995a, 35; 2001, 91) Although Williams' formulation is rather ambiguous, there is good reason to think that he would not restrict information and deliberation in the way Brandt does. As to the first, he is concerned with ruling out all false beliefs as possible grounds of reasons (think of his famous gin and tonic example) and it is arguable that this requirement often comes with the requirement to have all possible true beliefs. (Williams 1981, 102-4; Sobel 2001a, 469-472; though cf. Williams 1981, 103) As to the second, though Williams, in his Owen Wingrave example, rules out moving rhetoric (by Owen's family) as a form of correct deliberation, later he admits that "practical reasoning is a heuristic process and an imaginative one, and there are no fixed boundaries on the continuum from rational thought to inspiration and conversion." (Williams *ibid.* 108-9, 110) This way of depicting correct deliberation does not seem to rule out certain non-cognitive methods such as daily meditation, influence of literature or of our own self-censure and self-praise. (Velleman *ibid.* 357-8)

Nevertheless, there are problems with this theory too. To understand, we have to introduce a distinction between two understandings of the ideal advisor account. (Smith 1997, 304) On the *example* reading, the ideal advisor is seen as giving advice to himself, which then

serves as an example for the actual agent in the non-ideal world. On the *advisor* reading, the advisor advises not himself but the actual agent in his non-ideal circumstances. In my description I followed the latter reading, whereas Williams seems to endorse the first version. However, the advisor version is clearly superior. First, the idealization process through which the actual agent becomes an ideal advisor can change the agent: the desires, hence the reasons of the advisor may be different from that of the actual agent. For example, whereas the advisor may not want more information for himself, this is not true of the actual agent. (Railton 1997, 160 15n) Second, the authority of the advisor on the example reading is questionable. Since he is an ideal person in an ideal world, it is unclear what relevance his choice would have for the actual agent facing the problems of the non-ideal world. (Smith *ibid.*) He is not concerned with those problems; so why should the agent care about his ‘example’?

#### **2.3.4. Attempts to meet Rosati’s objection**

For these reasons I submit that the ideal advisor theory gives us the best account of condition *C* available in the literature.<sup>45</sup> This leaves Rosati’s objection standing and even more powerful; yet, it does not make her charge unanswerable. First, we have not yet considered the different ways

---

<sup>45</sup> This does not rule out that we can change Brandt’s or Williams’ theory in such a way that they become a version of the ideal advisor view. However, certain rationales they appeal to in defense of their theories may rule out such a transformation. Brandt, for instance, takes cognitive psychotherapy to unite the two, what he regards as most important, roles of moral philosophy. One is that a proper method should provide a standard for rational desires and aversions; the other is that it should provide a therapeutic method for discovering rational desires and aversions. The discovery is therapeutic because it is a form of (self)-therapy: the individual who undergoes the process of cognitive psychotherapy is supposed to acquire the desire or rid himself of the aversion by that very process. Therefore it might seem logical to exclude non-cognitive methods as well as unattainable information. Through the former the theorist could smuggle in his prior values or, because being too powerful, these methods could extinguish non-mistaken desires too. Unattainable information, on the other hand, would be of no real use of administering therapy to actual agents in the real world. See Daniels (1983), pp. 777 and Rosati (2000), pp. 785-6. In the case of Williams Sobel (2001b), pp. 226-33 has argued that Williams’ explanatory condition and existence internalism is inconsistent with the ideal advisor model. Others have made similar claims with respect to Williams as well as to the ideal advisor model in general. See Noggle (1999), Loeb (1995), section 3 and Enoch (2005), pp. 768-9. While I tend to agree with the claim about Brandt’s theory, I am more skeptical about the alleged disconnection of internalism and the ideal advisor view. However, I set this problem aside here as not being directly relevant for the present discussion.

the naturalist might try to incorporate an ideal of the person and, second, we have not yet considered the idea that the naturalist does not need an ideal in his theory. In this sub-section I will deal with the first strategy, then proceed on to the second and with this end my discussion.

We have left the presentation of Rosati's argument with the claim that naturalists cannot appeal to prior normative commitments as a means of inclusion. But even if this claim is accepted, naturalists still have several responses open to them. They are of two kinds: naturalists can try to show that an ideal of the person is derivable from the already existing features of their theory, or they can appeal to alternative methods of inclusion. Start with the first idea. In the case of the hypothetical desire reading of the Model, the ground for the approach is provided by the already mentioned thought that in perfecting the actual agent epistemically as well as cognitively, we also change his existing psychological make-up. In fact, in an argument that I cannot present here in detail, Rosati shows that the changes the agent goes through as a result of idealization, are very dramatic. (Rosati 1995a, 304-11; cf. also Noggle 1999) This is because the requirement that the advisor should be fully informed brings with it special forms of information processing. On the one hand, the advisor should not only have the information but he should also *appreciate* that information: the information should sink in and take hold of him. On the other hand, since assessing information also depends on the order one receives it, the advisor should consider information from all possible angles in all possible orders. These two requirements, argues Rosati, turn the advisor into someone who is dramatically different from the actual agent.

Now, the naturalist might say, if, through the idealization process, the agent changes dramatically, then perhaps these changes give us just what we need. The claim here is that the process of idealization produces just those traits in the advisor that we are looking for. But this is only an empirical thesis with no grounding in a proper psychological theory. Furthermore, to say that idealization produces exactly those traits that are desirable in an advisor amounts to taking a



stance on how the ideal personality looks like. And, at this point of the argument, this implies invoking substantive normative judgments, which would put the naturalist in a position he is not willing to embrace, or so I have argued before.

Nor does the actual desire approach do better. True, both Hubin and Noggle have candidates for an ideal of the person, namely the self understood as being composed of the actual desires they prefer. Recall the discussion in the previous chapter. Hubin gives priority to the intrinsic values of the agent, whereas Noggle prefers those desires that the agent most deeply identifies with. Both, moreover, take these desires to constitute the agent's self and thus secure his psychological identity over time (in fact, as we know, in the case of Noggle all this is built into an ideal advisor framework). (Hubin 1996, 47, 49-50; 2003, 327, 329; Noggle 1999, 314-6) The question is why this would answer Rosati's worries. After all, what Hubin and Noggle do is merely to equate an ideal of the person with the agent's actual self. But if there was anything behind Rosati's point, then it surely was included that someone like Sally can have serious doubts about even those character traits of her that constitute her intrinsic values or the core of her self.

So we turn to the second strategy. If an ideal of the person is not derivable from the theory itself and if a naturalist cannot appeal to substantive normative judgments, he should look for alternative means of inclusion. There are well-known candidates - conceptual analysis for analytical naturalists, linguistic intuitions and empirical research for non-analytical naturalists - as well as methods we may not be aware of yet. One could say, for instance, that our use of terms like 'advisor' or 'reason' follows our ideas about what features a person must possess to be a competent judge, and those features exclude certain disturbing factors and include others, or, alternatively, involve certain character traits and exclude others. If these moves are defensible, naturalists have a solution at hand. They will have an interpretation (of the ideal advisor or of what actual desires of the agent matter) that does not make use of normative judgments

concerning the inclusion or exclusion of certain motives, character traits and the like. And they may then use this interpretation to back up either the motives idealization produces or the actual desires they prefer, or to defend a separate set of character traits in addition to (or even as a critique of) the ones idealization generates or the actual desires they prefer.

I take this point. I certainly do not think that I have presented anything that would rule out the possibility of a possible naturalist answer along the above lines. It is telling in this respect that Rosati herself considers her objection not as a refutation of the ideal advisor view but more of an attempt to set a new task for it. (Rosati 1995b, 63; 2000, 810-1; 2003, 496, 527) Yet, this admission does not change the fact that at present there are no such accounts available in the literature. True, as noted in passing in the previous chapter, many naturalists find a role for certain motives in their theory.<sup>46</sup> Williams, for instance, at one point in his argument refers to ignorance and its undesirable effects on the agent's deliberation. (Williams 1981a, 104) Smith, as we saw, argues that the advisor has a maximally unified and coherent desire set as a result of the idealization process. (Smith 1994, 159-161) And Roderick Firth claims that our linguistic intuitions support the idea that the 'ideal observer' must be disinterested, dispassionate and consistent but otherwise normal. (Firth 1952, 335-344)

But none of these attempts give us what we need. Smith's description of the ideal advisor is just too thin for our purposes. A maximally coherent and unified desire set is compatible with advisors who are not, at least not without argument, authoritative for the actual agent. This kind of psychology goes well together with Stoics and Buddhists whose advices often we do not admire, since their unification and coherence are achieved by total disengagement. (Blackburn

---

<sup>46</sup> I am here not investigating whether they give any reason for such moves, nor whether they would be right to do so. Such issues, it seems to me, arise especially with regard to Williams' theory since he says very little about the meta-ethical background of his theory. For matters of interpretation see Parfit (1997), pp. 109-110; for problems with Williams' construal of his theory see Sobel (2001b), pp. 224-5.

1998, 117) And William's exclusion of ignorance does not in itself give us an ideal whose advice we would take to be authoritative, or even produce an example that we would follow. But besides this trait Williams mentions none other, in fact, he says explicitly that no further substantive constraints on the correct deliberator's psychology are allowed. (Williams 1995a, 37; 1995c, 190) Finally, though Firth's characterization is more substantial, it is so only from the viewpoint of what he intends to define, namely, rightness. Being disinterested and dispassionate are traits that are important for an observer whose job is to judge the rightness of an act, but they would be much less welcome in an account of what reasons there are for a given person to act. What we need in such an account is an advisor, not an observer and the latter, unlike the former, is interested in and perhaps even passionate about the agent.<sup>47</sup>

There is one last idea we need to consider. Although the objection we are dealing with comes from Rosati, her own preferred version of the ideal advisor theory may also provide the proper response to it. Her idea is that for *φ-ing* to be good for the agent, not only must the agent be able to care about *φ-ing* in some set of counterfactual conditions, but those counterfactual conditions must themselves answer to her concerns. (Rosati 1996, 305) That is, besides the first link between the agent's concerns and what she would choose in certain conditions, there must be a second link between her concerns and those conditions themselves. The decision, furthermore, whether the second link is secured is made in circumstances we already accept: they are presupposed by and central to our normative discourse.

Setting now aside what these circumstances are (Rosati calls them 'ordinary optimal conditions'), it appears reasonable that the epistemic and cognitive conditions the ideal advisor model operates with would be accepted there. (Rosati *ibid.* 308-9; cf. Railton 2003a, 54;

---

<sup>47</sup> In the text I do not consider another problem with Firth's proposal: that the particular method of inclusion he uses, namely appealing to linguistic intuitions is not among the favoured methods of naturalists nowadays. For a thorough criticism of the method see Brandt (1979), pp. 6-8.

Williams 1995a, 37) But the issue is whether an ideal of the person would be accepted too – and Rosati says nothing about this. As a result, her version of the theory offers more of a diagnosis of the problem, not a solution to it. It shows that that problem with the ideal advisor model as Railton and others formulate it is that it fails to secure the second link; in fact, in another formulation of the objection, Rosati says just this. (Rosati 1995a, 313-4) Otherwise, though Rosati may be right that her version best encapsulates the intuitions behind theory (a bit more about this later and in Chapter IV), it does not do much more. It offers us another means of including an ideal of the person but says nothing about how this inclusion would happen. And in lack of a properly defended account of inclusion, we cannot take Rosati's version as a response to her own objection. Again, she admits this. (Rosati 1996, 309)

### **2.3.5. Attempts to avoid Rosati's objection**

If inclusion is problematic, there is the obvious choice of trying to do without it. For such an attempt to be successful, the naturalist must argue that there is no need for an ideal of the person. Here are three such attempts: the first two concerns the ideal advisor theory only, the third is compatible with both actual and hypothetical desire theories. The naturalist can hold that the process of idealization is such that it will wash away the differences that arise from the different psychological make-up of agents. Railton suggests this when he says that since belief and desire co-vary systemically, i.e. those who have similar beliefs tend to have similar desires, fully informing agents will bring both their beliefs and their desires into line. (Railton 1997, 161 19n; 2003a, 57-8) And, as we saw, Smith holds that since idealization involves the systematic justification of desires, agents will converge on the same desires concerning a particular situation. (Smith 1994, 165; also Firth 1952, 322) If this is so, no matter how we picture ideal advisors they will always be the same, hence they will always give the same advice to their actual self.

Consequently, it is pointless for Sally to look for an advisor other than the one idealization produces: they will all tell her the same thing.

There are several problems with this suggestion. First, it relies on an unwarranted empirical claim. Nothing assures that beliefs and desires covary or that agents converge on their desires after idealization. To this, however, Smith says that the above point is conceptual not substantial, that is, it is a claim about our concept of reason and not about the world. The world might indeed disappoint us producing different ideal advisors; yet, this does not change the fact that this is not what we mean when we talk of ‘fully informed and cognitively perfect’ advisors. (Smith 1994, 166) But even if we accept Smith’s controversial claim, we can still say two things. First, we should not forget another point of his: that if all ideal advisors are the same, this means that our reasons are no longer anchored in our actual desires, which is exactly what the Model needs. Furthermore, we can point out that even if he is right and even if the world produces convergence, this will still not satisfy our needs. (Rosati 1995b, 57) For our question is which traits are important in an ideal advisor. Hence, even if advisors are the same, their advice may still not be authoritative for us because, say, they are all cold hearted, calculating or nihilistic. Of course, advocates of the view can defend the traits idealization produces. But, at this point of the argument, it is hard to see how they could do this without resort to normative judgments.

Certain of Railton’s remarks suggest another putative reason not to include an ideal of the person in our account of the ideal advisor. (Railton 2003a, 58, 60) He says that we should be careful not to distance the advisor too far from the actual agent because we need her to be someone who is *like* the agent. The idea seems to be this. If we keep the ‘non-belief properties’ of the advisor as close as possible to the actual agent, we thereby allow the agent to serve as her own ideal advisor. And since we ordinarily regard our own judgments as authoritative, at least insofar as the influence of our psychology is concerned, we would be more inclined to follow the

judgments of our fully informed self were that self to be construed in a way that does not disrupt its continuity with the agent. What drives Railton here is the intuition that generally stands behind the ideal advisor theory (it leads to motivational existence internalism, as I will explain in Chapter IV): that the agent's good or reasons should not be alien to the agent. For if they are not alien to the agent, they are also authoritative: or so the thought goes. And if we keep continuity with the agent, we will achieve just this.

I don't want to question Railton's internalist insight (I won't do that even in Chapter IV). Even so, Railton's argument does not produce the required result. Note first that there are two ways the ideal advisor account can keep continuity with the actual agent. (Rosati 1995b, 59-60) One is that whatever suggestions the advisor makes should be constrained by *facts* about what the actual agent is like: her physiological and psychological constitution, capacities, circumstances, history and so on. (Railton 1997, 142) These facts determine the agent's capacity to change, the costs of these changes and so on. (Railton 2003a, 54) And, no doubt, the advisor must take these facts into account when making a decision about what the actual self should desire. In this sense, then, any version of the ideal advisor theory should keep continuity with the actual agent. But there is another sense, and this is obviously what Railton is referring to above, in which continuity is a more controversial matter: it concerns how the facts should be *viewed* (and judged upon). What Railton suggests is that they should be viewed from an angle, which preserves as much as possible from the agent's actual self; whereas what Rosati suggests is that they should be viewed from an angle that contains an ideal of the person (which, importantly, is endorsed by the agent herself).

But it is hard to see why we should privilege Railton's viewpoint over the one proposed by Rosati. First of all, if we do so, we will lose the ability to allow the agent to pose certain questions about the reasons she has: this is just the problem Rosati's objection uncovers. Second,

the viewpoint Rosati proposes, i.e. the viewpoint that incorporates an ideal of the person is also meant to keep the continuity between the agent and the advisor: it keeps continuity with those ‘non-belief properties’ that the agent identifies with and throws out those that she does not endorse. (Rosati 1995b, 60) Third, there is good reason to think that the ideal advisor as Railton reads him would not be continuous with the agent after all. Recall Rosati’s previously mentioned, though in detail not presented, argument. It is supposed to show that the idealization not merely changes the agent but perturbs his psychology drastically. Hence, instead of getting someone who is continuous with the agent, we get someone who is dramatically different from him, who is a stranger to him. And to establish an appropriate connection between advisor and agent, the best we can do is to appeal to an ideal of the person, which the agent shares and our account of the advisor incorporates.

This takes us to the third attempt. It makes use of a simply intuition that might be taken to back Railton’s preference. For a proper presentation, let me switch from reasons to the agent’s good since it is in this respect that the idea typically appears. It goes like this. We can make a distinction between the agent’s good and personal ideals such as moral or aesthetic ideals. We can further add that while the former is attached to the agent’s present standpoint as defined by his present motivational capacities, the latter is *about* which standpoint to prefer. Rosati’s question clearly belongs to the sphere of personal ideals since it prefers a standpoint that incorporates an ideal of the person. Hence Rosati’s objection does not point to concerns about the naturalist construal of the agent’s good as it appears in the ideal advisor view. As a result, though the naturalist account does indeed leave questions open, these questions are not about the agent’s good; insofar as this notion is concerned the ideal advisor view is tolerably revisionist.

This is, I believe, the crux of the matter in the present debate. Although advocates of the ideal advisor theory do not in press mention this objection (Rosati reports that the objection was

made by Railton in personal communication), both Hubin and Noggle explicitly defend their view in this way. (Rosati 1995b, 59 note 45; Noggle 1999, 323-6; Hubin 1999, 41-2) I, however, don't think that this response settles the controversy in the naturalist's favour. As a way of ending discussion let me offer two reasons. First, as noted, the response is typically made in discussions about the agent's good. But it is not clear whether the same kind of distinction can be made in the field of reasons. It seems that insofar the justification of conduct is concerned questions about standpoints, i.e. about what sort of person to be may also be relevant and legitimate.<sup>48</sup> In fact, it is not at all clear why the claim would be true even about the agent's good. This, it seems to me, is a rather elusive matter in which decision does not come easily: it needs an argument, not just a declaration of position. Second, rejecting the idea that determining the agent's good (or reasons) involves questions about what sort of person to be is also in the business of choosing standpoints. For this is exactly to say that the preferred standpoint is provided by the agent's current motivational system. And this puts us back to the situation we have left Railton's suggestion with: that a reason needs to be given that justifies preference for the agent's current standpoint. But, as we saw, no such reason is presented.

### **2.3.6. Summary**

Let me sum up. The previous discussion has shown that (1) naturalist accounts of the Model leave important questions about reasons open because of their lack of an ideal of the person, that (2) though including such an ideal is not impossible, no acceptable account has yet been proposed, and that (3) trying to do without such an ideal is not a plausible solution to the problem. As a result, we can conclude that the Model is intolerably revisionist and that therefore the triviality charge follows. This last point shows, however, that an important shift has taken

---

<sup>48</sup> Except Hubin (1999), pp. 42 who deals with reason and thinks that such questions are preconditions to the agent's reasoning, not elements of it.



place since our first encounter with Parfit's argument. While Parfit's charge is intended it to be a knock-down argument against naturalism per se, the condition of tolerable revisionism leaves open the possibility that one day a proper naturalist account of reasons, one that is compatible with the Model, will be proposed.<sup>49</sup>

#### ***IV. Summary and what follows***

In this chapter I have attempted to refute the Model through its naturalist meta-ethical background. I started out from a suitable characterization of the Model along naturalist lines. I then presented an argument, following Derek Parfit, which claimed that naturalism, including the Model, makes reason-claims trivial. My discussion showed, however, that this is only so if we can prove that naturalist reductions are always eliminative and not vindicatory. Vindication requires, I argued, at least one condition to be met: the reduction proposed should be tolerably revisionist. In the remainder of the chapter, I showed that naturalist advocates of the Model have difficulty in meeting these conditions. Hence vindication, at least for the Model, remains an unfulfilled promise. In the next chapter I change fronts and attack the Model from a different angle. I will argue against the idea that the reason-relation must make reference to the agent's desires. Instead, I will show that desires are themselves based on reasons, which they transmit but to which they cannot add. This is the task that I turn now to.

---

<sup>49</sup> Zimmerman's (2003) idea that we should *historicize* ideal preference theories may be one such attempt. The thought is that in addition to the usual characterization of the ideal advisor we should also add a historical condition concerning the development of the advisor with perhaps an ideal educator watching over the process. Zimmerman thinks that doing so would go some way to answering Rosati's charge. See Zimmerman (2003), pp. 374, 390-2. The question is whether it would go far enough and whether we can characterize the historical condition (and especially the ideal educator, if we invoke one) without reproducing Rosati's problem (or any other problem idealization might invoke, see note 15 above).

## CHAPTER III: REASON AND DESIRE

In the previous chapter we have attacked the Model through its meta-ethical background. I have used Derek Parfit's triviality objection to argue against the Model, claiming that though it is not an objection *per se* as Parfit claims it to be, it is nevertheless effective against the Model. This time I will approach the Model from a different angle; I will see whether it is right in claiming that the reason-relation must make reference to the agent's desires. Since on the second and third version of the Model, it is desires that are responsible for giving us an account of the relation, the move to deal with the two versions on one occasion is again warranted.

### ***I. Reason-based desires: the first premise***

The Model presupposes a particular picture of the reason-relation. We know the central idea. On its second version, it claims that some consideration  $q$  speaks in favor of action  $\varphi$  by virtue of being related to a desire of agent  $A$  that the action satisfies. On its third version, we have the additional element that only desires the agent would have in condition  $C$  are those that count and that this situation is understood in terms of rationality. In either case, it is desires that carry the weight of giving us an account of the relation. There are several attempts in the literature to refute this idea. Some of these arguments – five, by number - I consider in Appendix III. I there argue that none of these attempts are successful for different reasons, reasons that mostly have to do with the differences between the three versions of the Model. This leads me to the thought that another argument is needed: this is what the idea of reason-based desires offers for us.

The idea is built up from two premises. (Anscombe 1957, 76; Dancy 2000b, 35-43; Scanlon 1998, 41-50; Raz 1999a; 1999c, 260-264; Parfit 2001, 20-5; Heuer 2004, 49-53) The first claims that desires are states that we have reason to have. The second argues that desires do not add to the stock of reasons the agent has for having them. Together the two premises entail the following conclusion: desires are based on reasons, which they transmit but to which they cannot add. And this, advocates of the position claim, means that the reasons we have to do what would subserve desires are entirely derived from the reasons to have those desires. Desires are redundant in our account of the reason-relation: the Model is refuted.

In the rest of this chapter I am going to examine this idea. We will follow a simple procedure. I will start with the exposition and defense of the first premise; this will be the topic of the present section. Then, in the second section, I will present the second premise in detail and will also consider objections that were made or that might be made against it. Here I will also separately examine the first version of the Model that we have set aside so far: Dancy's Advice Point. I do so because this view can be used to master an argument against the second premise. Finally, I will make some remarks concerning the topic that will occupy us in the next chapter: the theory of motivation.

### **3.1.1. Issues not discussed**

The claim that the first premise makes is related to many issues in contemporary meta-ethics. Therefore in the present sub-section I would like set aside those problems the resolution of which is not necessary for us to proceed. First, I won't consider the question which is the primitive: reason or value. There is an ongoing debate here, which is roughly about the following. Both camps in the discussion agree that it is the natural features of the object of the desire that give us reasons to have the desire; the question is *how* they do it. On the classical view the natural

properties give the object the property of being good, and its being good may then give us reasons to want the thing. On the other view, the object's being good is the same as its having natural properties that would, in certain contexts, give us reasons to want the thing – Scanlon calls this the 'buck-passing' account of value. (Scanlon 1998, 97) In my exposition of the first premise, I won't take side on this issue and will use the term 'reason' and 'value' interchangeably.

Nor will I deal with the meta-ethical background of the Model. Not because we have refuted it already (though, I hope, the previous chapter did make some progress in this respect), but because an appeal to it would be an extraneous substantial argument in favor of the Model and would thus fall outside the scope of this chapter. So, to make my claim more explicit, I am not going to consider the naturalist idea that the property of being good is reducible to facts about desire-satisfaction. The property of value or that of being a reason can be a natural property that is in line with the Model or it can be a natural property that goes against the Model, or it can simply be a non-natural property, or, perhaps, non-cognitivists are right and there is no property involved at all. This issue is no concern for us here.

Also, I won't deal with a typical response to the first premise according to which all reasons for having a desire can be accounted for in terms of desires: they are desire-based reasons themselves. Parfit, for instance, denies this view because he thinks that the Model has no resources to account for a particular class of reasons: those he calls 'intrinsic reasons' below. (Parfit 2001, 23) He may or may not be right about this, it is again no concern for us. Given the nature of the first premise, a defense against potential counter-examples would also eliminate objections of this kind; hence we can set this argument aside.

Finally, there is an influential codification of the first premise, which I will use only in part. I already mentioned it above: it comes from Parfit. (Ibid. 21-2) He makes two distinctions.

The first is between object-given and state-given reasons to have a desire. The former arises from the value of the *object* of the desire; the latter is grounded in the value of the agent's having the desire, i.e. in his being in the *state* of desiring.<sup>50</sup> In both cases we refer to features of the desire: there are facts about the object of the desire or of the state of desiring that make the object or the attitude valuable. This is where the second distinction enters the picture. For the features concerned can be valuable in two ways. If they are good in themselves, we are dealing with *intrinsically* valuable features of the object of the desire or of the state of desiring. If they are good only as a means to something else, we speak of *instrumentally* valuable features of the object of the desire or of the state of desiring. To take Parfit's example, my desire to end your suffering has an object that is in itself good, and it may also have the good effect of allowing you to enjoy life again. Further, my wanting your suffering to end may be in itself good, and it can also have good effects, such as your being comforted with my sympathy.

In my discussion I will set aside Parfit's first distinction between object-given and state-given reasons. Although there are interesting philosophical issues involved here, I won't attempt to settle these questions here.<sup>51</sup> I do this for a technical reason: the counterexamples that are used

---

<sup>50</sup> Speaking of the 'object' of desire might be ambiguous. One is tempted to say that it is the *thing* what we want and indeed this is the easiest way to refer to the object of desire. Nevertheless the object of the desire is often not a thing in the ordinary sense but some event, process, or state of affairs. And even when it is a thing properly speaking, it is not the thing itself but our using, consuming or having some other relation to the thing that forms the object of the desire. See Parfit (2001), pp. 21 and Scanlon (1998), pp. 379n.

<sup>51</sup> The issues are the following. There is first the question whether we can separate object-given and state-given reasons. Rabinowicz and Rønnow-Rasmussen (2004), pp. 406 argues that we cannot because to every property *P* of the state of desiring corresponds a property *P'* of the object of the desire: if a desire for object *a* would have a property *P*, then, ipso facto, *a* has (or would have, if it existed) the property *P'* of being such that a desire for it would have the property *P*. Second, there is the question whether state-given reasons exist at all. Parfit, among others, argues that state-given reasons to have a desire are in fact object-given reasons to *want* to have the desire. State-given reasons do not exist: they collapse into object-given reasons. See Parfit (2001), pp. 24; also Gibbard (1990), pp. 37 and Skorupski (ms b). Third, there is the question about the ways state-given reasons are related to action. In the case of object-given reasons the answer is simple: the agent satisfies the desire in order to get its object. State-given reasons, however, are attached to the state of desiring; hence they do not necessarily convert into reasons for action. What they always direct us to do is to *stay* in the state of desiring; whether they also urge us to satisfy the desire is not obvious. Of course, this 'staying' can also be understood as action and in this peculiar sense state-given reasons are always related to action. Finally, there is one big controversy that connects all these issues: the debate about the primacy of reason over value. Those who reject the buck-passing view typically do so by citing examples

against the first premise concern only object-given reasons (though I will also mention, when relevant, state-given reasons). Parfit's other distinction, however, invites a bit of attention already at this point. It is a somewhat misleading distinction. In one sense, the difference between instrumental and intrinsic reasons clearly exists: when we cite the reasons for having a *particular* desire we do make a distinction between these two types of considerations. Parfit's example above is a case at hand in this respect. At the same time, most advocates of the first premise (including Parfit) argue that instrumental reasons get their normative force from intrinsic considerations. Reasons for and against a desire form a *chain of justification* at the end of which stands a final reason that, according to many (including Parfit), can only be an intrinsic reason. (Ibid. 24; Skorupski ms b) Those who hold this view must then say that in another sense there are only intrinsic reasons to have a desire: every instrumental reason we may cite for having a desire is based on an intrinsic reason that establishes its normative status. Hence, whenever I will make reference to instrumental reasons in the upcoming discussion, this qualification should be kept in mind.

### **3.1.2. First counterexample: affective desires**

Let us turn now to the potential counter-examples. Start with affective desires. The idea comes from Ruth Chang. (Chang 2004, 58, 64-5) She describes these desires as cases when the agent wants something just because a particular object appeals to him. He just 'feels like' doing (or pursuing, or attaining) the thing in question. His attraction is typically immediate and

---

when the object of the desire is not valuable, yet the agent has reasons to want it. See Rabinowicz and Rønnow-Rasmussen (2004), pp. 402-4. These examples stem from the mentioned special character of state-given reasons: that they may direct us to desire something without urging us to satisfy the desire. Consequently, if there are state-given reasons to have a desire, then 'buck-passers' can say that it is only object-given reasons that are related to the goodness of the object in the right way. Also, if state-given reasons are in fact object-given reasons to want to have a desire, buck-passers can again make the needed distinction. For a good overview see Rabinowicz and Rønnow-Rasmussen (2004), pp. 408-419.

spontaneous and its object can be trivial as well as complex. But there is one common feature all these desires share: that the element of attraction, the ‘affective feel’ they come with cannot be brought under any ‘evaluative category’. The agents who have these desires do not think they have reasons to have them: they are thus living counter-examples to the first premise.

However, many if not all of these ‘feels like it’ cases can be explained in a way that is consistent with the first premise. Take the well-known phenomenon of innocent whims. When we do something on a whim, we indeed rarely give an account of why we did what we did. Normally, instead of a further explanation, we just shrug and say: ‘I just felt like it, I was in the mood to do so.’ But often this does not mean that we cannot actually explain what stands behind our desire. It is just that our whim was so curiously misplaced or strangely timed or had such a strange object that we know others would not appreciate our reasons. Therefore, in an effort to avoid further embarrassing questions, we want to shortcut discussion with the brisk response: ‘I just felt like it.’ For instance, if I fly to Oxford just to have a brunch at Christ Church, others may stand shocked at my unreasonable behavior. And then to avoid their questions or just to get rid of them, I may give the above explanation. But, if I really wanted, I could have given a proper account of my act citing considerations like that the College is beautiful, that it is trendy to have Sunday lunch there, that I can afford such trips and so on. Although these reasons may sound crazy to most people but they are reasons, even if apparent, for me.

Take now suicidal and destructive whims. Here too there seems to be an explanation available that doesn’t contradict the first premise. Again, the idea can be best illustrated with examples. While driving on a busy highway, someone may imagine – with fascination, not with horror – driving into oncoming traffic. Or someone walking on the edge of a cliff may imagine, with similar fascination, simply stepping off. For destructive whims we can begin with the example of Gary Watson. A squash player, after having suffered an ignominious defeat, desires

to smash his opponent in the face with the racquet. (Watson 1975, 210) Also, one may have a fancy to kill someone in the heat of a debate, or destroy some beautiful work of art. We may say that these whims do not provide reasons even on the Model because, say, being not entirely innocent they would be extinguished in condition *C* or because they are not among the agent's intrinsic desires. Although this reaction would, I think, be not far from the truth, there is an even better response available to us. (Hubin 2003, 325) We can question whether these whims are indeed desires. It seems that what people do in these examples is *toying with the thought* of desiring something. They 'try it on for size' and, in seeing how they feel, learn something about themselves. And these 'toyings' are not or at least need not taken to be desires, not even in the functionalist sense; hence they need not trouble us.

Another large class of affective desires can be brought under the first premise with the help of a distinction. The idea comes from Chang herself. Sometimes agents do not desire a thing for any of its particular feature but rather they want the thing *as such*. To use Chang's terminology, our being attracted to things can be feature-bound as well as feature-free. (Chang *ibid.* 80) And the same distinction can be made in terms of reasons: not only can the particular features of a thing provide reason to desire it but also the thing as such. And I think this is what often happens when we say we did it just because we felt like it; we signal the fact that we couldn't point to any particular feature of the thing as desirable: doing it as such made sense in the given situation. Chang's example about someone who has a sudden desire to turn a cartwheel while walking down the sidewalk is a case at hand. (*Ibid.* 66) No particular feature of this act appears as desirable to the agent, the act as such shows itself in a favorable light. And Chang is right: agents in such a case rarely represent the object of the desire as valuable to themselves. But this is because they cannot single out anything about the object that they value in particular, and not because they do not have the desire for a reason. Attributing them a normative thought then



becomes a matter of flexibility: whether we allow for this kind of mistake or not. I see no reason to deny the former option.

Finally, often when we refer to what we felt like doing we have a fairly definite decision situation in mind. Chang, following Scanlon's and Raz's treatment of similar cases, also deals with this idea. There are cases when we cannot make up our mind between equally desirable options. Chang, for instance, cites Scanlon's case about an agent who, standing at crossroads, must decide which road to take. Since the options are equally desirable to him, the agent does what at the moment appears to be attractive to him. (Ibid. 70-3) There are also cases with less definite descriptions. Often we just hang around in life, driven solely by what we feel like doing in the given moment. In neither case, however, need we suppose that the agent has no reason to desire the object of his choice. It is more plausible to say that there are reasons to take any option, they are just not decisive; hence the agent let his desire have its way. (Dancy 2000b, 36) To this Chang may reply that the agent has in fact *two* desires here: a reason-based desire and a momentary inclination attached to it. But this sounds implausible: how can we make sense of this separation? Once we allow that there is only one desire present, however, our initial interpretation takes its place. And though this tie-breaking role can still give desire a role to play in the theory of reasons, this is an issue that concerns not the first but the second premise; till then we can set it aside.

This is still not the end of the story. We should pay closer attention to what Chang says. In her account of affective desires, she doesn't deny that there can be reasons for these desires; what she points out is that there *need* not be. Her reasoning goes like this. First, the 'objective' claim that desires are based on reasons implies the 'subjective' claim that when we have a desire

we *judge* that we have reasons to have the desire.<sup>52</sup> We have so far taken the truth of this claim for granted because this is indeed how most of its advocates put the first premise. (Raz 1999a, 52-6; Dancy 2000b, 35-6) But, second, this permits us to say that even if there are conceivable reasons to have affective desires, often this is not how we, the agents *experience* them: we don't have corresponding normative thoughts. In fact, third, we can go further than this. We can say that it would be *superfluous* to suppose that when the agent finds herself with an affective desire, she must also think that she has reason to have the desire. In fact, in some special cases, attributing such a thought would actually be *impossible*. (Chang *ibid.* 66) This is what Chang's real argument amounts to.

I think Chang underestimates the resources available to advocates of the first premise. First of all, it is not obvious that the equivocation of the subjective and objective reading is indeed supported.<sup>53</sup> On the face of it, the idea is strange, though understandably invoked. It is strange because the first premise is a normative claim about the valuable states of affairs that provide us with reasons, whereas the claim under discussion is a psychological claim about what reasons agents *think* they have. It is nevertheless understandably invoked because if it can be shown that no agent can have a desire she thinks she has no reason to have, it would be awkward to insist that nevertheless there are such desires (see Raz 1999a for this sort of argument). But what if we fail to prove the first part of this claim? Can we not say that though the agent *sees* no reason to have a desire, there may still *be* reasons to have it? This is, I think, the question that both advocates and opponents of the first premise must ultimately answer.

---

<sup>52</sup> Here, again, I set aside the corresponding meta-ethical debate between cognitivism and non-cognitivism about what this judgment consists in. Chang herself distinguishes between two understandings of normative judgment. The first takes judgment to be cognitive, that is to say, to be a normative belief. This is the view she attributes to Raz. The second interprets judgment as a quasi-cognitive attitude such as a tendency to judge that one has reasons to have a desire, do or believe something (hence it is not properly speaking a judgment, it can only become one). This is the idea she attributes to Scanlon. As much as I can see, she may be right about this (I will say a bit more about this in Chapter IV), but this distinction has no real relevance for us here. Hence I set it aside.

<sup>53</sup> I thank János Kis for drawing my attention to this point.

There is a parallel debate here concerning reasons for action (and we should not forget that the reasons treated here are supposed to transform into reasons for action): that between motivational existence internalists and externalists. Here I don't want to present these positions in detail (I will do that in the next chapter), what interests me are those insights that are typically cited in favour of internalism. For they seem to support the subjective reading. Railton's thought, mentioned in the previous chapter, that the agent's good or reasons should not be alien to the agent, or Williams' well-known claim that reasons have an explanatory potential both suggest that reasons for action are considerations the agent can recognize as reasons, though perhaps only in idealized conditions. (Williams 1981a, 102) Externalists, however deny these claims, sometimes exactly on the ground that such recognition is unnecessary for a consideration to serve as a reason. (e.g. Bond 1983, Chapter 2) So the question is whether a parallel externalist position can be defended in the case of reasons for desires. For if it can be, then the existence of affective desires need not pose a threat to the first premise.

But let us set aside this issue. Even then Chang's argument need not work. Take first her claim that it would be superfluous to suppose any kind of normative thought on the part of the agent. An advocate of the premise can agree with this claim. It is nowhere near the idea behind the premise that when we desire something we *consciously* think we have reasons to do so. This would be absurd. Chang is certainly right. Many of our desires involve an attraction that is spontaneous and automatic and we often act on these desires without ever thinking about the value of their object. But, recall Scanlon's idea from Chapter I, this poses no irremediable obstacle to the first premise. (Scanlon 1998, 24) We can just invoke the background structure, which, being composed of standing normative judgments, allow us to unconsciously and unreflectively form and act on desires on the basis of reasons that are not themselves desire-

based. Hence, insofar as this is what Chang has in mind, we can accept her claim without giving up the premise.

The question, therefore, is what happens while the desire persists or when the agent, after having acted on the desire, thinks about its object in retrospect (assuming honesty and setting aside post-rationalization, though, admittedly, this is not a psychologically trivial matter: e.g. Haidt 2001 and Nisbett and Wilson 1977). And here too the first premise allows for considerable flexibility; in fact, this already follows from my internalism-inspired presentation of the subjective reading. The agent may not easily see why he desires something and can act on the desire without ever seeing it. Perhaps there wasn't enough time involved: if her desire is a real whim not just a mere toying, he may immediately act on it without any thinking. Also, some desires may be the consequence of deeply hidden, unconscious and repressed tendencies. (Quinn 1993, 249-250) To bring to surface the underlying rationale behind these desires, thus to give the agent the opportunity to endorse them, i.e. to find their object reason-giving, deep contemplation, even psychiatric treatment may be necessary. In sum, while the desire persists, the agent may not see anything in its object because that would happen only in a 'cool hour', after special treatment, thorough contemplation or further informational input; in a situation like condition *C*, for instance.

Of course, Chang can dig in her heels and insist that agents can but need not possess corresponding normative thoughts when having affective desires. But I think that at this point the debate reaches a dead end. For advocates of the first premise will in turn claim, as Chang rightly remarks, that desires in the object of which the agent sees nothing, not even retrospectively and after contemplation, therapy, better timing or more information are desires the agent finds *alien*. They are urges like Quinn's man's desire to turn on radios or Rapoport's boy who couldn't stop washing his hands, we have encountered in Chapter I. And though these desires may indeed not

be based on reasons (though more about this later), they don't have a place in the Model either. Chang accepts the latter claim, but denies that these desires are indeed urges. In the absence of an account of urges, which will come later, we can only test intuitions here. But this should at this point suffice. For I think the contrast between the two positions becomes clear when we look at Chang's examples for the other case: when it is impossible to attribute to the agent the requisite normative thought. (Chang *ibid.* 67-8) There are three such impossibilities: rational, psychological and logical.<sup>54</sup>

In the first case, a diabetic is squaring off against a piece of chocolate whose taste he is attracted to, even though from previous experience he knows that he doesn't like the taste of chocolate and, having just tested his blood sugar level, he knows that eating the chocolate would put him a diabetic coma. In the second case, a teenager is attracted to breaking the rules in spite of the fact that, being raised by a strict disciplinarian, he has the deeply ingrained belief that cutting class is taboo, which in turn makes him incapable to think that he has reason to play truant. In the third case, an eccentric philosopher has a desire for what there is no reason to have. To me who am an advocate of the first premise, the first two cases sound clearly like an alien urge. They are temptations one is fighting against and it is just as reasonable to interpret these temptations as alien intruders from the agent's point of view as the interpretation Chang suggests. The third case is trickier since it is hard to speak of a temptation the philosopher finds alien. Yet, here I would indeed suspect - as Chang expects me to do - that there is some larger project or

---

<sup>54</sup> There is a fourth case, which I don't mention in the text because, in my view, it stands apart from the other three. Here Chang makes a point that others have also noticed: infant children don't possess the concept of reason, hence they cannot form normative beliefs. See Chang (2004), pp. 67; also Copp and Sobel (2002), pp. 258 and Mele (2003), pp. 78-9. But I don't think this is particularly troubling. First, we can follow Scanlon and deny that these children have desires. See Scanlon (2002), pp. 340. This would require us to distinguish desires from the appetitive states that infants and animals possess, that is, to make a distinction within the functionalist theory of desire. Alternatively, we can do without introducing new distinctions by saying that infant children and animals, whether or not they have desires, are simply not susceptible to rational guidance. That is, they can be moved by their desires but they are just too young to be guided by reasons: they haven't reached the age of reason, so to speak.

background rationale behind the philosopher's rejection of all reason-giving objects. In other words, something that would give him reasons to desire things he has no reason to desire. Of course, everyone can make up his mind whether this is the right interpretation; all I want here is the contrast between the two philosophical approaches.

### **3.1.3. Second counterexample: hedonic desires**

Turn next to hedonic desires. They are the likings or dislikings of our own present conscious states such as a desire for a cold shower, an aversion to the sound of squeaking chalk, or an aversion to the touch of velvet. Parfit thinks that these desires are not based on reasons. (Parfit 2001, 26) We only have to go through the potential grounds of these desires to see what led him to this thought. First, we can safely assume that the object of these desires is not intrinsically valuable: the feeling of cold water on our skin and the sound of squeaking chalk are not things we regard as valuable in themselves. Then the question arises whether they are instrumentally valuable. This is possible but contingent. It might be that I desire to pull down the chalk on the blackboard in order to draw the attention of my class. Yet, nothing guarantees that we find similar considerations for every hedonic desire. Similarly with state-given reasons: the state of desiring these sensations can hardly be intrinsically valuable, while reference to the effects of desiring is a contingent matter. The first premise is compromised.

At this point, a defender of the premise can point out that we forgot about a crucial feature of these desires, namely that they are hedonic. That is, even if we abstract away from all possible but contingent instrumental grounds, and even if we accept that neither the state of desiring nor the object of desire is intrinsically valuable, it will still be the case that the agent has these desires for a reason: that their satisfaction is pleasurable. And pleasure, the argument continues, is a

consideration that is a reason in its own right, ungrounded in desire.<sup>55</sup> (Quinn 1993, 243; Raz 1999a, 58; Scanlon 1998, 44; Heuer 2004, 50) But opponents are ready with an answer. Already in his ‘definition’ Parfit introduced the core idea. Hedonic desires, he said, are the likings and dislikings of our present conscious states that *make* these states pleasant, painful or unpleasant. (Parfit *ibid.*) In other words, what makes the object of these desires pleasurable or painful is the desire itself, and the making relation here is understood as that of constitution: the fact that I can’t stand the sound of squeaking chalk just consist in the fact that I dislike it and so on. (Chang 2004, 76) As a result, pleasure is not a reason in its own right but is a reason that it itself based on desire. The premise is back again.

Parfit takes the truth of his claim for granted; yet, it is far from obvious that he is right. After all, pleasure and desire are distinct existences; this would be hard to deny. But then the relation between them could be of different kinds and Parfit’s idea about constitution is just one among the potential candidates.<sup>56</sup> One historically influential approach, for instance, holds that all

---

<sup>55</sup> This statement raises two questions that are not, to my knowledge, taken up by advocates of the first premise. First, it is not clear *what* figures as the reason when we say pleasure provides the reason, i.e. what ontological category we opt for (this is the element I marked as ‘q’ in the structure of practical reason in Chapter I). Since many who hold this view do not like the idea that reasons are mental states, they should try to say that it is the proposition or the state of affairs involving pleasure that serve as reason here. This is not an unprecedented view; cf. Katz (1986), Chapter 4, pp. 105-6. The second question concerns the meta-ethical background of the idea. The issue here is whether non-naturalists can also claim that pleasure provides reason for action that is not based on desire. They would have to say that the property of value is still a non-natural property that supervenes upon the phenomenology of pleasure. This is an important question because many who like the idea that pleasure is reason-providing just by virtue of its qualia are not naturalists. Cf. also Sobel (2005), pp. 442 note 12.

<sup>56</sup> There are other issues involved here. First, there is the question whether pleasure is a general concomitant of desire-satisfaction and frustration of its non-satisfaction. There is evidence that the answer is negative. See Kagan (1992), pp. 170; Raz (1999a), pp. 58-9; Heuer (2004), pp. 50; Katz (2006), pp. 22-3 and (forthcoming); also Schroeder (2004), pp. 27-35 who brings examples from neuropsychology in addition to the cases I mention here. I may be unaware that something I strove to achieve was realized as a direct result of my activity. Due to my lack of awareness I did not experience pleasure; yet, my desire was satisfied. To claim the opposite, one would have to claim that satisfaction of desire is intended in a psychological sense, i.e. the agent must *feel* that his desire was satisfied. But this is certainly not the case. Satisfaction here is understood in the logician’s sense: the question is simply whether or not the state of affairs that form the object of the desire obtains. As to the other side of the relation, it is also possible to experience pleasure without preceding desire: pleasant surprises, for instance, are not necessarily preceded by the agent’s desiring the object of the surprise. Nor need frustration follow the non-satisfaction of desire. For instance, if the opportunity for satisfying a desire does not arise, the agent need not feel frustration over this. Second, there is the question whether pleasure is always among the reasons to have a desire. However, it is easy to find real life examples that confirm the opposite conclusion. Many of our desires, especially altruistic ones, are based

pleasurable experiences share a phenomenological commonality, which we call ‘pleasure’. This common core and pleasure, moreover, are only contingently related, both metaphysically and causally. That is, it is possible both in some possible world as well as in our actual world that a pleasurable experience is not liked by the agent. Consequently, the value and reason-giving power of pleasure lies in this phenomenology and not in the agent’s desire. Parfit clearly rejects this idea. His claim about constitution puts him at the other end of the spectrum: he assumes a particular theory of pleasure, one we might call desire-based.<sup>57</sup> (cf. also Parfit 1984, 493) This theory holds that the only thing common to pleasurable experiences is that they are liked by the agent, while they are occurring. Accordingly, in its account of pleasure’s value and reason-giving power this theory attributes a central role to desire: it is the desire that confers value on pleasure and makes it reason giving.

Parfit, however, nowhere defends his choice. For a proper defense we have to turn to David Sobel. Sobel’s strategy is to show that the desire-based theory is the best available account of pleasure. His first target is the idea I described above, which I will call the phenomenological account. His rejection rests on two reasons. One is the traditional objection that no matter how carefully we introspect our mental life we cannot find the common phenomenological state the alternative account claims to exist. (Sobel 1999, 230; 2005, 444) But his real problem is not this. (Copp and Sobel 2002, 271-2; Sobel 2005, 444-5) On the phenomenological account, he claims, it can happen that a phenomenological state occurs without the agent’s having a favourable response to it. This, we saw, is indeed the case. And, Sobel goes on, advocates of this account

---

on object-given intrinsic reasons other than pleasure. The desire to end famine in Africa, or the desire to help the drowning boy are all instances of this phenomenon: it is the fact that action would bring famine relief, or save the boy that gives us reason to desire it and not the fact that it would bring us pleasure or save us pain. For further examples see Sidgwick (1907), pp. 49-52; also Quinn (1993), pp. 243.

<sup>57</sup> Alternatively, Parfit could argue for a hybrid account on which pleasure is constituted by the agent’s desire in one case but is not so constituted in another case. But this would be just another choice from the possible theories of pleasure; in fact, a very difficult and unusual choice. Hence in the text I assume that he would go for the simpler and more general desire-based account of pleasure.



claim that the state gives reason to the agent to experience it even in this case. This, again, is true. But, Sobel says, this is implausible. For this implies that an ordinary sensation, which pleasure is taken to be on this picture, is reason providing even when it is not liked by the agent. The analogy here, Sobel claims, is with sensations like pins and needles or the taste of chocolate or of strawberry ice cream. But to hold that sensations like these can provide reason for agents to experience them regardless of the agents' response to them is something we say "only when joking", he says.

However, even if we accept Sobel's rejection of the first alternative, we could still opt for a theory, which attempts to strike a middle ground between the phenomenological and the desire-based account. The idea is to accept that pleasure and desire are inseparable: an experience is pleasurable (if) only if the agent is desiring it while it is occurring. But then add that this is only because desire is taken as a response to or evidence of pleasure's value where pleasure is still regarded as a phenomenological state common to all pleasurable experiences. The result is a version of the idea, mentioned in Chapter I, that desire can co-vary with valuable states without itself conferring value on them. As Scanlon puts it, "When we have reason to bring about an experience in virtue of its being pleasant, what we have reason to bring about is a complex experiential whole that involves, say, having a certain sensation while also desiring that this sensation occur. So these cases remain ones in which the quality of the experience (considered broadly) is a reason to bring it about, rather than cases of having a reason to do something because it will fulfill some desire." (Scanlon 2002, 340; cf. Chang 2004, 64, note 13)

But Sobel is not happy with this more concessive approach either. His refutation takes the following form. (Sobel 2005, 449; cf. Trigg 1970, 121-3) He argues that if the desire-condition is introduced the phenomenological state cannot ground the reason because it has no content other than being a state that is desired (by 'content' Sobel must mean phenomenological, not

propositional content). His explanation is this. Reasons do not supervene on their grounds, i.e. the requisite phenomenological state can occur without being desired by the agent: hence no pleasure, hence no reason. This is a problem, he goes on, because we don't know what, in the absence of desire, could ground the reason. All we know is that the phenomenological state counts as pleasure and provides reason when desired but does no such thing when not desired. On closer inspection, therefore, the second approach collapses into the desire-based account.

To this, however, an advocate of the second approach can say that Sobel is mistaken: there is desire-independent content here provided by the phenomenological state that this approach also invokes.<sup>58</sup> The reason why I think Sobel does not consider this response is because he thinks it only produces an insolvable dilemma for the responder. (cf. Sobel *ibid.* 451) If, in response to his claim, we say that the phenomenological state does have content other than being desired, we will be forced to endorse the phenomenological account. We will have to say that the phenomenological state provides reason even in the absence of corresponding desire. If, on the other hand, we deny that the phenomenological state has any such independent content, we will have to accept Sobel's conclusion. The first option, as has been argued previously, is implausible, whereas the second collapses into the desire-based theory. Hence, in the absence of powerful rivals, the desire-based account is the best available theory of pleasure. The first premise is in trouble again.

---

<sup>58</sup> Sobel here clearly repeats the argument of Sobel (1999), pp. 232-3, but *that* argument was premised on the fact that the phenomenological view of pleasure is rejected. Sobel there was considering a view according to which pleasure is not a feeling state but is merely a non-sensory dimension of experience, i.e. that it has no phenomenological content, which is inseparably related to desire. He attributed this view to Kagan (1992), pp. 172-9 and Katz (1986), Chapter 2, 47-8, Chapter 4, pp. 105, 112, rightly in the first case, wrongly in the second, in my view. Once we introduce a phenomenological element, however, we will give desire-independent content to the concerned phenomenological state thereby making ground for the response considered in the text.

There are problems with both parts of Sobel's argument.<sup>59</sup> Start with his rejection of the phenomenological account. His first point is that our failure of introspection shows that there is no phenomenological state common to all pleasurable experiences. But today's science and philosophical research both have resources to remedy this failure of introspective psychology. (Katz 2006, 1.3, 2.3.4) Without going into details, there is scientific evidence that introspection of affective, as opposed to, for example, sensory experience is especially prone to errors of omission. And, as we shall soon see, there is good reason to think that the phenomenological state we identify with pleasure is best understood as a feeling-state. Furthermore, instead of endorsing the early empiricist, Lockean view that we form the concept of pleasure *solely* by ostension to mental items discovered exclusively and known exhaustively by introspection, we can opt for a direct reference to pleasure, a shared but fallible ability preserved since early childhood. Finally, we can borrow Ned Block's distinction between phenomenal consciousness and cognitive awareness. The idea would then be to identify pleasure with 'bare' pleasure, i.e. an immediate phenomenal experience, instead of our consciousness of pleasure. If this is accepted, failure of

---

<sup>59</sup> In addition, there is a problem with the positive part of Sobel's argument, namely his endorsement of the desire-based theory. For the kind of desire pleasure is related to must be a desire that the agent has *while* the experience is occurring. It is therefore a *future* desire considered in relation to the act of bringing about the relevant experience. And, remember, the Model only invokes actual or hypothetical desires of the agent but future desires are not admitted as reason-giving – or, at least, this is what we have assumed so far. Scanlon (2002), pp. 339 notes this problem claiming that though this extension of the Model is 'quite coherent' (?), it does eliminate the connection between reason and motivation, which advocates of the Model typically want to maintain. And Parfit (2001), pp. 21 seems to suggest that for him the extension is unacceptable because he thinks that reasons that are allegedly given by future desires are actually value-based reasons. This also explains why Parfit, who is otherwise a defender of the first premise, allows for the exception of hedonic desires. Both Chang (2004), pp. 77-9 and Sobel (2005), pp. 454-5 take up this issue, but their responses are not convincing. Sobel's brief remark is just unclear to me, whereas Chang's more detailed reasoning does not address the real issue. By referring to personal correspondence with Parfit, she interprets the problem as that of extending the Model to desires directed at present states as opposed to ordinary desires, which are directed both at present and future states. Then she shows, to my mind convincingly, that we cannot really distinguish these two types of desires. Yet, she does not, except for a short footnote on Scanlon, consider the question whether future desires directed at future states should also be admitted in the Model. In the footnote mentioned she only says that the same strategy can be employed here too; yet, she does not actually provide the argument. Hence, until this argument is provided, I regard this problem as not yet adequately answered. (We should also note that in her discussion of Parfit Chang takes for granted the desire-based theory of pleasure, a surprising move given her earlier rejection of Sobel's argument on the ground that something like the second, Scanlonian approach to pleasure is right.)

introspection poses no problem since by construction we don't have direct cognitive access to phenomenal consciousness.

Sobel's implausibility charge fares no better. First of all, in his treatment he identifies garden-variety sensations, such as pins and needles or taste of chocolate, with pleasure. But among those who adhere to the phenomenological view of pleasure many deny that the two are the same; and others, who otherwise are not advocates of the theory, also reject the analogy (for a good overview see Katz 2006, 1.1. and note 3). The more accepted view is that pleasure is a simple and undescribable feeling in momentary consciousness. (Katz 1986, Chapter 2, 47-8; 2006, 1.1) And if we make this distinction between sensation and affective experience, it is not necessarily implausible to hold that pleasure gives reasons for agents to experience it just by virtue of its intrinsic nature, i.e. just by the way it *feels*. Furthermore, Sobel also forgets to distinguish the non-hedonic aspects of an experience from its hedonic aspect. But as C. D. Broad has pointed out, it is possible to claim that the value of a pleasurable experience stems not solely from its hedonic aspect (which Broad takes to be a quality but we need not to), but rather from the *combination* of this aspect with the other non-hedonic aspects (making up, I take it, some kind of an organic experiential whole). (Broad 1930, 232-7) Hence the claim is not that e.g. the taste of chocolate understood as a sensation in our mouth is valuable without our favorable response to it. Rather, the idea is that it is valuable because it combines this sensation with the hedonic and other non-hedonic aspects of the same experience.

Finally, there is a way the phenomenological account can integrate the second reading. It can do this with the help of a distinction between immediate liking and full desire. There is evidence that 'liking' and 'wanting' are different neural core processes where the former is not necessarily conscious (though may be phenomenally conscious: see Katz 2005 referring to Block's distinction above), but may enter into larger ones, such as the latter, that are. (Katz 2006,

3.3 referring to Berridge 2004) And Leonard Katz points out that liking, unlike wanting is a pre-intentional state, which need not have an object, and which can but need extend to, intentional and object-bound, desire. (Katz forthcoming; 2006, 2.3.3 and note 35; Scanlon 2002, 340 also seems to suggest this when he says that the appetitive states of human infants and animals are different from ordinary desires) Of course, this distinction between liking and wanting takes us beyond the widely-shared functionalist theory of desire since these states presumably have the same direction of fit. Nevertheless, I suggest that, for the sake of argument and further insight, we assume for the moment that the distinction can be made.

Then two further ways of understanding the phenomenological account becomes open to us. In one case we understand pleasure as an experience that contains within itself an immediate liking of itself by itself.<sup>60</sup> (Katz 1986, Chapter 2, 48; Chapter 4, 104 may suggest this; though cf. Katz 2006 note 35) Here it is a question how such a self-referential liking can be objectless. But we can explain away this apparent contradiction by assuming that the liking is also slippery or potential, lacking object sometimes.<sup>61</sup> Alternatively, though this requires more by way of metaphysics and may make the view collapse into the second interpretation below, we can deny the distinction between liking and being liked. In the other case we identify pleasure with a stance of affective openness, welcoming or immediate liking, which can but need not be self-referential as the immediate liking of the first interpretation. (Katz 1986, Chapter 5, 134-5; 2006, 2.3.3) Both versions are useful for two purposes. They help us account for objectless but pleasurable moods

---

<sup>60</sup> Katz takes inspiration for this view in Sidgwick's remark that pleasure is a feeling, which, when experienced by intelligent beings, is at least implicitly apprehended as desirable. See Sidgwick (1907), pp. 12; Katz (1986), Chapter 2, 48 note 36; also Christiano (1992), pp. 274-6. Interestingly, Sobel, on the basis of the same passage, takes Sidgwick to be the primary representative of the desire-based account. He does so because Sidgwick introduces the above account by reference to his failure to introspectively find a phenomenological state common to pleasurable experiences. See Sobel (2005), 444. Already in the text, however, I note that this may not be the right interpretation of Katz's thinking. In fact, in personal communication he made it clear to me that he no longer endorses this position.

<sup>61</sup> An unpublished draft of Katz (1986), Chapter 2, 48, note 27 contained this idea but was later abandoned. Again, just like in the previous case, Katz told me that he no longer accepts the view.

as well as the pleasures of children without the capacity for propositional representation, which a purely intentional account such as Sobel's desire-based theory cannot accommodate. But for us the more important use is that they provide us with another way to explain why a momentary feeling-state can be valuable without the agent's favorable response (i.e. full desire) to it.

Sobel's rejection of the second approach also raises questions. The idea, recall, was to maintain that pleasure is an identifiable phenomenological state while making it strongly bound up with desire. Sobel's criticism started from the assumption that we can nevertheless imagine cases when the state occurs without the responding desire. But this requires explanation, an account of the inseparability involved in the relation between desire and pleasure. To keep contrast, Sobel must think that while on the phenomenological account the relation is contingent both metaphysically and causally, on the second approach the contingency involved is only metaphysical. That is, though in our actual world pleasure and desire never come apart, there can be a possible world where this is not the case.

But it is not clear why the relation cannot be metaphysical. Such 'tracking' views can allow for a metaphysical relation, at least it is not obvious why this would be impossible; the question is why this is not possible in the present case. Perhaps, this wouldn't work with an account that requires desire always to occur when the requisite phenomenological state is present. But we can say that desire need not be a response while the feeling is occurring: it can be mediated through reflection and appear only after (and, perhaps, only hypothetically) the feeling has occurred. In other words, we would have something like a dispositional theory of value with desire tracking the intrinsically valuable feeling-state called pleasure in some idealized condition *C*. (cf. Kagan 1992, 175-7) Again, it is not obvious why this form of tracking cannot be an instance of (weak or strong) supervenience.

There are two further ways of restoring metaphysical inseparability. First, on both approaches that distinguish between liking and desire, pleasure does not seem to be separable from liking because it contains this liking within itself (first version) or is identical with that liking (second version). (cf. Katz forthcoming) Second, Scanlon suggests that desire plays a role in pleasure by *affecting the experience itself*. (Scanlon 2002, 340) Sobel interprets this as the claim that though the initial phenomenological state has no content independent of desire, it gains such content as a *consequence* of being desired. In this way, we would re-introduce the phenomenological element in the picture, while also accounting for the role of desire. Sobel, however, thinks that it is meaningful to ask whether the new state itself must also be liked. And this question, he argues, reproduces the original dilemma: either we answer ‘yes’, and then the new state by itself won’t be able to ground reasons or we answer ‘no’, but then we have to face the implausibility of the phenomenological account. (Sobel 2005, 450-1)

The issue here is whether Sobel’s question is indeed meaningful. Probably, his idea is that if we introduced desire in the first case, the same logic forces us to do this again. The question is what this logic is about and whether it indeed applies to the new case. I think the logic is simple: we need desire in the first case because, having no other content, this is the only way we can ensure that the phenomenological state has the ability to provide reasons. But then this also shows the problem with Sobel’s reaction. For in the second case we do have such a content, hence it is hard to see why we would need the help of desire again. Sobel’s re-introducing the question is anything but an obvious step at this point. In lack of further explanation, however, Scanlon’s idea remains a viable response to Sobel’s objection.

### 3.1.4. Third counterexample: urges

Finally: urges. I have previously admitted that these desires may not be based on reasons but added that then they have no place in the Model either. Here I will explain why this is so. Start with some examples, most of which are familiar to us. Rapoport's boy's desire to wash his hands, Quinn's man's urge to turn on the radio are known for us from Chapter I. (Rapoport 1989; Quinn 1993, 236) Then there are the cases, which I mentioned there but did not spell out. They are Watson's mother who has an urge to drown her bawling child in the bath, Ayer's kleptomaniac woman who steals everything silvery she sees as well as Harry Frankfurt's unwilling addict who would do anything to get the drug, though he effectively hates himself for this. (Watson 1975, 210; Ayer 1954, 20; Frankfurt 1971, 87) In all these cases, the agent who has the desire admits that the object of his desire is devoid of value. Nor is the state of desiring, intrinsically or instrumentally, valuable; in fact, as we shall soon see, the opposite verdict is true.

One might say, however, that we forgot about something. For there is one consideration that holders of these desires sometimes appeal to: the pleasure that the satisfaction of the desire brings. If this is so, and if we accept my previous argument about the right account of pleasure, we found the reason we were looking for: urges turn out to be reason-based after all. But in response one can point out that if there is any pleasure involved here, a description of that pleasure must make reference to the urge itself. This is because urges are desires the satisfaction of which brings no pleasure other than the relief of the discomfort of having the urge. Urges, or at least some them, have an essentially uncomfortable phenomenological character, and this is why it is pleasurable to satisfy them. This shows that the pleasure referred to here is different from the pleasure one experiences when an ordinary desire is satisfied.<sup>62</sup> For in the case of ordinary desires

---

<sup>62</sup> I take this idea from Sidgwick (1907), pp. 46. It should be noted, however, that the context of my discussion differs from Sidgwick's: this is why I make no explicit reference to him in the text. What Sidgwick is dealing with is

Attila Tanyi, THESIS, Page 120



the expected pleasure is always conceived as a positive future result of satisfying the urge: it is about *getting* the object of the desire. Whereas in the case of urges, the pleasure attainable through their satisfaction, is represented purely negatively as compared to the state of desiring: it is about *not* having the urge any more.

We thus get the following picture of urges. Satisfaction of the urge may bring no pleasure and then there is no consideration to refer to as grounding the urge, given the lack of other, object- or state-given reasons. I don't know if there are any such urges, but I would like to leave the possibility open. Alternatively, there might be some pleasure involved in satisfying the urge but a full account of that pleasure must make reference to the urge itself. In both cases we get the same result. Since urges are not based on reasons, which are not themselves provided by the urge, if we are willing to accept that we have reason to do what satisfies urges, we have no choice but to admit that this reason is provided by urges themselves. The first premise is compromised again.

We have two options. Either we reject the claim that urges are not based on reasons, which are not themselves provided by the urge or we argue that they don't provide reasons for action. The first option seems to be readily available. For we are dealing with pleasure here and the upshot of the previous sub-section was exactly that pleasure is a reason in its own right, ungrounded in desire. True, here we deal with a peculiar kind of pleasure. But the peculiarity of this pleasure is a result only of the *causal* relation between pleasure and desire, and says nothing, if our previous argument was successful, about the *value-conferring* relation between the two. Hence here too we could say that even in this admittedly special case, it is the pleasure,

---

the general view that desire always has pleasure as its object, and within this he considers the particular Lockean theory that desires are painful states to be in and that therefore their motivational impulse consists in nothing more but an effort to remove the pain. I agree with Sidgwick in his rejection of both ideas. My point in the text is only that there are desires, namely urges, in the case of which this Lockean understanding of pleasure is correct. But I in no way claim that urges have this pleasure as their object and I certainly do not think that Locke's view is true of desires in general.

understood as a simple and undescrivable feeling in momentary consciousness that provides the reason and not the urge to which it is causally related. And the alternative but equally reason-granting picture of pleasure even allows, as we saw, that the causal relation is not contingent.<sup>63</sup> Hence, regardless of the metaphysical status of the causal relation, our 'old' idea is back: urges are no exceptions to the first premise since they are desires that are based on the reason their pleasurable satisfaction provides.

Of course, one might point out in response that I have admitted the existence of urges the satisfaction of which brings no pleasure. But I also signaled that this is only a theoretical possibility, and even if there are such urges, they are probably not numerous. In any case, there is a better response available to those who want to question my claim.<sup>64</sup> They can point out that even if pleasure in principle provides reason, this is not so in this case: for the agent who has the urge does not see the invoked pleasure as reason providing. Those who have urges see them as alien intruders, impulses they cannot control; they do not endorse them on the ground that their satisfaction brings relief. This is exactly because of the peculiar nature of the pleasure involved. Those who have these urges know that satisfying the urge is a futile enterprise, the only purpose of which is to end suffering, while knowing that the desire will reproduce itself and then everything starts again. Hence these people feel encaged, seeing only the futility of having such a desire, the never-ending suffering (or, at least, annoyance) it causes. If they nevertheless endorse

---

<sup>63</sup> Not everyone agrees. Fred Schueler in his treatment of Schiffer's account (see note 15) seems to suggest that the causal connection between the desire and its phenomenology, hence the pleasure its satisfaction produces is only contingent. See Schueler (1995), pp. 91-97.

<sup>64</sup> There are alternative responses, which are consistent with the objective reading of reason-based desires. One is the Broadian idea (or, more precisely, the idea I have attributed to Broad) that what carries value is an organic whole that is composed of the hedonic and non-hedonic aspects of the sensation in question. Along these lines we can argue that certain sensations, mental states etc. are not valuable - even though they are pleasurable - due to the non-hedonic aspects they have. Think of malice, for instance. The other idea endorses particularism and decomposes aspects of the situation into reasons (favourers/contributors), enablers and intensifiers in the way mentioned in Chapter I, note 7. Then we can claim that though in certain cases pleasure is present, due to the absence of necessary enablers it cannot serve as a reason for the agent.

the relief the satisfaction of their urge brings as reason providing, they become starkly different from the agents in our examples: they become *willing* addicts and compulsives. This is a possibility, true; but this is not what we are interested in here.

In other words, the idea is to make use of the ‘subjective’ interpretation of the premise we have encountered while discussing Chang’s argument: that the claim that desires are based on reasons implies the claim that people who have these desires *judge* them to be based on reasons. This does not happen in the case of urges, consequently, they are not reason-based; the first premise is in trouble again. But this is not so. We can accept the subjective interpretation as well as the claim that it applies in the present case without doing any damage to the premise. For the premise is about desires that provide reasons, while urges, understood in this way, are not like that. In the absence of any endorsement by the agent, these urges are the mental equivalents of physical reflexes, purely functional states, to quote Quinn. (Quinn 1993, 236-7) They are not intentional states, i.e. attitudes about some content for which reasons can be given.<sup>65</sup> (Chang 2004, 69) As a result, they cannot provide reasons for action; they are irrelevant our purposes. (Quinn *ibid.* 247; Raz 1999a, 51) And advocates of the Model agree. They regard the existence of urges as a *challenge* to the Model and employ idealization or other methods to filter them out. (Smith 1994, Chapter 5; Hubin 2003, 326-8)

---

<sup>65</sup> This is important because one might want to interpret urges along the lines of Stephen Schiffer’s notion of reason-providing desire. Then an urge would be understood as a self-referential intentional state the sole object of which is the removal of the discomfort its possession consists in. See Schiffer (1976), 199; Platts (1991), pp. 73 also notes the self-referential character of certain desires. But while the feature that makes these desires self-referential is one they share with urges – namely that they have an essentially uncomfortable phenomenological character - these desires do have intentional content and what makes them peculiar is exactly that we need both their content and phenomenological character to identify them. See Schiffer (1976), pp. 199. Whereas Chang, as I emphasize in the text, denies that urges have content at all in addition to their phenomenological feel. I don’t know if there are indeed self-referential desires in Schiffer’s sense. The examples he gives are those I categorize as urges in the text such as cravings or bodily appetites. If I am right, however, then these desires lack intentional content, which Schiffer claims they have. But even if we were to accept Schiffer’s view of them instead, the claims I make about the reason-providing status of pleasure as well as about the reason to satisfy the urge because it distracts the agent from living his life remain.

This picture can be further sophisticated. We can admit that in certain cases there is reason to satisfy the urge, while maintain that this reason is not provided by the urge nor is it a reason the urge transmits. We can say that since it is uncomfortable to have many of the urges, especially the more serious addictions and compulsions, the agent who has the urge might not be able to live his life the way he would like to.<sup>66</sup> Consequently, the agent has a reason to satisfy the urge, but this reason is not provided by the urge but by the fact that having the urge distracts him from doing other things.<sup>67</sup> (Scanlon 1998, 44; Raz 1999a, 55; Chang 2004, 74-5) Let me sum up. There are two ways to understand urges. Either urges are reason-based where the reason is provided by the pleasure that the satisfaction of the urge brings. In this case we go along with the objective reading of reason-based desires and admit that there is a reason to satisfy the urge, namely the reason the urge transmits. And, in the light of what has been just said, we can now add that this reason may be overridden by a further, not-desire based reason, namely the consideration that the agent live his life without distraction. Alternatively, we can endorse the subjective reading, admit that these desires are not based on reasons, but then claim, in agreement with opponents of the first premise, that such desires are alien and do not provide reason to act. Here, again, we can add that there may be a not-desire based reason present nevertheless, namely the reason we have begun this paragraph with.

At this point, however, a further problem offers itself. For those who understand urges along these lines also list instinctive desires, most typically the appetites like hunger or thirst

---

<sup>66</sup> Obviously, not all urges are like this since there are milder, less intensified urges that are not particularly uncomfortable states to be in, thus there is no corresponding reason of the mentioned kind present in their case. Cf. Schueler (1995), pp. 88; Schiffer (1976), pp. 202. In those cases, however, we can still stick to our original idea and insist that even these milder urges are alien for the agent and provide no reason for action, or maintain the claim that these urges are based on the reason pleasure provides.

<sup>67</sup> We might here add, following Sidgwick and Broad, that perhaps the best way to put an end to an urge is not through satisfying it by getting its object, but through suppressing it by e.g. diverting attention from it. This is then another aspect in which these desires differ from ordinary desires. See Sidgwick (1907), pp. 46 and Broad (1930), pp. 189 who interprets and comments on Sidgwick's discussion.

here. (Schiffer 1976, 199, 202; Parfit 2001, 26; Chang 2004, 69) And, one might think, there is no problem with *these* urges; they in no way resemble the urges we have so far considered. But on closer inspection we realize that this appearance is due to a failure to distinguish two different states that are conflated under the label ‘appetite’. Take the case of hunger. Hunger has a two-tier structure. On the surface we find the familiar desire to eat, which indeed stands in stark contrast with the urges mentioned earlier. But this desire *is* based on reasons, reasons that are provided by the pleasures of eating: pleasures that, unlike the relief of the discomfort of having an urge, are represented positively as compared to the prior state of desiring. The gourmand, for instance, whose primary aim are these pleasures often willfully stimulates his hunger and prolongs its satisfaction in order to get more of them. (Sidgwick 1907, 45)

But the state that lies at the core of hunger, the state on which the desire to eat is parasitic (at least most of the time: since we can conceive of a gourmand who eats even when he is not at all hungry), is not a desire that is directed at the pleasure of eating. As Sidgwick puts it, hunger is best understood as a direct impulse to eat. (Ibid. 46) Hunger, then, is not the same as the desire to eat: it is a basic disposition that depends on the physiological needs of humans. Nor is it, however, just another ordinary desire. Instead, it is a state, which is relevantly similar to the urges we have considered earlier. To begin with, satisfying this sort of ‘bare’ hunger generates pleasure but this pleasure is nothing else but the relief of the discomfort of being hungry. Sidgwick denies this on the ground that hunger is not normally a painful state to be in; it is painful, he says, only when the agent is sick or when the satisfaction of the appetite is abnormally delayed. (Ibid.) But if we accept an inclusive reading of pleasure and pain, which takes these terms to be synonyms to ‘feeling bad’ and ‘feeling good’ (for this usage see Katz 2006, note 1), then Sidgwick’s point loses ground. For on this reading even mild irritation and the relief of this feeling count as ‘pain’

and ‘pleasure’ respectively. And it is, to my mind, hard to deny that hunger and appetites in general have a phenomenology that involves at least this much of ‘suffering’ and ‘pleasure’.

If this point is accepted, all of our earlier responses become open to us. We can first try to say that the pleasure the satisfaction of bare hunger brings can serve as a reason to have that hunger. Very probably, however, we will find this response awkward and instead will agree with Chang that hunger is not the kind of mental state for which reasons can be given. (Chang 2004, 69) But then we will also have to share Chang’s rationale for this claim: that hunger is not an intentional state, i.e. it is not an attitude about any kind of content. It is not something that stands under the agent’s control and in this sense even the agent’s own hunger is alien to the agent. Hence, however surprising this may sound, hunger, being just one of our animal instinct, is no more *ours* and is no more reason-providing than the urges mentioned earlier. If nevertheless the agent endorses his hunger, this is because she finds either the pleasures of eating valuable, thereby transforming bare hunger into a proper desire to eat, or because she wants to go on with her life, which is impossible without eating, thereby invoking a reason, which is not itself based on his hunger.

## ***II. Reason-based desires: the second premise***

With this we have ended our examination of the first premise. We saw that the suggested counterexamples – affective desires, hedonic desires and urges (including appetites) – do not pose a serious threat to the premise. Most of the time, these desires turn out to be reason-based on closer inspection. And in the rest of the cases they are either not taken to be reason providing even on the Model or are reason-providing but in a way, which is consistent with a rejection of the Model.

Let us turn to the second premise now. Recall, it says the following: desires are not merely based on reasons but they also cannot add to those reasons. There are two applications. (Dancy 2000b, 36-7) First, the fact that one desires something does not imply that the agent has more reason to do the thing desired. If I love climbing for all sorts of reasons - the air I breathe, the beautiful landscape, the excitement I feel - the fact that I love climbing does not give me additional reason to climb. We can call this the *primary* application. Second, if desire is not an additional reason to do the thing desired, nor is it an extra reason to do what subserves the desire. This is the *secondary* application of the premise. Note that since one way of subserving a desire is just to do the thing desired, if we accept the secondary application, we also get the primary application through the backdoor. Both applications are disputed; let us examine them in turn.

### **3.2.1. The primary application**

The primary application says that we have no more reason to do something just because we desire it. Many have questioned this claim, however. There are two interesting cases. The first goes like this. There are two or more prospects, all are supported by reason but only one of them is desired by the agent. As to the balance of reasons, it is supposed that the reasons for the desired option either evenly match with or are inferior to the reasons for the option(s) not desired by the agent.<sup>68</sup> Some of the ‘feeling like it’ cases are good examples. Scanlon’s example with the agent standing at crossroads can be interpreted in this way, and Raz has a similar case with a hungry agent who has to choose between a banana and an equally desirable pear but only wants the banana. (Scanlon 1998, 48; Raz 1999a, 62) Also, imagine a young and extremely talented piano

---

<sup>68</sup> For completeness’ sake we should note that reasons can be evenly matched in three ways: they can be equally strong, incommensurable or on a par. The first is self-evident, the second and the third require further explanation. Incommensurability claims that the strength of reasons sometimes cannot be compared, whereas reasons are on a par when neither of them is stronger than the other while being comparable and unequal. For the former see Raz (1999a), pp. 182; for the latter see Chang (2002). Both phenomena are disputed: for a recent attempt to explain away incommensurability see Gert (2003), pp. 23-4.

player who is bored by playing piano and wants instead to become a philosopher.<sup>69</sup> He knows that his enormous talent gives him all the reason to prefer the glamorous life of a piano player to the simple life of philosophy for which he might have no talent at all. But, as a matter of fact, he is positively inclined to do the latter and finds it impossible to bring himself to pursue the former. All three cases appear to invite the same diagnosis: the agent has more reason to take the desired option just because he desires it.

There is, however, an alternative interpretation of these cases that attributes a role to desires without admitting them as reason-giving forces. It is based on the distinction I discussed in Chapter I: the role of desire as a necessary (and, perhaps, sufficient) condition for the existence of reasons (e.g. as a tracking ‘device’) and its role as an element in the favouring relation (i.e. as a truth-maker). (Dancy 2000b, 15-20, 26-7; Sobel 2001a, 473; 2001b, 233; Enoch 2005, 764; Scanlon 1998, 48-9; Raz 1999a, 64) It is only accepting the latter that would lead us to the conclusion that agents in the above examples have *more* reason to do something because they desire it. The former reading only admits that certain reasons (or all, according to those who interpret motivational existence internalism in this way) cannot *exist* for the agent unless the agent has a desire connected to them: they exist *only if* the agent has the relevant desire but they do not exist *because* the agent has the relevant desire.

Once we have this distinction in view, we can deploy it in defense of the second premise. We can say that in all the above cases we have two sets of considerations present. On one side, there are the considerations that serve as reasons for the agent: the particular features of the chosen road, the taste of pear and the pleasures of eating it, and the characteristics of a life spent with philosophy. They are, however, not based on the agent’s desire: that the agent desires them only signals that they are reasons for the agent but they are not reasons because the agent desires

---

<sup>69</sup> The example comes from János Kis and Gręg Bognár.



them. Instead, they are reasons because they are all valuable activities, experiences and aspects of objects. The other sets of considerations – the particular features of the other road, the taste of banana, and the life spent as a famous piano player – are, on the other hand, do not function as reasons at all since the agent has no relevant desire connected to them. So what we have in the above situations is not one course of action that the agent has more reason to do than another course of action, but one course of action, which is supported by reasons and another, which is not supported by reasons. And this is not a picture that would in any way be inconsistent with the second premise.

The second scenario is less easy to deal with. Here we are asked to repeat the above cases as to the balance of reasons but now we should suppose that the agent desires all options. Moreover, and this is the crucial addition, he does so to a different degree: his overall desire, i.e. what he most wants to do lies squarely on one side. Then the claim is made that the agent has more reason to do what he most desires to do just because of that. This is, I believe, the version Raz and Scanlon had in mind when formulating their examples, and others have also agreed that desire can play this sort of tie-breaking role when the reasons present evenly match. (Quinn 1993, 250-252; Bond 1983, 16-8) In this case we cannot employ the previous strategy since, by stipulation, all reasons that are present have a desire connected to them. But there is another solution open to us. Since on the present scenario one has more reason to do one thing just because this is what he most wants to do, it follows that we should give some interpretation of the *strength* of desire. At the first look, we are in trouble here because we have to make sense of a seemingly paradoxical claim. Namely that agent *A* prefers  $\phi$ -ing to  $\psi$ -ing even though he does not think of  $\phi$ -ing as in any way better than  $\psi$ -ing (in fact, he might think it worse). But it is possible to meet this task.

There are three candidates for measuring the weight of a desire. The most obvious choice is the pleasure that comes with the satisfaction of desire and the frustration its non-satisfaction brings. These are indeed factors that can tip the balance of reasons in favour of an act and/or against another. But, as we saw, these considerations are reasons in their own right: they are normatively independent factors that are often among the reasons to have the desire. Hence they are illegitimate considerations to cite as grounds for the Model. It could be that the weight of a desire is measured according to what the agent is willing to forgo for the satisfaction of the desire. For instance, if I give up a big chance of promotion in order to spend more time with my children this means that I really love them. The problem here is that if desires are based on reasons, it follows that their strength understood in this way also depends on the weight of the reasons that back them. (Raz 1999a, 61-2) That is, the things one is willing to forgo for the satisfaction of one's desire is dependent upon the reasons one has to have the desire in the first place. It is these reasons that determine whether the agent should have that desire or not and for what price, so to speak. Consequently, this alternative must also to be ruled out.

We haven't run out of candidates yet. In fact, the third alternative is the most promising. What we are doing here is looking for an attribute of desires that, through determining the strength of desire, can tip the balance of reasons in favour of the option the agent most wants. Once we see the issue in this way, the solution appears to be obvious: the desire in question can do this because it has an overriding *motivational force* propelling the agent towards one particular prospect. This would also explain away the paradoxical nature of the second scenario. If agent *A* is more moved to  $\phi$  than to  $\psi$ , then in this sense she can prefer  $\phi$ -ing to  $\psi$ -ing even though she in no way values  $\phi$ -ing more than  $\psi$ -ing. This is no doubt a controversial claim. Many reject the idea that motivating force would be identical with normative force (e.g. Parfit 1997); but I set this move aside here. Nor do I want to deny the empirical claim that the stringency of reasons

underlying the desire may not be equal in strength with the motivational force of the same desire. Although I think there is room for suspicion here too, the case of tie breaking seems to pose no trouble, and I suppose that the claim is also defensible in the case when reasons do not evenly match.<sup>70</sup>

We could still say, though, that the motivating force of the desire is normative *only to the extent* that it follows the stringency of the reasons that back the desire. This can be claimed to follow from the point made earlier: if desires are based on reasons, people who have those desires want the strength of their desires to reflect the stringency of the underlying reasons. Therefore, upon realizing that some desire moves them out of proportion to the stringency of the reasons it is based on, they will not attribute normative significance to the extra motivational force. Although this solution may appear to be theoretically attractive, this impression vanishes once we consider our examples. In general, if the reasons on all sides evenly match but the agent is more moved to take one option than others, it is strange to say that we should not allow this extra motivational force to play a normative role: that we should say that the agent ought not to take the option he is most moved to. Take the piano player example. If the piano player would do anything to become a philosopher even while admitting that the reasons for embarking on an international music career are stronger, then it would be unattractive not to allow him to fulfill his ‘heart’s desire’. There are ethical theories that are not troubled by claims like this; yet, I find such theories too demanding for human purposes.

Let me sum up. We are saying that desires can add to the reasons they are based on in virtue of the motivational force they have out of proportion to the strength of the reasons that

---

<sup>70</sup> We could say that the ‘frequency’ of the motivational force of desire can fluctuate both spatially (from occasion to occasion) as well as temporarily (from time to time). As a result, when the agent finds himself in a particular situation at a particular time it may well be possible that he is more inclined to  $\varphi$  than to  $\psi$  even though in general he does not think that he has more reason to  $\varphi$  than to  $\psi$ .

back them.<sup>71</sup> To meet this challenge we have to be able to accommodate this extra motivational force in a way that is line with the second premise. Here is a suggestion. We can say that there is a scale along which we normatively evaluate the extra motivational force of desires as appropriate and inappropriate, with the degree of appropriateness varying from situation to situation. This scale, moreover, is governed by higher-order reasons that are not themselves based on desires. For instance, in the case of tie breaking we can distinguish two types of situations. If the available options are trivial or not important - as when the agent is choosing between alternative routes or between a banana and a pear – the extra motivational force of the agent’s desire gains normative relevance. But when the available options are important – as when one must choose whether to save one drowning stranger or two – what the agent is most inclined to do does not count. In the first type of situation the higher-order reason can be that of living a spontaneous life, whereas in the second case the interest of others takes precedence.

In this way we can find a place for the extra motivational force of desires without doing damage to the second premise. Since the normative role of this force is conditional on higher-order reasons that are not themselves based on desires, desires do not make a genuine contribution to the normative analysis of the agent’s choice. Chang who also considers this sort of response is less enthusiastic, however. But what she says does not seem to me to be of much relevance. She points out that in the case of tie-breaking the extra motivational force of desire, she calls it ‘affective feel’, is relevant also in choices between careers, lovers, places to live, and so on. (Chang 2004, 85) I see no difficulty here. In the case of unimportant, trivial choices we referred to the higher-order reason of spontaneity as governing our choice. Why can we not come

---

<sup>71</sup> Raz (1999a), pp. 61 does not accept this claim. But this is because he thinks that resolving conflicts on the basis of motivational force would make practical reasoning impossible. In the text I do not take up this objection (which, by the way, would only lead to the rejection of the Model), but in Appendix III I show that it does not work (it is the fifth objection on my list there).

up with a higher-order reason here as well? We can appeal to self-realization as the operating higher-order reason, for instance, and then repeat everything what we have said above. Chang, as far as I can make out, doesn't have a reply.<sup>72</sup> What one could perhaps do on her behalf is to claim that these higher-order reasons are based on desires. But given our analysis in the previous section, they would have to be based on urges. It is unlikely that this claim could be true: the first application is defended.

### 3.2.2. The secondary application

Let us turn to the secondary application now. It claims that if desire does not add to the reasons for doing the thing desired, nor does it add to the reasons for doing what subserves the thing desired (and, as noted, one way of doing this is by doing the thing desired, so we would get the primary application back as well). On the face of it, this application looks implausible. Take the extreme case when one has no reason to  $\varphi$ . The application says that because the agent desires to  $\varphi$  and because  $\psi$ -ing makes it easier for him to  $\varphi$ , he has reason to  $\psi$ . Dancy gives the following example. (Dancy 2000b, 37) Imagine that some shameful act that would immediately bring the agent's marriage to an end but which he still has some desire to do. It seems perverse, Dancy suggests, "to claim that my desire to  $\varphi$  does add to the stock of reasons for  $\psi$ -ing just because  $\psi$ -ing subserves  $\varphi$ -ing, and so gives me good reason to do something because it subserves doing something that there is every reason not to do."

There are two ways to get around the implausibility charge. One can claim that, just as with the previous application, here too the application only covers certain cases. The problem

---

<sup>72</sup> This is not exactly so. She does give some sort of a reply. She says that "In so far as doing what one feels like is relevant to what one should do, 'feeling like it' can rationalize as an independent reason in its own right, not as a reason that is conditional on the other reasons being of trivial importance." See Chang (2004), pp. 85. I must confess I don't understand this claim. If desires are relevant only because they are *made* relevant by higher-order reasons, then they are not independent reasons, full stop. I can't find anything in Chang's remark that would meaningfully question this thesis.

with this suggestion is that examples can nowhere be found in the literature, so there is nothing to evaluate (perhaps this is because the two applications are conflated and the same examples are supposed to back both applications). The second way to defend the application is by invoking the so far neglected first version of the Model, the idea that Dancy's Advice Point articulates. To repeat, it goes like this. (Ibid. 34) What the Model claims, Dancy says, is that reason-attributions are like giving advice to a friend: perhaps his aim makes no sense by my light (even, perhaps, by his light too: so it is just an urge), I tell him what he should do *given* what he aims at. This reading of the Model, one might claim, would give us the second application; in fact, it would prove the truth of the application over the entire range of desires.

It would not. From Chapter I we know how the Advice Point comes about. We take the instrumental principle understood as a consistency requirement on human action and desire, then claim that there is reason to be instrumentally rational, and finally accept what it seems to come with such a claim, namely bootstrapping reasons into existence.<sup>73</sup> Here is the reasoning in more detail. Very roughly, the instrumental principle says that if you desire some prospect, rationality requires you to do whatever subserves that desire. If there is reason to conform to the principle, it follows that if you desire some prospect, you have reason to do whatever subserves that desire. Fully spelled out, the claim is this: the fact that  $\varphi$ -ing subserves your desire that  $p$  constitutes a reason for you to  $\varphi$ , if you desire that  $p$ . And this is just what the second application needs.

There are two ways to refute this argument. One is to endorse the claim that reason is understood in terms of rationality and deny the reverse option on which the argument is partly based. But I myself don't accept this claim; moreover, charity to the Model also dictates that we don't take such an 'easy' way out. This leaves us with the option of rejecting bootstrapping. I see

---

<sup>73</sup> We should note that there are other problems with this claim. These too were mentioned in Chapter I. The most important is where exactly the reason(s) to be instrumentally rational comes from. See Broome (2005), Kolodny (2005), section 3, and Dancy (forthcoming b) on this question.

two problems here. One is that bootstrapping is not necessarily an attractive prospect even for advocates of the Model. Recall the discussion in Chapter I. Just as bootstrapping can lead to self-justificatory beliefs, it can also lead to self-justifying desires. This is a point Broome and Quinn makes in the argument I examine in Appendix III. They want to say that this makes the Model self-defeating but, as I point out there, they need be right about this because it is not clear whether advocates of this version of the Model would not be willing to bite the bullet. Advocates of the other two versions of the Model would certainly resist the idea of self-justifying desires since they do rule out certain desires as reason-giving. But it is not at all clear that advocates of the Advice Point would also protest. After all, their original idea, put in terms of rationality, also has the consequence that all desires count as normative for the agent. So I suggest that we set this issue aside here.

Even so, there is a second problem. We just have to employ the general strategy, known from Chapter I, against bootstrapping. In the present case it looks like this. (Hare 1971; Darwall 1983, 15-17; Dancy 2000b, 42-3; Broome 1999, section 6; 2004, 29; Greenspan 1975; Hill 1992, 23-4; Blackburn 1998, 242-3) What the instrumental principle claims is that one ought not to desire some prospect *and* be unwilling to do whatever subserves that desire. Since this is a disjunctive, wide-scope ought of the form “You ought, if you desire that *p*, to do what subserves that desire”, there are two ways of satisfying it: one can do so by doing what subserves the desire or by giving up the desire. And from this complex obligation one cannot detach a simple obligation to act: there is no reason to do what subserves the desire. What reason dictates is to do what subserves the desire, if one has the desire, or not to do what subserves the desire but then give up the desire. The second application is in trouble again.

From Chapter I we know what could be the possible responses to this argument. The option of denying that detachment is indeed impossible I don't treat here since I have said

everything I can about this matter back in Chapter I. The other option, namely Kolodny's argument is more interesting, although in this case too I cannot provide a detailed discussion of the issues the argument raises. Kolodny's argument, recall, has the following structure: at least certain rational requirements are what he calls process-requirements and at least some of these requirements are also narrow-scope. Since I have explained these terms in Chapter I, I am not going to repeat this explanation here. Our question, then, is whether the particular requirement as encoded in the instrumental principle is also a narrow-scope process requirement. I don't want to question that the requirement can be given a process reading, I think the issue here is rather whether requirements of rationality involve process requirements in general. This again is a question I discussed in Chapter I, trying to show that Kolodny may be wrong on this point. So this is one thing. As to the second part of the claim, i.e. the claim that the requirement in question is a narrow-scope requirement, it is worth noting that Kolodny himself does not take the requirement to be a "simple narrow-scope requirement". (Kolodny 2005, 541) What exactly this statement means requires further examination before we can decide what the truth of the matter is.

Finally, there is Kolodny's own explanation of the normativity of rationality. The Transparency Account, as he calls it, also explains rationality in terms of reasons but without bootstrapping. Again, we know the story from Chapter I. Kolodny's idea (shared and, according to Kolodny, further developed by Scanlon) is that all rational requirements can be derived from two core requirements. These are "If one believes that one ought to have *A*, then one is rationally required to have *A*" and "If one believes that one lacks sufficient reason to have *A*, then one is rationally required not to have *A*." (Ibid. 557) Kolodny rightly points out that his account does not lead to bootstrapping because it does not claim that there is reason to be rational. It is called the Transparency Account exactly because it derives the normativity of rationality not from



reason, but only from what it *seems* to the agent is a reason: from an appearance, not from a reality. Of course, as I pointed out in Chapter I and as Kolodny himself admits, the success of the account hinges on an argument, not yet provided that indeed all rational requirements, including the one the instrumental principle articulates can be derived from the core requirements. If such an argument can be provided, and Kolodny strongly believes it can be (in fact, he refers to Scanlon as having answered most of the problems such a view encounters), then even if the instrumental principle does not give us a wide-scope requirement, bootstrapping, hence the second application still won't follow.<sup>74</sup>

Recently, however, Stephen Darwall has come up with a suggestion that might offer some hope for the second application. (Darwall 2001, 144-51; 2003, 440-2) Before ending our discussion, we should consider this argument. Darwall's idea is that integrity and self-respect can enable the agent's desires to provide reasons for acting *within* the constraints of a general ban on detachment. He gives two examples. One concerns an agent with the deeply held moral conviction that abortion lacks moral justification and who now has to decide whether to have an abortion or not. Darwall claims that even if one holds the opposite moral view one should, when asked for advice, say that if the agent really feels that way, then he has a reason not to have the abortion - and this is because having the abortion would violate his integrity. The other example is about a middle-aged daughter whose parents are trying to get her to eat broccoli, which she hates. In this case, Darwall says, she can claim that she has a reason not to eat broccoli on grounds of self-respect – for not allowing herself not to eat broccoli would be a failure to respect herself as a free and equal person.

---

<sup>74</sup> This may sound mysterious. At the point where Kolodny makes this statement he gives two examples: one is the instrumental principle, the other is the requirement of logical consistency. But he only takes up the case of the latter, arguing that it can be accounted for a disjunction of three narrow-scope requirements, and says nothing about the instrumental principle. See Kolodny (2005), pp. 541-2. Hence follows my statement in the text.

Observe, first, that Darwall's wording of his claim is misleading. He suggests that there are *some* desires, which get exempted from the ban on detachment on grounds of self-respect and integrity. In fact, his reference to integrity, especially the example about moral conviction, may suggest that this happens because these desires are so crucial that they cannot be given up. They form the agent's deepest convictions and ends and therefore abolishing them would put an end to the agent's personality. But this is certainly not what Darwall has in mind. First of all, it would go against the main thrust of his argument. As his second example demonstrates, his point is a general one that applies to all ordinary desires and aversions. It is that there is a fundamental difference between theoretical and practical reason. Unlike the theoretical sphere where we can discount the agent's perspective, i.e. give up certain beliefs, in the practical realm we are called to being true to ourselves and others in ways that can make our desires to be a source of normative reasons. Of course, as a matter of fact it might be true that the agent cannot give up certain ends of his. But this sort of psychological incapacity, as Bernard Williams has pointed out, abolishes the normativity of those ends. (Williams 1981c, 128-9; 1995b, 53) Since having such ends allows no room for failure – one cannot fail to do what one is irresistibly driven to do – they make one's reason depend on what one actually does or will do. And this, arguably, would abolish their reason-giving potential.

I conclude that Darwall's claim should be taken to be a wholly general point, one that doesn't reject the ban on detachment on the basis of incapacity to give up some end. Instead, it claims that giving deliberative weight to our desires is a form of respect for ourselves as independent persons; hence we have reason to do what subserves them *even if* we otherwise share the ban on detachment. But this much wouldn't do. In his reply, Dancy puts the point well: someone accepting the ban can accommodate claims about integrity and self-respect by regarding them as *grounds* for the otherwise disjunctive, wide-scope oughts we are concerned with. (Dancy

2003, 474-5) That is, we can say that the agent ought not to have an abortion because that would violate his integrity; and the daughter ought not to eat the broccoli because that would be a way of disrespecting herself. But the 'ought' invoked here is still a disjunctive, wide-scope ought and the ban on detachment is still in place: the Advice Point gains no significance leaving the second application intact. What we get is only a proposed reason why we should conform to a particular requirement of rationality and this is only the beginning of the story, not the end of it.

#### ***IV. What's left: the theory of motivation***

The aim of this chapter was to show that best way to reject the Model, if one doesn't want to get muddled up in debates about naturalism, is to present and defend the reverse thesis: that desires are based on reasons, which they transmit but to which they cannot add. The Model reconstructs the reason-relation in a particular way: it says that the relation must make reference to a privileged subset of the agent's desires. What we now claim is that this cannot be right, not if the two premises hold. For they claim that desires merely transmit reasons on which they are based and do not add to them. This, admittedly, is an *external* refutation of the Model; yet, if it is successful, this should not bother us too much. Together with my previous argument that naturalism fails to account for the normativity of practical reason-claims, this gives us good ground to doubt the appropriateness of the Model. This is not the end of the story, though; far from it. We still have to face an influential defense of the Model that applies results from the theory of motivating reasons in order to arrive at conclusions in the theory of normative reasons. Accordingly, in the last chapter I will try to defeat the Model by presenting a plausible cognitivist theory of motivation. If this 'feast' can be accomplished, then we will be in a good position to say that the Model should be substituted for a theory that does not ground reasons in the human desires.

## CHAPTER IV: A COGNITIVIST THEORY OF MOTIVATION

The previous chapters have established that the Model can be attacked both through its naturalist background as well as its picture of the favouring relation. Nevertheless, advocates of the Model might still claim that it articulates something crucial about our understanding of practical reason. For there is a consideration that is often cited as the savior of the Model: that only desires are capable of motivating people. In this last chapter I investigate the consequences this claim has for the future of the Model. As before, due to the similarities mentioned in Chapter I, here too I will act on the assumption that there is no need to distinguish the second and third version of the Model.

### ***1. The motivational defense of the Model***

So far we have left unexamined an influential defense of the Model. The idea is to apply results from the theory of motivating reasons in order to arrive at conclusions in the theory of normative reasons. The argument has two premises. The first establishes a sufficiently strong connection between normative reason and motivation; this is called the Internalism Requirement (IR). The second premise locates the source of this motivation in human desires; this is called the Humean Theory of Motivation (HTM). In what follows I will give a detailed presentation of these premises. This then will enable me to specify the kind of objection that I am going to present in the rest of the chapter.

#### 4.1.1. The Internalism Requirement

The Internalism Requirement (IR) is based on a particular form of normative internalism, which, following Stephen Darwall, we might call motivational existence internalism.<sup>75</sup> (Darwall 1983, 54; 1992, 155; 1997, 308) It claims that the agent *has* a reason to act (if) only if, necessarily, he has (actual or hypothetical) motives to act, the necessity being conceptual or metaphysical. The rationale that is most often cited as the background for this form of internalism is Railton's intuition, mentioned in Chapter II. According to Railton, the agent's good – here, her reasons - must suit her, must be able to engage her, thereby avoiding alienation. Normativity is located in the domain of motivation: normative force is motivational pull experienced by the agent. We also know that there are two main ways to flesh out this intuition: one that employs idealized, i.e. hypothetical motivation and one that sticks to the agent's actual motives in some way. In the first group we find attempts like Brandt's idea of cognitive psychotherapy, Williams' notion of the ideal deliberator (though, recall, he appeals to a further rationale, what I have earlier dubbed the 'explanatory constraint'), Railton's ideal advisor view and Rosati's two-tier internalism. In the other group we have Noggle's privileging those motivations of the agent that he identifies with, while Hubin opts for motivations that are intrinsic.

However, as it stands, existence internalism is potentially misleading. Misunderstanding can arise from four sources. First, existence internalism should be distinguished from

---

<sup>75</sup> There is considerable confusion about the IR in the literature, which is mostly due to the fact that a number of different internalisms are introduced under different titles. Here I only want to refer to one particular taxonomy of internalisms since it aims to be exhaustive but differs from mine in certain respects. It comes from Shafer-Landau (2003), pp. 142-5. He follows Darwall on motivational judgment and existence internalism, but applies them only to the moral case. Within existence internalism he singles out Williams' theory, which he calls 'internalism about reasons' and he also mentions moral rationalism, which he calls 'reasons existence internalism'. Within judgment internalism he lists three different views. The first is 'motivational judgment internalism', the second is 'reasons judgment internalism', the third is 'hybrid judgment internalism'. The first is just the judgment internalism mentioned above, whereas the third is the same as Smith's practicality requirement (see note 3 below). The second is somewhat unusual. It states that, necessarily, a person who sincerely judges an action right has reason to perform that action. In the text I am not dealing with this version just as I set aside the other two forms of internalism Shafer-Landau mentions - rational egoism and prudential internalism - as not being relevant for my discussion.

motivational *judgment* internalism. On this view, sincere judgment that one has reason to do something is necessarily connected to (actual or hypothetical) motivation, the necessity being conceptual or metaphysical. (Darwall *ibid.*) Here too we can distinguish two large groups depending on the kind of motivation employed.<sup>76</sup> What we might call robust judgment internalism refers to motivation that the agent experiences in any actual situation, whereas so-called moderate judgment internalism speaks of motivation experienced in some idealized situation. Moderate internalism can have the same, conceptual or metaphysical, status as robust internalism; or at least this seems to be the consensus in the literature and I am not going to question this here.<sup>77</sup> What is more important is to see that there is a metaphysical difference between judgment and existence internalism. On existence internalism motivation is either part of what it is for a normative proposition to be true of the agent, or a necessary condition for the holding of a normative fact; whereas on judgment internalism motivation comes with judgment and has nothing to do with normative facts and truths. Consequently, the two views can but need not go together. In particular, non-cognitivists deny existence internalism (since they deny that

---

<sup>76</sup> Another often cited distinction is between weak and strong judgment internalism where the former takes motivation to be overriding (or silencing), while the latter does not (the distinction can be applied to existence internalism as well but, to my knowledge, no one reads existence internalism in its strong form). However, the purposes that judgment internalism is used for, such as the argument in favour of non-cognitivism, clearly doesn't need more than the weak version. It is silent on the strength of the accompanying motivation, and even allows that it varies from agent to agent. See Brink (1989), Chapter 3; Svavarsdóttir (1999), pp. 166. Accordingly, only few philosophers hold the strong view, most notably Richard Hare (but only with regard to moral judgment) and Allan Gibbard (but he adds that he can also accept the weak reading). See Hare (1981), pp. 20-24, 214-216; Gibbard (2003), pp. 152-154. John McDowell also gives a version of the strong view (he speaks of silencing), but, see note 4 below, listing him as a judgment internalist may be a mistake. See McDowell (1978), pp. 25-9 and (1979), pp. 334-5.

<sup>77</sup> It is also an issue how one characterizes the ideal situation. Some speak of the absence of practical irrationalities, e.g. Smith (1994), pp. 61 on what he calls the 'practicality requirement', others refer to 'normality', e.g. Dancy (1993), pp. 25 or Dreier (1990), pp. 11, or use temporal concepts, e.g. Simpson (1999), pp. 203. Not everyone takes these disorders to be irrational, however and reference to irrationality can also lead to the charge that judgment internalism only reveals certain norms of rationality and poses no conceptual or metaphysical constraint (thereby extending Korsgaard's criticism of Williams to the this case). On the other hand, appealing to normality or saying that the agent is sometimes not motivated to act is too vague: these notions are underexplained and underinterpreted. But since it is not my job here to come up with a defensible formulation, I suggest that we should simply speak of the absence of general motivational disorders and set issues of formulation aside.

normative facts exist), while some existence internalists are also judgment internalists (since they hold that normative beliefs motivate in themselves or produce motivation in some other way).

The above understanding of existence internalism as a claim about the existence of normative facts and truths is itself ambiguous, however. This is because of a well-known phenomenon: that motivation may be a necessary (and, perhaps, sufficient) condition for the presence of a normative fact without being identical with or constituting that fact. As we saw earlier, many have pointed out this difference.<sup>78</sup> Dancy introduces what he calls internalism as a necessary condition for the existence of reasons as opposed to a claim about the grounds of reasons. (Dancy 2000b, 15-20, 26-7) And one very influential argument for existence internalism is the epistemic claim that the possibility of a person's caring about a thing is necessary evidence of its goodness because in the absence of such evidence, nothing could show it to be good for him.<sup>79</sup> (Lewis 1989, 121; Smith 1994, 187; Rosati 1996, 315-21; Railton 1997, 143 17n; 2003a,

---

<sup>78</sup> Although I do not list him here, Darwall also makes the distinction when he distinguishes constitutive (he later calls it 'metaphysical') and non-constitutive existence internalism along the lines set out in the text. See Darwall (1992), pp. 152, 157-160 and (1997), pp. 308-9. He, however, gives a different example for non-constitutive existence internalism than the ones I use in the text. He introduces a new type of internalism, he calls it 'perceptual internalism', which holds that it is impossible for a person to know directly or perceive the truth of a normative proposition without being moved. According to Darwall, this view – he associates it with the intuitionism of Samuel Clarke and Richard Price and the sensibility theories of John McDowell and David Wiggins – is a form of existence internalism rather than of judgment internalism because it speaks of direct knowledge of normative facts as opposed to normative belief or sincere normative assertion. I find Darwall's claim unclear. On the one hand, reference to knowledge as such need not make perceptual internalism different from judgment internalism since one can take knowledge to be nothing else but justified belief. Later in the text I will explain why the particular understanding of knowledge employed here, namely perceptual knowledge makes the required difference. Darwall, however, gives no such explanation. On the other hand, it is also not obvious why reference to knowledge would make this internalism a version of existence internalism. Again, later in the text I will explain why this may be so. But given that existence internalism is a metaphysical claim about properties and truths, the required connection does not seem to me to follow from the mere fact of cognitive contact with normative facts.

<sup>79</sup> The other philosophical rationales for internalism that Rosati mentions also seem to be compatible with a non-constitutive reading of existence internalism. This is because this reading keeps the idea that something can only be good or reason-giving for the agent if she can care about it, and three of the four remaining arguments – from judgment internalism, on the basis of the 'ought' implies 'can' principle, and of autonomy – seem to be based on this view. See Rosati (1996), pp. 310-3, 320-4. The only exception may be the argument she dubs metaphysical since here the claim is that normative facts exist only because there are creatures who can respond to them. But, as Rosati herself notes, this in itself does not lead to a constitutive form of existence internalism since one can still hold that normative facts are non-natural (as Moore did), or even claim that non-natural normative facts and certain sensibilities that have both an affective and perceptual aspect are interdependent (as sensibility theorists do). Finally, note that Williams' explanatory constraint mentioned in the text (and in previous chapters) that normative reasons

62-3) But such an epistemic grounding leads directly to ‘tracking’ internalism, to the claim that the agent’s motivation only indicates the presence of normative facts and nothing more. (Sobel 2001a, 473; 2001b, 233; Enoch 2004, 764) And the tracking version of the Model, as I have emphasized throughout this dissertation, is not what we are interested in.

This is still not enough. Even if we make the previous distinctions, we can still get confused. The points that concern us now were originally made with respect to Williams’ theory but they are relevant for all those idealized accounts that make reference to rational deliberation. Williams, recall, argues that his theory is sub-Humean because, though it refers to hypothetical motivation, this motivation must relate to the agent’s actual motivational set since it is this set that constrains her deliberation. But, as Korsgaard has pointed out, one can understand rationality as involving the disposition to be moved by genuine reasons; in other words, the disposition to be moved by certain reasons is constitutive of rationality. (Korsgaard 1986, 11; cf. Velleman 2000c, 172-4) William’s internalist constraint therefore does not exclude any theory because we first have to decide what reasons are genuine before we assess their motivational power: the constraint voices content skepticism not motivational skepticism, as Korsgaard puts it. (Ibid. 23) The other, also familiar point comes from Smith. He claims that rational deliberation includes the systematic justification of desires and this process, he argues, would lead us to converge on the same desires about a given situation no matter what actual desires we started out from. (Smith 1994, 164-174)

It is not my task to evaluate these proposals. Both arguments trade on the idea that reason should be understood in terms of rationality, and though in previous chapters I have claimed that this idea is not without alternative (nor am I sure that it actually gives the right reading of Williams), we don’t need to go into details here. The idea gives us the third version of the Model

---

must have an explanatory potential is also compatible with the non-constitutive reading. For what makes reasons able to explain the agent’s action is their connection to motivation but this in itself says nothing about the metaphysical status of that connection.



and, as before, I am not going to question its rationale. This takes us to the last issue. It points out that normative internalism cannot straightaway be applied to the sphere of morality. In the moral case, existence internalism is given in the form of *moral* reasons, whereas judgment internalism speaks of *moral* judgments. However, to connect this to normative internalism we need a further premise, which is called *rationalism*. (Nagel 1970, Chapter 3; Brink 1989, 37; Parfit 1997, 103; Smith 1994, 62; Svavarsdóttir 1999, 175; Shafer-Landau 2003, 143) With respect to existence internalism, it claims that if it is right for an agent to act, then there is a reason for that agent to act in those circumstances. And in the case of judgment internalism, it says that if an agent judges that it is right for him to act, then he also judges that he has reason to act in those circumstances. Rationalism is needed to establish the truth of moral existence and judgment internalism. Hence these internalisms do not form a subset of normative internalism and must be kept separate from it.

#### **4.1.2. The Humean Theory of Motivation**

If the IR in one form or another is true, we still need to find out what moves people to act. This is where the Humean Theory of Motivation (HTM) enters the picture. On the HTM the following three claims are taken to be true. First, in order for an agent to be motivated a desire and a suitable instrumental belief must always be present: these two mental states are necessary for the occurrence of motivation. Let us call this the *Existence Criterion* (EC). Second, the agent's motivation always consists in the co-presence of a desire and an instrumental belief: these two mental states constitute the agent's motivation. We can christen this the *Motivational Criterion* (MC). Finally, third, desire and belief are distinct existences: they are distinct mental states in the sense that they are independently intelligible from each other. Let us call this the *Independence Criterion* (IC). The HTM needs all three theses. It is not enough if desires are omnipresent in

human action since it could still be the case that what motivates is not a desire but some cognitive state (which shows that the MC involves the EC). And even if what motivates involves a desire, it could still happen that as a mental state the desire cannot be prone apart from some cognitive state with which it together comprises the agent's motivating reason (which shows that the IC is distinct from both the EC and the MC).

Although the HTM pictures the agent's motivating reason as consisting of two distinct elements, it holds that the desire component has a dominant role: it is the one that provides the motivational push necessary for action. Desires are essentially motivating: they always provide some motivation to act whenever they are present and do so in their own right. Beliefs, on the other hand, can be present without motivating and even when they do motivate they don't do so in their own right: they only channel the motivational force arising from the desire to the action that satisfies the desire. This asymmetry can be accounted for in two ways. On the causal reading beliefs are passive, causally inert states of mind, hence it is difficult to see how they could cause anything into being. But this is certainly too strong. Although the asymmetry follows the Humean picture in which belief/reason can only find the means for desire/passion but cannot oppose it, this only implies that belief cannot cause *action* but says nothing about its causal powers in producing *mental states*. Or at least this seems to be the position of modern-day Humeans as well as anti-Humeans (Cohon 1988, 103, 105-6; Smith 1994, 73, 160, 179-181, 193, 213-4; Hubin 1996, 43; Parfit 1997, 101-3, 117; Dancy 1993, 8-9, 12; 2000b, 79; Shaver 2006, 5-6; though cf. Cuneo 2002), though it may or may not be an accurate position of Hume's own views about the subject.<sup>80</sup> I will say more about the former issue later in the text; here, however, this much should suffice.

---

<sup>80</sup> Opinions differ. Dancy says that Hume himself appears to attribute some causal power to belief when he pictures it as a lively idea and contrasts it with imagination, whereas others think that such an admission would be incompatible

Attila Tanyi, THESIS, Page 146

That we don't need the causal reading is also underlined by the other account of the asymmetry, which is compatible with attributing causal power to belief. (Smith 1987; 1988ab; 1994, Chapter 4) It is based on premises we are mostly familiar with. The first is the dispositional or functionalist account of belief and desire: one who has a belief that  $p$  is disposed to lose that belief in the presence of a perception with the content that *not*  $p$  (downward/mind-to-world/thetic direction of fit), whereas one who has a desire that  $p$  is disposed to bring about that  $p$  in the presence of the same perception (upward/world-to-mind/telic direction of fit). The second is the teleological view of action: to act intentionally is to be in a goal-directed state. And the third is the claim that to have a goal is to be in a state with which the world must fit: it is to be disposed to bring about some result in the world. These three ideas are then taken to entail that, first, desire and belief are distinct existences (hence the IC), that, second, intentional action must be accompanied by a desire (hence the EC), and that, third, desire must be the active part of the agent's motivating reason (hence the MC). We can call this, following Jay Wallace, the *teleological argument*. (Wallace 1990, 359)

All three elements of the teleological argument are controversial. We know this from previous discussion; only the third premise can be unfamiliar for us. Not everyone is happy with the dispositional account of desire, some reject it (Copp and Sobel 2001; Schroeder 2004, Chapter 1; Platts 1979, Chapter 10), others, we saw in Chapter I, only suggest refinements of it (Velleman 2000a; Humberstone 1992). The teleological account of action is also not universally accepted (e.g. Stocker 1981; though cf. Smith 2004a, 161-163), even though the teleological argument only needs it in its weak form. It only says that every intentional action has a purpose, not that every event that has a purpose is an action: having a purpose is a necessary but not

---

with Hume's anti-rationalist argument. See Dancy (2000b), pp. 79; Stroud (1977), pp. 157; Parfit (1997), pp. 105-6 and Radcliffe (1999), pp. 101-3 for the opposite view. There is no need for us to decide this issue here.

sufficient requirement for an event to qualify as intentional action. Finally, the idea that to have a goal is to be in a state with which the world must fit, though perhaps the least controversial, also has its opponents. (Humberstone 1992, 63; Bittner 2001, 13-4) In this chapter, however, I am not going to question these premises. I do this partly out of charity to advocates of the HTM, partly because in the end I will offer a theory of motivation that does not require any such move, but mostly because discussion of any of these issues would take us too far, often beyond the theory of motivation.

#### **4.1.3. The argument**

With the Internalism Requirement and the Humean Theory of Motivation at hand we can draw up the following motivational defense of the Model:

- (1) The IR is true: a consideration acquires the status of a reason through its relation to the agent's motivation (normative existence internalism is true);
- (2) The HTM is true: a motivating reason is composed of a desire and a belief with the desire being the dominant force due to its upward direction of fit;
- (3) Hence the Model is true: normative facts are identical with or constituted by facts about the agent's desires.

As always, there are two ways to reject this argument. One can refute either or both of the premises or deny the inference from the premises to the conclusion. The first premise is rather controversial both in its constitutive and non-constitutive reading, but I am not going to question it here (for attempts see Bond 1983, 32-41; McDowell 1995; Millgram 1996, 203; Parfit 1997, 102; Svavarsdóttir 1999, 172; Dancy 2000b, 154-7; Zangwill 2003, 146; Shafer-Landau 2003, Attila Tanyi, THESIS, Page 148

176-189; FitzPatrick 2004). As to the success of the inference, it of course does not work if we read the first premise in its non-constitutive form. It is, however, more difficult to deny it when, as above, it is understood as a claim about constitution, though, we saw, there are some who nevertheless deny it. (cf. also Brink 1997, 8n; Heuer 2004) But, again, I won't deal with these arguments.

Instead, in the remainder of the chapter I will concentrate on the second premise: the Humean Theory of Motivation. I will show that the HTM does not offer us a plausible theory of motivation and should be substituted for a properly formulated cognitivist theory. This is not a novel approach either (there isn't much novelty to be found here in general); indeed, I will arrive at my favoured theory through the examination of other attempts.<sup>81</sup> In doing so I will touch upon some of the issues that concern motivational judgment internalism (since I will ultimately claim that normative beliefs motivate), but will mostly deal with problems specific of motivational theory. Then, as usual, at the end of the chapter I will give a summary of the argument presented. In this case, however, this summary will also mean the end of my investigation insofar as this dissertation is concerned.

---

<sup>81</sup> There is one attempt that I am not going to consider: the denial of psychologism, i.e. the idea that motivating reasons are mental states. This attempt, more precisely its end-product, is familiar for us from Chapter I. It requires us to show that (i) motivating reasons are identical with the agent's reasons and that (ii) due to certain features of these reasons they cannot be but states of affairs. In Chapter I, I have tried to show that (i) is true and this claim will appear later in my present discussion too; yet, for all I have said there and will say here, I don't take its truth to be confirmed. Establishing (ii) is even more controversial. For those who argue for this position see Dancy (1993), Chapter 2; (2000b), Chapter 5; Garrard-McNaughton (1998); and Bittner (2001), Chapters 4, 7; for an attempted refutation see Mele (2003), Chapter 3, especially sections 4-5. Consequently, since the rejection of psychologism would take us deep into debates about the ontology of motivating reasons and the proper form of action-explanation, I won't deal with it here. My question will therefore be confined to asking if motivating reasons are mental states, what those mental states are.

## ***II. Refuting the HTM***

### **4.2.1. Nagel's pure ascription theory**

Let us start with the Existence Criterion. Although at first it may appear to be uncontroversial, not everyone shares this opinion. Here is a famous passage from Thomas Nagel:

“The claim that a desire underlies every act is true [...] only in the sense that *whatever* may be the motivation for someone's intentional pursuit of a goal, it becomes in virtue of his pursuit *ipso facto* appropriate to ascribe to him a desire for that goal... It may be admitted as trivial that, for example, considerations about my future welfare or about the interests of others cannot motivate me to act without a desire being present at the time of action. That I have the appropriate desire simply *follows* from the fact that these considerations motivate me; if the likelihood that an act will promote my future happiness motivates me to perform it now, then it is appropriate to ascribe to me a desire for my own future happiness. But nothing follows about the role of the desire as a condition contributing to the motivational efficacy of those considerations. It is a necessary condition of their efficacy to be sure, but only a *logically* necessary condition. It is not necessary either as a contributing influence, or as a causal condition. “

(Nagel 1970, 29-30)

Moral and prudential desires, Nagel claims, are only logically necessary conditions of action: they aren't necessary either causally or motivationally. And it is because they are necessary only in the sense that action is always accompanied by a desire to carry out that action. That is, it is appropriate to ascribe to the agent a desire for that goal no matter what *in fact* motivated him - even if no antecedent desire was present among the grounds of the action. In short, Nagel claims that in certain cases desires are mere *logical* consequences of the *fact* that one is motivated to act

by one's belief that some action is prudent. Such a desire is, to borrow Bond's term, a 'logical shadow' that has no independent psychological existence. (Bond 1983, 12) The EC is refuted.

Nagel's 'pure ascription' view (I borrow the label from Dancy 1993, Chapter 1) entails a hybrid theory of motivation. According to Nagel, we can have two kinds of scenario in explaining human motivation. One in which the desire is the motivator and another when the belief itself motivates with the desire being ascribed as a mere 'logical shadow'. But hybridity has its price: Nagel must explain why these beliefs are special. Unfortunately, he says nothing about this, so we must act on our own. The natural candidate is reference to the *content* of these beliefs. However, it is unclear whether we can demarcate moral and prudential beliefs just by virtue of their content. And if we cannot, which seems clearly to be the case with prudential beliefs, we end up with the awkward claim that the motivational capacity of certain beliefs changes from situation to situation. (Dancy 1993, 21; 2000b, 93-4) But even if such a demarcation is possible, an appeal to content will still not offer the explanation Nagel needs. In the face of the teleological argument, reference to content just doesn't seem relevant to the explanation of the motivating potential of moral and prudential beliefs. After all, other beliefs too have content, so we need an account of what it is that moral and prudential content have and, say, evaluative content lacks.

There are two other ways of completing the Nagelian enterprise. According to the first, the missing account of the peculiarity of moral and prudential beliefs can be found in the other beliefs that accompany them. The idea is that both beliefs exist in a *nexus* of beliefs that somehow endow them with motivating force.<sup>82</sup> There are many problems with this suggestion. To begin with, the beliefs that accompany moral and prudential beliefs must be ordinary beliefs about the natural world; hence on Nagel's hybrid view they are themselves motivationally inert.

---

<sup>82</sup> I owe this suggestion to Edmund Henden.

But then it is difficult to see, first, how their co-presence can explain the motivational capacity of moral and prudential beliefs and, second, why it is that their ‘ability’ is only activated with regard to moral and prudential beliefs and not to other, say, evaluative beliefs. An answer to these questions may be to invoke some sort of a special relationship among the beliefs concerned, for example, some kind of organic unity they form together. In fact, this may go as far as involving other beliefs in picture, such as evaluative ones, thereby broadening Nagel’s rather restricted view. Yet, we would even so be short of an explanation that gives an account of the special relationship referred to.

The third suggestion makes use of the *phenomenology* of our motivational experience. It points out that before theorizing about motivation most of us would find it plausible that certain beliefs can move us all by themselves without the help of any desire. And these beliefs are typically those Nagel mentions: prudential, moral and perhaps evaluative beliefs. Examples abound. When someone has a duty or responsibility to do something, or when it is prudent to do something one often acts without wanting to act – or, at least, this is what, upon inquiry, the agent reports about his state of mind. If there is a desire present in such situations, it is normally a desire that pulls the person in the opposite direction, appearing more as a temptation than an attractive prospect. (Schueler 1995, 30; Shafer-Landau 2003, 123) Other examples can also be cited. When we make inferences on the basis of some evidence, it is unlikely, though, of course, not unimaginable that we do so because we have a desire to make correct inferences. (Scanlon 1998, 35-6) Finally, there are well-known cases when someone is mistaken about her desires. In Shafer-Landau’s example the agent goes to law school because this is what his father and grandfather did and because this is what everyone expects him to do. But in school he is bored, does desultory work, the library enervates him and so on. One summer he then takes a job as a carpenter, loves it, quits law school and becomes a professional carpenter. (Ibid. 125) It seems



plausible to suppose that the agent went to school and stayed there for a while out of respect for his family, or to meet social expectations or just because he thought he had a desire to become a lawyer. No desire is needed to explain his action, and indeed there was none present.

There are two issues here. One is whether there can be desires present in these cases, the other is whether their presence is plausible and/or needed. The answer to the first claim is clearly affirmative. Unless we endorse the phenomenological account of desire – the view that to desire is to have a feeling or sensation of some sort (Stroud 1977 Chapter 6; Platts 1979 Chapter 10; 1981) – we should find no trouble in attributing a desire to the agent, even though it is not one he can discover by introspection. And those who like Smith accept the functionalist theory of desire, are uniform in their rejection of the phenomenological view (for Smith’s argument see Smith 1994, 104-11) This, of course, is a substantial debate, which is far from over. Yet, it is charitable to advocates of the HTM if we suppose that the phenomenological view is defeated, since this allows them to go on.<sup>83</sup> This then takes us to the second claim: that it is just not plausible and/or superfluous to attribute a desire to the agent. As Shafer-Landau rightly points out, the explanation that invokes no desire is more natural and simple in the above cases. (Ibid. 123-6, 140) Hence the burden of proof is on advocates of the HTM: they must show that their position is right, i.e. that there must be a desire with real psychological existence present whenever intentional action takes place. (Ibid. 124)

Shafer-Landau’s insistence, however, seems idle. For we have the teleological argument at hand, which claims to support just the above conclusion. Now, as noted, the argument’s premises are controversial and Shafer-Landau also goes some way to argue against the teleological view of action. (Ibid. 135-6) But he admits that there can be arguments that show the

---

<sup>83</sup> Thanks to Daniel Friedrich who pushed me hard to admit that there is much to be going in favour of the phenomenological reading. My concession in the text is largely due to his efforts.

view to be the right position; in any case, discussion of these issues would take us beyond the theory of motivation. Shafer-Landau, moreover, accepts the rest of the argument, so the question arises what he takes to save Nagel's position. His solution is to invoke a distinction by Fred Schueler between 'desire proper' and 'pro attitude'. (Schueler 1995, 29-31, 34-5) The former refers to mental states that are real with the only thing common to them is that they are intrinsically motivating; otherwise, they can be states with a strong phenomenological presence but can also be 'calm passions' that have no such presence. 'Pro attitudes', on the other hand, only denote the fact that the agent was motivated. It is a placeholder sense of desire that is inspired by Nagel's point (Schueler makes this explicit): it refers to anything that motivated the agent including desires proper but also beliefs, convictions and the like.

Shafer-Landau thinks that by using the distinction we can neutralize the force of the teleological argument. We can admit that whenever an agent acts intentionally, she wanted to act because the 'want' referred to here is the pro attitude sense of desire and is thus compatible with the pure ascription view. (Ibid. 138-140) I don't share this optimism. Smith puts the argument in terms of mental states that invariably accompany intentional action and shows that, if we are willing to follow his logic, these mental states must be desires. It is hard to see how we get from this position, which uses states, hence real existences - what else can have a direction of fit? (cf. also Smith's understanding of pro attitude in Smith 1994, 117) - to the view that there is nothing psychologically real and desire-like present in some instances of intentional action. I certainly don't see how Schueler's distinction, which merely conceptualizes Nagel's point, can muster such a shift in focus. No doubt, in Shafer-Landau's own dialectic the distinction is helpful since up to that point he was only considering the support motivational phenomenology can give to the idea that motivation does not need the presence of desire. The distinction gives a framework to

his position; but I don't see how it advances *our* discussion, which has started out from Nagel's observation.<sup>84</sup>

#### 4.2.2. Motivated desires

I conclude that a plausible theory of motivation must meet the EC: it must have a place for desire as a real psychological existence. Therefore those who want to refute the HTM must reject either the Motivational or the Independence Criterion. For a well-known attempt of the first kind we can turn to Nagel again. He writes:

“Many desires, like many beliefs, are *arrived at* by decision and after deliberation. They need not simply assail us, though there are certain desires that do, like the appetites and in certain cases the emotions... The desires, which simply come to us are unmotivated though they can be explained. Hunger is produced by lack of food, but is not motivated thereby. A desire to shop for groceries, after discovering nothing appetizing in the refrigerator, is on the other hand motivated by hunger. Rational or motivational explanation is just as much in order for that desire as for the action itself [...] If the desire is a motivated one, the explanation of it will be the same as the explanation of [the agent's intentional pursuit of his

---

<sup>84</sup> The situation would be different were Schueler to provide good reasons why we should accept his distinction even in the face of the teleological argument. But I cannot find any such reasons in his book. There are two claims he makes in introducing the distinction. See Schueler (1995), pp. 31-4. The first is the familiar claim about the failure of introspection, which, as noted in the text, is useless unless one adopts the phenomenological view of desire. The second point concerns practical reasoning. Schueler claims that in order to describe the content of our reasoning in terms of desires, we must employ the pro attitude sense of desire since in considering what to do we often appeal to considerations that have nothing to do with our desires (e.g. our responsibilities, convictions and the like). There are three problems with this claim but they are all rooted in the same thing: that the content of practical reasoning concerns normative reasons, not motivating reasons (a fact that, of course, Schueler is also aware of: see his discussion of the practical syllogism at Ibid. 97-108). On this basis, advocates of the HTM can make three responses. First, as I make it clear in Appendix III, reasoning need not be about desires, that is, it need not mention desires even if one adheres to the Model. Second, even if one thinks that reasoning must mention desires, this itself is a normative claim, not a description of how people actually reason. As such it requires argument of course, but there is no antecedent reason to suppose that no such argument can be given. Finally, third, since reasoning is about normative reasons, there is nothing contradictory in claiming that while normative reasons are not provided by desires proper but by pro attitudes, it is only desires proper that can move people to act.

goal], and it is by no means obvious that a desire must enter into this further explanation.” (Nagel 1970, 29)

Although this claim appears together with the view we have just examined and though Nagel doesn't distinguish them, they are separate claims. (Schueler 1995; Dancy 2000b, 93-4) His new objection concerns the plausibility of the MC. It is the following.

Motivated desires are the product of the agent's decision and deliberation based on considerations the agent faces in the course of deliberation. Since these considerations could be of all sorts, there is no antecedent reason to suppose that all desires have at least one desire among their motivating grounds. In particular, prudential and moral desires need not be motivated by other desires. For example, in Stephen Darwall's case Roberta after watching a documentary about exploited workers acquires the desire to join the boycott against responsible firms. This desire is entirely a result of Roberta's awareness of the unjustifiable suffering of the workers: no desires are present among its motivating grounds, not even an abstract desire to relieve suffering. (Darwall 1983, 40) As a result, when we explain why Roberta joins the boycott we make no reference to her desire. The relevant material for the explanation of her action is the same as the material that explains her desire: her awareness of the unjustified suffering of the workers. Although Nagel's idea admittedly needs more sharpening, its point, even in its present form, remains clear: the MC is refuted.<sup>85</sup>

The problem with Nagel's proposal concerns his concept of the 'rational and motivational explanation' of desires. (Lenman 1996, 293-4; Dancy 2000b, 82-4) There are two ways to understand this idea: causally and rationally, i.e. in terms of motivating reasons. Nagel's

---

<sup>85</sup> As it stands, Nagel's view defines motivated desires too narrowly and the classification it offers is not exhaustive. Both problems are handled well in Schueler's (1995), pp. 17, 21-8 discussion. Since my criticism does not concern these details of the theory, I set these modifications aside in the text.

distinction between producing and motivating a desire shows that he is aware of this. But merely asserting the rational reading is not enough. For the HTM has the teleological argument on its side claiming that belief cannot motivate because it has the wrong direction of fit. And this argument is compatible with the causal reading: even if a belief causes a desire, the desire will still have the right direction of fit necessary for action.<sup>86</sup> Therefore if Nagel wants the rational reading, he must face the teleological argument head on; since he does no such thing, his objection is irrelevant for the HTM.

Let me put this point in a different way. Here is a more detailed version of the teleological argument, put in terms of motivating reasons. (Smith 1987, 58; 1994, 116) The beliefs that supposedly explain desires are in fact the motivating reasons to have those desires (first premise); to have a reason is to have a goal (second premise); to have a goal is to be in a state with which the world must fit (third premise); the state with which the world must fit is a desire (fourth premise); hence the beliefs that explain desires are in fact desires or complexes where the desire has the dominating role (conclusion). The consequence is this. Desires can be motivated in two ways. There is first the idea that belief can cause a desire into being; in the present case this is taken to mean that one may arrive at a desire through reasoning that is not helped by pre-existing desires of the agent. Although the second part of this claim is more controversial than the first (as I will soon explain), it is not something advocates of the HTM must reject. The second way desires can be motivated is by motivating reasons: this is what the ‘new’ teleological argument is

---

<sup>86</sup> In fact, an advocate of the HTM can put all this in the framework of rationality. He can understand rationality as, among other things, involving assumptions about the causal connection between certain beliefs and corresponding desires. Smith, recall, has this idea in mind when he speaks of a rational tendency towards coherence between normative beliefs with certain content and desires. See Smith (1994), pp. 179; (1995), pp. 168; (1999), §6; (2004a), pp. 53. And recently Robert Shaver has argued that Sidgwick’s theory of moral motivation is best understood as claiming that moral beliefs cause corresponding desires in every rational being. See Shaver (2006), pp. 6-7. Note, moreover, that understanding motivated desires in this way answers the charge of hybridity (see Dancy 2000, pp. 81 who uses the objection against this position as well). The response is simple: it is normative, evaluative or moral beliefs that produce desires because it is only in their case that the content of the belief combined with the property of rationality gives rise to a desire. See Shaver (2006), pp. 9-10 who makes this response on behalf of Sidgwick.

about. Since Nagel gives us no reason why we should interpret his views as a version of the rational reading, we have no ground to reject the HTM on the basis of his discussion.

I think we can also explain what might have led Nagel astray. His account of motivated desires confuses two separate theses. The first is the HTM, whereas the second is what, after Alfred Mele, we might call the Antecedent Motivation Theory (AMT). The AMT holds that, in actual human beings, all motivation non-accidentally produced by practical reasoning issuing in a belief favouring a course of action derives at least partly from motivation already present in the agent before he acquires the belief. (Mele 2003, 89) The HTM and the AMT are clearly distinct. The latter speaks of the causal production of motivation, but does not specify what that motivation consists in; the former specifies what motivation consists in, but says nothing about the causal antecedents of a particular motivating state. Hence an advocate of the HTM can accept, in fact this is how I presented it above, that there is such a thing as motivation *ex nihilo*: motivation produced without the contribution of antecedent motivating states. (Millgram 1996, 207; Hubin 1996, 42-3) But, contrary to what my presentation may suggest, those who accept that beliefs have desire-producing causal powers need not also think that the AMT is false. They can still hold that though it was a belief that caused the desire, this act wouldn't have taken place without the assistance of antecedent motivation. Mele explicitly admits this possibility and it is also implicit in Smith's and Williams' account of deliberation. (Mele 2003, 94-5; Smith 1994, 166-174; Williams 1981a, 101-2, 104-5)

Nagel thought that his argument refutes the HTM, whereas at best it throws some doubt on the truth of the AMT.<sup>87</sup> Yet, and this is where my promised explanation comes into view, it is

---

<sup>87</sup> He is not alone with this mistake. Jay Wallace also argues that the debate between the 'Humean' and the 'rationalist' is about the truth of what he calls the 'desire-out, desire-in' principle according to which "processes of thought which give rise to a desire (as 'output') can always be traced back to a further desire (as 'input'), one which fixes the basic evaluative principles from which the rational explanation of motivation begins." See Wallace (1990), Attila Tanyi, THESIS, Page 158

probably not without reason that he committed this mistake. After all, in his book he was primarily interested in normative reasons and the AMT does have a bearing on normative reasons. The connection is simple. (cf. Mele 2003, 103) If one rejects the AMT, then even if he accepts the IR and stays within the confines of the HTM, he will still give up the Model. He will hold a position similar in its result – but, importantly: not in its presuppositions – to Smith’s idea: that one’s reasons are not relative to one’s actual desires because motivation or at least certain kind of motivation, can be produced without the help of the agent’s current ‘motivational set’.<sup>88</sup> It is no surprise that Williams took this to be the typical, Kant-inspired external reason position and answered it by invoking exactly the AMT. (Williams 1981a, 109; 1995c, 229 note 3) So the AMT is an undoubtedly important theory, and, in the light of what has been just said, perhaps I should have included it already in my presentation of the motivational argument. Yet, Nagel’s idea does not amount to a full-blown argument against the AMT, especially not if we consider the available forceful defenses of the theory. (e.g. Mele 2003, Chapter 4) In any case, my aim now is to investigate the HTM and, if my reasoning is correct, Nagel’s point has no bearing on *this* theory. So we’d better go on.

---

pp. 370. But Wallace’s principle is clearly nothing else but a reformulation of the AMT; hence its truth or falsity has no bearing on the fate of the HTM. Wallace, moreover, just like Nagel seems also to be interested in normative reasons and this might explain his mistake too. Or this is at least what certain passages of his article suggest. See (Ibid). pp. 368-370.

<sup>88</sup> The difference is this. Smith accepts the AMT but holds that because deliberation involves the systematic justification of desires (and beliefs), no matter what desires (and beliefs) an agent deliberates from, he will always arrive at the same normative conclusion in a given situation. The present account, on the other hand, claims to reach the same conclusion because it holds that the agent’s deliberation can proceed independently of his current motivations/desires. See FitzPatrick (2004), pp. 290-1 on the two approaches. Note also that there is a third position, again mentioned previously, with the same result: the view that there are standings desires constitutive of rationality. See Harrison (1976), pp. 13-15; Broad (1930), pp. 146, 179; Korsgaard (1986); Velleman (2000a). All three views can be taken to be Kantian and, for instance, Korsgaard seems to have held all three positions (but is now endorsing the first approach, though her grounds are different from Smith’s). See FitzPatrick (ibid) and Shafer-Landau (2003), pp. 172 on this.

### 4.2.3. Besires and desire entailing beliefs

We now know that besides meeting the EC a successful refutation of the HTM must also confront the teleological argument. There is a much-discussed way to do this. One can accept that desires differ from beliefs in that they have an upward direction of fit, yet claim that they are not independently intelligible from accompanying beliefs. That is, one can accept the EC and the MC but reject the IC. There are two ways to do this. Many argue that virtuous agents possess so-called 'besires', unitary states that function as believing-attitudes toward the proposition that *p* is right/virtuous, but serve as desiring-attitudes toward the proposition that *p*. (McDowell 1979; McNaughton 1988, Chapter 7; Wiggins 1991; Ross 2002; Zagzebski 2003) Others allow that the two states have separate existence but claim that possession of the moral belief conceptually entails the possession of the desire. (Phillips 1977, 146-7; McDowell 1978; Pettit 1987, 531)

This suggestion is not open to the objections that beset the previous proposals. It easily takes care of the teleological argument since that argument is premised on the separate intelligibility of beliefs and desires. And though it may look *ad hoc* and therefore open to the charge of hybridity to claim that certain beliefs are in fact besires or conceptually entail desires, advocates of the position offer some explanation for why this should be so. Most typically, they claim that some affective capacity serves as a persisting background condition to the possession of moral beliefs. (Phillips *ibid.*; Little 1997, 71) In fact, they sometimes put this in an evolutionary context holding that certain desires stand as preconditions to developing the rudiments of moral concepts. (e.g. Zagzebski *ibid.*)

This is then coupled with a robust metaphysical background. According to so-called sensibility theorists, certain ethical properties and sensibilities with both a cognitive and an affective aspect are interdependent. Neither can exist and be understood without the other, even though the property itself is not identical with or constituted by the sensibility (this is why



Darwall 1992, 159 classifies this position as non-constitutive existent internalist). On another view, there are so-called thick affective concepts that unite an affective, emotional aspect with a cognitive aspect that tracks certain ethical properties in the world. Again, neither side can be understood without the other, though the ethical facts concerned are not identical with or constituted by the agent's emotional reaction. (Zagzebski *ibid.* 111-113)

However, on the face of it at least, this suggestion is open to a knockdown objection. (Smith 1987, 57n; 1988b, 594; 1994, 123-5; Mele 2003, Chapter 5; Shafer-Landau 2003, 147-155) In my presentation of the IR I distinguished robust and moderate judgment internalisms. The ground for the distinction (not given there) and the reason why the latter is more plausible than the former is that there are motivational disorders - cases of listlessness caused by depression, fear, illness, physical tiredness and the like - that can break the connection between moral judgment and motivation. Therefore the most we can say is that these judgments are connected to motivation *ceteris paribus*: only when the above motivational disorders are not present. However, on the approach under discussion, this is clearly not the case. Since the agent's motivating reason is either a desire or a desire-entailing belief, moral judgment, understood here as a belief, and desire cannot be pulled apart. Robust moral belief internalism is true: moral belief connects to motivation *simpliciter*, thereby flying in the face of our common moral experience.

In response, advocates of the approach typically claim that the listless agent doesn't have a clear cognition of the situation. True, since he is an ordinary person we can reasonably suppose that he *once* had that view, but now his focus on the morally relevant elements of the situation is *clouded* by the disorders mentioned. (McDowell 1978, 28; 1979, 334; Platts 1979, 262; McNaughton 1988, 129) But this is puzzling. If he once had that view, then presumably he can also retain it now; hence we should attribute to him a proper moral belief. John McDowell admits this. "Failure to see what a circumstance means, in the loaded sense [in the sense of having a

besire/belief-entailed desire]”, he says, “is of course compatible with competence, by all ordinary tests, with the language used to describe the circumstance.” (McDowell 1978, 22) Nevertheless, he insists, “the relevant conceptions are not so much as possessed except by those whose wills are influenced appropriately.” (Ibid. 23) That is, we should simply *define* the virtuous person’s cognition of the situation as the possession of besires/belief-entailed desires. But, as it stands, this is a mere dogma, which, in light of our moral experience and McDowell’s previous admission, we have no reason to go along with.<sup>89</sup>

However, there is a way out of this difficulty. One can claim that there is more to moral cognition than belief: cognitive differences are not reducible to differences in propositional content. (Little 1997, 73-6; Zagzebski 2003, 119-121) An analogy from sensory perception illustrates the point well. We are familiar with the phenomenon that a picture can be organized under different gestalts, and that gestalt changes make for differences in our experience of the picture: once we see a rabbit there, another time we depict a bird. Moreover, such changes can bring with them differences in propositional knowledge: for instance, we can suddenly come to see the dots *as* Marilyn Monroe’s visage in the pointillist painting. But, and this is the crucial point, when the opposite process takes place, when we suddenly loose our ‘grip’ on the picture, we need not also loose the propositional knowledge we have previously acquired. That knowledge can survive shifts or losses in experimental gestalts: we can still retain our belief *that* Marilyn Monroe is there even though we can no longer see it *as* her visage. Hence seeing and believing are not the same: the former includes the latter *plus* something more.

---

<sup>89</sup> Garrard and McNaughton (1998), pp. 45-59 claims that in the absence of an account of the nature of motivation this argument is a case of *petitio principii* and that, when making this argument, we do not yet have such an account at hand. However, those who point to cases of listlessness are normally advocates of the HTM (either as motivational externalist or internalist), and the HTM can have such an account. Namely, it can say that to be motivated to act is to desire to act where desire is understood dispositionally. Although this will later put them in trouble, at this point of the argument this remark is perfectly in place.

The suggestion is that the same happens in the moral case. The virtuous person's cognition of the situation is like depicting Monroe's face in the picture: every element of the picture suddenly drops into line and the agent sees the action *as* virtuous and not just believes *that* it is virtuous. Talk of perceiving or conceiving is not accidental. Its aim is to highlight the fact that correctly perceiving that some act is right is more than believing that it is right: it involves the latter *plus* something more. And, specific of the moral case, this 'something more' is inherently motivating. There is thus nothing dogmatic about McDowell's position. The agent can retain his belief and thus rightly say that some act in the given circumstances is virtuous, but, at the same time, due to certain disorders, lack the relevant conception of the situation and thus fail to be moved to act.<sup>90</sup> Desire is independently intelligible from belief, but not from virtuous perception: robust perceptual internalism (to borrow a term from Darwall 1997, 308) is true.

Even so, there are reasons to go on. True, these reasons are *methodological* rather than philosophical in character.<sup>91</sup> There are three such considerations. First, I noted in the beginning

---

<sup>90</sup> More precisely, on the Aristotle-inspired sensibility theories at least, moral perception presupposes the right character, which is taken to be the result of proper moral upbringing, conversion or some other non-rational process. But we can set aside these otherwise important details of the theory – its importance is well illustrated by McDowell's answer to Williams in McDowell (1995), which basically claims, on this ground, that there is motivation *ex nihilo* – for it has no effect on the point made in the text.

<sup>91</sup> This is not to say that there are no attempts in the philosophical direction; in fact, there are many. But these attempts are either unsuccessful or are indecisive or their discussion would take us too far. Here is a brief survey. Lewis (1988) and Collins (1988) have argued that the existence of *besires*/desire-entailing beliefs is inconsistent with Bayesian decision theory. But, as Smith (1994), 221n points out, this is true only of instrumental *besires* and does not show that there can be no non-instrumental *besires*/desire-entailing beliefs. See also Price (1989) for an independent criticism of the Lewis-Collins position. There is also a temptation to claim that while the HTM is standard or austere, *besires* or desire-entailing beliefs are bizarre or at least not in line with 'psychological reality'. See Smith (1987), pp. 56-7; Mele (2003), pp. 120-3. However, the second part of the claim is mistaken. First, as Pettit (1987), pp. 532 points out, the aim of a theory of motivation is to provide a philosophical conception of folk psychological states, not a psychological theory built up from scratch. Both the HTM and the modified position does this in its own way, hence pointing to the austerity of the HTM is a non-starter. Second, there is also ground for doubting that the HTM would be the standard view, whereas the modified position would be bizarre. See Little (1997), pp. 66-69 who argues that Hume also held the view, and Millikan (1995) who claims that primitive as well as some human representational states are bidirectional. In his short treatment of the subject, Shafer-Landau (2003), pp. 137 note 11 also rejects the *besire* theory on the ground that it goes against the traditional account of individuation, which is done by the content of mental states. But this is only true of *besires*, not of desire-entailing beliefs; therefore the objection is not decisive. Finally, recently Jacobson-Horowitz (2006) has considered a possible objection to the theory on the ground that of the two directions of fit of *besires*/desire-entailing beliefs the *thetic*, downward one is dominant because the belief operator is larger in scope ('that it is morally required that p' v. 'that p'). This would make *besires*/desire-entailing

that in order to avoid the charge of hybridity those who reject the IC must give an explanation why there are besires or why the claimed conceptual entailment exists between belief and desire. And, as I have tried to show, while there is such an explanation, this explanation invokes substantial claims in epistemology and metaphysics. Consequently, the besire/desire-entailing belief position is both metaphysically and epistemically loaded, and it would be better to have a theory, which does not come with such a burden. The modified position, moreover, can be inferior to other refutations of the HTM in other respects as well. Since it restricts analysis to the virtues and uses a special form of cognition, it can be superseded by an account of motivation that deals with the domain of normativity as a whole and/or makes use of only beliefs. This is so because of a simple methodological principle: of two successful arguments against a given position the one that uses more of the central ingredients of the targeted position is reasonably considered better than the one that uses less.

Of course, here the question arises whether it is possible to interpret the theory in such a way that brings it closer to the HTM. I don't think so. As to the view of cognition employed, we saw that, on pain of falling prey to the objection about the plausibility of moderate judgment internalism, the besire/desire-entailing belief position must have a broader understanding of the cognitive than what is exhausted by propositional knowledge.<sup>92</sup> As to the sphere of application, it

---

beliefs more belief-like and therefore motivationally impotent. However, in response she has argued that in a sense relevant to their role in practical reasoning it is the telic, upward direction of fit that dominates. Her idea is that since in reasoning it is the content of beliefs that count, therefore besires/desire-entailing beliefs with normative content push their upward direction of fit in the fore, so to speak. Although I am skeptical about this argument (it again seems to confuse normative reasons with motivating ones), discussion of her views would take us too far (nor am I sure that I am right in my views).

<sup>92</sup> Another reason why advocates of the position must be more flexible about the cognitive is because this is a way, in fact perhaps the only way, they can account for weakness of will. Since they hold that motivational failure must always be traced back to cognitive failure, the only option they have is to say that when the agent thinks it is right to act, the reason why he does not act is because his perception is not perfect. Their original interpretation of this idea was that the factors that in the case of listlessness sap away the agent's motivation entirely, in this case blur his view only *to a degree* so that morality ceases to provide sufficient motivation for action without leaving him cold entirely. See McNaughton (1988), pp. 130-2 as well as McDowell (1978), pp. 91-4 on incontinence. But within the confines of the modified position such a move may no longer be available to them. For the analogy with gestalt changes

might seem that here more flexibility can be allowed. But again I am skeptical. For it is not without reason that advocates of the position are explicit about the constraints of their theory. (McDowell 1979; Zagzebski 2003, 114) They typically shy away from extending their theory to thin ethical concepts such as ‘right’, ‘ought’, or ‘good’ and stick instead to thick virtue and vice concepts. Linda Zagzebski is explicit about the reasons that inspire this move. (Ibid. 119-123)

She defines the process she calls the ‘thinning of moral judgment’ as comprising of three levels. On the first people use thick concepts in the perceptual way and are always moved to act as prescribed. On the second level, they recall propositional knowledge of these concepts but are not necessarily moved to act, though they often have a flicker of motivation remaining. Finally, on level three there are thin moral concepts and the situation is more or less the same as on level two. This process permits Zagzebski to account for weakness of will (level 2 judgments with a flicker of motivation remaining from level 1), listlessness (level 2 judgments with no motivation remaining from level 1) and amorality (level 3 judgments without knowledge of level 1 concepts). And these solutions may not only be attractive, but also the only available ones given the well-known problems classical approaches to amorality encounter (more about this in section three), and the effect the modified position may have on their usual account of weakness of will (see the above footnote on this).

#### **4.2.4. The Cognitivist Theory of Motivation**

We now have the recipe for the theory we are looking for: it must meet the EC and tackle the teleological argument while respecting the distinct existence of beliefs and desires. Drawing on

---

suggests (though, I am not sure whether it also proves it decisively) that perception does not come in degrees: one either sees Monroe’s face in the picture or one does not – there is no middle way. If this is so, advocates of the position must endorse the process of the thinning of moral judgment mentioned in the text and its account of weakness of will. See Zagzebski (2003), pp. 122. I thank Eric Brown, Greg Bognár and András Miklós for suggesting this problem to me.

the view Jonathan Dancy calls pure cognitivism I will present such a theory in three steps. The first is to distinguish the state of being motivated from that which motivates, and claim that the agent's motivating reason is identical with the latter. (Dancy 1993, 19; 2000b, 85) The argument for this position is itself composed of three sub-steps. The first points out that the distinction is not merely a terminological point, as some seem to have suggested (Garrard and McNaughton, 1998, 55). The idea is that whereas what motivates has a role in the explanation of action, the agent's being motivated, that is to say, the motivatedness has no such role. As Dancy puts it: "[...] though necessary for the eventual action it [here he refers to desire, but as the second step will make it clear, he identifies desire with motivatedness] is not a cause of that action; it does not pull its own weight in the causal story." (Dancy 1993, 20)

Now the question is why Dancy thinks that the motivatedness, unlike that which motivates has no role in the explanation of action; in other words, what we need is a philosophical rationale: this is the second sub-step. Dancy's logic seems to be this. The reason why the motivatedness has no role in the explanation of action is because it is itself (or, at least, part of) what is to be explained: it is the explanandum, not the explanans. And, further, this is so because the motivatedness is on the same causal and/or explanatory level as the action itself: explaining the latter is the same as explaining the former. "The explanation of motivation", Dancy says, "must be structurally similar to the explanation of action" because "often the only thing necessary to take us from motivation to action is the absence of contrary motivation, or the fact that contrary motivations were 'weaker' than this motivation." (Dancy 2000b, 85)

Now, this may sound contradictory – first we say that action and explanation is on the same level, then we speak of moving from motivation to action – but there is a way of making sense of it. Dancy himself suggests the solution when he points out that his position is analogous with the volitional theory of action. (Dancy 1993, 36 note 4 referring to Davis 1979, 38-41) On

that theory volition is seen as part of a bodily action: the action begins with the volition, which then normally but not necessarily causes the bodily movement. Therefore the volition is not what explains but is what is to be explained (roughly: action equals volition that causes bodily movement); it is the causing not the cause. Dancy's thought, as mentioned, is analogous with this theory, that is, he must think that motivatedness is also part of action, which can but need not cause the bodily action itself (roughly: action equals motivatedness, which, after having worked upon by volition, causes bodily movement). (cf. Dancy 2000b, 86) I find this idea attractive but, admittedly, it is a controversial one.<sup>93</sup> Its discussion, however, would take us deep into the theory of action; therefore I suggest that, in order to be able to continue with our presentation, we provisionally accept the thesis. Once this is granted, however, the third-step automatically follows. (Ibid.) For it simply points out that if explaining action entails explaining motivatedness by that which motivates, then this last factor cannot be motivatedness itself since that would be explaining of itself. Consequently, there must be something else that explains and this cannot be else but *what* motivates: the agent's motivating reason.

The first is perhaps the most important of the three steps since we need it in order to get anything started. Moreover, it must be emphasized that motivatedness as concerned above is understood to have a distinct and real psychological existence.<sup>94</sup> The first step has no problem

---

<sup>93</sup> John Stuart Mill appears to hold a similar view when in his notes on his father's *Analysis of the Phenomena of the Human Mind* he says the following: "I believe the fact to be that Desire is not Expectation, but is more than the idea of pleasure desired, being, in truth, *the initiatory stage of the Will*. In what we call Desire there is, I think, always included a positive stimulation to action, either to a definite course of action which would lead to our obtaining the pleasure, or to a general restlessness and vague seeking after it. The stimulation may fall short of actually producing the action: even when it prompts to a definite act, it may be repressed by a stronger motive, or by knowledge that the pleasure is not within our present reach, nor can be brought nearer to us by any present action of our own. Still, there is, I think, always, the sense of a tendency to action, in the direction of pursuit of the pleasure, though the tendency may be overpowered by an external or internal restraint." See Hubin (1999), pp. 44 note 19 who quotes Mill (1963), pp. 215, Italics are mine.

<sup>94</sup> We need to make this clear for otherwise there will be a constant temptation to confuse the account with Nagel's first theory that we have started this section with. Dancy is mostly clear on the difference, but there is one point – in Dancy (1993), pp. 20 in brackets, just after the sentence quoted in the text concerning the first sub-step - where he says that we need desire in the explanation of action only for logical purposes. This led Shafer-Landau (2003), pp.

admitting this; in fact, as we shall see, such an admission will count as the strength of the emerging theory. Turn now to the second step. Its role is to identify desires with motivatedness leaving the role of what motivates open. This appears to be true by default since in the beginning we characterized the HTM as claiming that desires are essentially motivating states: states that always motivate, whenever they are present. Still, one might claim that there are counterexamples to this claim, desires that *contingently* motivate: they also motivate in their own right but can be present without motivating. The existence of such desires would not compromise the HTM because, first, most of our desires would still be essentially motivating, and, second, the required contrast with belief as a state that neither contingently nor essentially motivates would remain. Hence if we can find such desires, they would offer sufficient counterexamples to equating desiring with being motivated without giving up the HTM.

Alfred Mele has recently suggested that certain desires - he calls them state-desires - offer good examples. Here are his three cases. Ann, a Pistons fan, hopes that they won last night's basketball game. But, Mele says, she understandably has no motivation to try to bring it about that they won. (Mele 2003, 22) In another example Mele is standing at the airport and realizes that he should have called from home to see whether Angela's plane departed on time. So now he desires that her plane departed on time, but, again, this in no way motivates him to bring it about that the plane left on time. (Ibid. 26) Finally, Connie is a big Giants fan who hopes that they are

---

138 to interpret Dancy (1993) as offering a version of Nagel's theory. However, in the very next sentence Dancy makes it clear that this is not what he has in mind. Also, Locke (1974), sections 2-3 and (1982), pp. 243 uses almost the same ideas as Dancy himself (and, we should note, does that earlier than Dancy), but does so in a framework governed by a distinction similar to Schueler's. His main claim is that if we use desire in the philosophical sense, which basically corresponds to Schueler's pro attitude sense, then we should end up with Dancy's conclusion: that desire is not what explains but what is to be explained: it is the causing, not the cause. This is a strange marrying of ideas. On the one hand, if Locke has in mind a Dancy kind of view, then it is hard to see why he needs the distinction at all. For, as pointed out in the text, given the structure of Dancy's argument, it just doesn't matter if we think of motivatedness as having a real psychological existence. If, on the other hand, he doesn't share Dancy's idea, then his argument is just the same as Nagel's and it encounters the same problems. The only thing that could make it different would be the claim that the agent's motivating reason never involves a desire. Besides, however, that Locke does not want to claim this (on his view what Schueler calls desire proper *are* causes of action), it is also doubtful how he can get this conclusion. See Schueler (1995), pp. 37-8 on this.



going to beat their traditional rivals next month. But she is also a space pilot on a long journey and thus knows that there is no way to find out about the results of the match. She is, further, a special person who, when she knows that she has no way to find out how matters of minor to moderate importance to her have turned out, tends not to think or daydream about those matters. But then, Mele notes, it is difficult to see what else her motivation could manifest itself in. (Ibid.)

However, these examples pose no threat to a thesis that equates desiring with motivatedness.<sup>95</sup> The first two desires Mele cites have a feature in common: they are desires for the uninfluencable. It is impossible to satisfy them and this is why the agent is supposed to have no motivation to do so. But others, who have also recognized the existence of these desires and were, in some cases at least, even tempted to use them against the second step, also realized that there is a standard response to any such desire. (Schroeder 2004, 16-7; Dancy 2000b, 87-8) We can say that though the agent has no motivation to do the action desired, he might still have motivation to do the action that he believes conduces to the desired result. That is, in the above cases we can say that Ann can be motivated to check the morning papers for the result, whereas Mele can be moved to check the monitors at the airport. It seems that equating desire with motivatedness needs no more than this; the first two examples are harmless to the second step.

Mele, however, rejects this idea. He says that there is a difference between a desire's *providing* motivation and *encompassing* it and that both Anne's and Mele's desire is of the former kind. But even if we accept this distinction, it is hard to see its relevance for the case we

---

<sup>95</sup> In the case of desires for the uninfluencable it may cause further trouble that on certain readings at least such desires don't qualify as desires at all. Think of Velleman's account in Chapter I according to which desires differ from other conative attitudes, such as wishing or hoping, in that they are only directed at what is attainable, i.e. what it is possible to get or pursue. This is their 'constitutive aim', to use Velleman's term. See Velleman (2000a), pp. 116-7. It is interesting that Mele himself endorses this position (or something very similar to it) at Mele (2003), pp. 135-6. He says that desiring *A* precludes being convinced that one cannot *A* (action-desire) and that desiring *x* precludes being convinced that *x* is impossible (state-desires). Observe also that in the concerned examples he is speaking of hopes and hopes on the refined version of the functionalist theory of desire, do not qualify as desires. Mele certainly thinks that this causes no problem for him, but it is not clear to me how he wants to reconcile this account with his examples.

are discussing. For the difference between encompassing motivation and providing it is that in the latter case the desire constitutes motivation to act *indirectly*, through its contribution to the motivational strength of another desire. (Ibid. 15, 20) To take Ann's case: her desire that the Pistons won last night constitutes motivation to check the morning papers by increasing the motivational strength of her desire to check the morning papers. Hence indirectly though, but in desiring that the Pistons won last night she is also being moved to check the morning papers. And the same is true of Mele's desire that Angela's plane departed on time and his act of checking the monitors. And, again, a proper defense of the second step needs no more than this.

Connie's desire, however, requires different treatment. In her case the above strategy doesn't work. By stipulation, her motivation can only channel itself into daydreaming, and this, again by stipulation, is also not open to her. Again, others have also noticed this type of desire. In Schroeder's example the agent hopes that the committee will make up its mind in his favour, but is in no way willing to do anything to influence the committee's decision and thus bring about the desired result. (Schroeder 2004, 17) Nevertheless, there is a way to accommodate these desires. Observe that in the described cases the agent's desire, due to an internal or external restraint, is directed at something that cannot in fact be achieved. In this respect the desires are similar to the previous cases. The difference is that in their case the agent's motivation can find no alternative outlet; this is why both Schroeder and Mele think that there is no motivation present. But it seems reasonable to deny this claim. For often when we have such desires we also feel frustration *exactly because* they are about things we cannot bring about or cannot avoid. And this sense of frustration shows that there is motivation here, namely motivation that was frustrated.<sup>96</sup> (Dancy 2000b, 88)

---

<sup>96</sup> There is only one point in Mele's discussion of Connie's case where he appears to consider a suggestion that is vaguely similar to mine. He imagines someone claiming that "Connie's desire for a Giant's victory does provide  
Attila Tanyi, THESIS, Page 170

This, again, is another controversial thesis. We have not proven it right yet since that would require the denial of two auxiliary theses: that whenever there is a feeling of frustration, there is also at least a flicker of motivation; and that it is impossible to have a desire such as the ones above in the case of which the agent feels no frustration. Obviously, I have not provided any argument to show that these claims hold and, in fact, I doubt that any philosopher can. What at best we can do, I believe, is to consider potential counterexamples such as the ones above and try to explain them away. I think my treatment of the above cases is plausible, and therefore we would need further counterexamples to keep this debate going. Schroeder has one such case, so we should consider this before ending our discussion of the second step. He brings in Galen Strawson's weather watchers, beings who have no disposition to cause their owners to act, but who nevertheless have desires that the weather be one way or another. Moreover, as Schroeder points out, such creatures can also have feelings of despair if their hopes about the weather are not realized: that their desires are not motivationally potent does not rule this out. (Ibid. 20) I agree but for the same reason I also think that we cannot rule out the possibility that weather watchers feel frustration not only because their hopes are not realized, but because they cannot do anything to realize them. Again, this is a stipulation that is not, strictly speaking, proven right; but, again, it is a plausible stipulation and only a psychological, not philosophical theory can prove it wrong.

---

motivation to fantasize about their winning and the psychological facts about her that I mentioned simply *prevents* that motivation from being manifested behaviorally." But, he continues, "this claim has the ring of a desperate move to save a theory." See Mele (2003), pp. 26-7. I agree, but this is not the point I am making. What I say is that Connie's underlying motivation *does* manifest itself: it shows itself in the frustration she feels over not being able to do anything about her desire. Can Mele say that she is a special person so that she not only does not daydream in situations like hers but does not even feel frustration either? He certainly can, but this would too be a 'desperate move to save a theory'. It would be so tailoring the situation that in the end we see Connie not as a real human being but as some creature without feelings, emotions, in fact, any kind of real psychological make-up. Given the fondness Mele elsewhere shows for 'psychological realism', I doubt that he would be happy to embrace this result.

This takes us to the last step. It draws the conclusion: since desiring is being motivated to act, what motivates cannot be the desire but something else, namely a belief. More precisely, there are two such beliefs. The first is about how things are, thus ensuring that the agent has some idea *what* makes his action necessary; the second is about these things constituting reasons for the agent, thus ensuring that acting *is* necessary. This account is different from Dancy's own suggestion that the second belief represents the world as it will be when and if the action is successfully completed. (Dancy 1993, 14) But I think my proposed candidate fares better than his. To mention one advantage, it is able to handle apparent exceptions to Dancy's account. (Ibid. 34-6) One is the case where there is something the agent wants to be doing rather than to have done; the other is the case where, in response to someone's distress, the agent expresses his sympathy. In both cases we can attribute to the agent a belief concerning the reasons his performing some act or someone else's distress provides.

Nevertheless, there are cases that fall outside the scope of my account. As noted in previous chapters, very young children, non-human animals and some mentally handicapped adults may not possess the concept of reason, thus they cannot have normative beliefs either. But we can deny that there exists continuity between toddlers/animals/mentally handicapped and mentally able grown ups.<sup>97</sup> (cf. Shafer-Landau 2003, 132) We can, moreover, do this on evolutionary grounds. We can say that what marks the latter group from the former is the presence of fully functioning reflective capacities. What makes us humans special is that we are capable of standing away from our ends, or at least from those ends that we put under scrutiny, and thus make our choice. And, as noted in Chapter I, it is this space of reflective consciousness

---

<sup>97</sup> There may be another way to explain away this problem, perhaps while also putting it in the evolutionary framework I am going to present in a moment, if we accept the suggestion made in the previous chapter: that there is a difference between the attitude of liking and the attitude of full desire. In fact, recall that Scanlon actually puts the distinction while invoking the case of children and animals in Scanlon (2002), pp. 340.

that the notion of a reason, in fact, normative notions in general emerge. Hence it is plausible to claim that we should not expect the same sort of motivational structure in the case of mentally healthy adults as the one we find in the case of animals, toddlers or the mentally handicapped. But this causes no real problem. For if the previous two steps were right, we can simply constrain the agent's motivating reason to the first, ordinary element of the pair of beliefs introduced above, thereby accounting for the mentioned discontinuity.<sup>98</sup>

There is one last issue to deal with. We have the somewhat unusual case of people who have the second belief but not the first. For instance, one may vaguely recall that he should do something but can't remember why. But normally we indeed do that thing (because, say, we trust ourselves), so we do have some motivation. And even when the lack of the first belief stops us from doing the thing, this need not mean that we didn't have *some* motivation to do it. It may just change the pattern of motivation in various ways. (Garrard and McNaughton 1998, 54) For example, if I forgot why I should do something, I may also be unsure how much reason I have to do that thing. Hence I do not do it, but not because I am not moved at all but because I decide that it makes more sense to do other things until I remember what the reason was. In fact, if the action is part of a lengthier series of actions, I may decide to do something opposite to it, in order to find out what the reason was and thus be able to continue the train of action. Here, again, it is not the case that I am not moved to do what I originally intended to do; it is only that this motivation manifests itself in an unusual, indirect way through another action.

---

<sup>98</sup> In fact, this may also explain why toddlers and the mentally handicapped often have such a sporadic attention. In absence of normative concepts, there is nothing that can anchor their attention by keeping it focused on one particular thing. And when this is not the case, when their attention is focused on something, some other explanation is available that doesn't require us to postulate the use of normative concepts. In the case of very young children, we might say that such attractions are steps on the way to acquire the concept of a reason; they are part of the child's normative development, so to speak. Whereas in the case of mentally handicapped adults focused attention is normally the product of the problem that put the person in his present situation in the first place.

### **III. Defending the CTM**

#### **4.3.1. The CTM and previous objections**

The three steps just outlined and defended add up to what we may call the Cognitivist Theory of Motivation (CTM). In the remainder of this chapter I will first show the advantages of the CTM over other accounts. Then I will defend it against objections that haven't been considered so far.

The CTM is able to handle all the objections other theories have stumbled upon. It can admit that whenever action occurs there is a desire present that has full psychological reality: the EC is met. Further, by allowing no role for desire in the constitution of the agent's motivating reason, it also avoids the charge of hybridity. Since desire never pulls its weight in the explanation of action (besides being what is to be explained), it follows that beliefs never need the help of desire to explain action. Therefore the pressure that led to the charge of hybridity - that if one allows that the HTM is true of certain areas of human motivation, but claims that it does not hold for others, one must explain why these beliefs are special - evaporates. Furthermore, the CTM can be applied to normativity as a whole, while making use of only belief and desire. Nor does it come, or at least it need not come with heavily loaded epistemological and metaphysical presuppositions. Its emphasis on the pair of beliefs that are responsible for human motivation is compatible with a wide range of meta-ethical theories of morality or of practical reason.

Also, the CTM has no trouble with tackling the teleological argument. The agent's motivating reason is constituted by a pair of beliefs: they rationally explain the agent's desire. The teleological argument only shows that desire must be present in intentional action as the sign that the agent is motivated, i.e. that the EC is true. We can simply reverse Smith's 'new' teleological argument in favour of a cognitivist reading. (Dancy 1993, 29-30; 2000b, 91) It goes like this. Smith's premises establish the following conclusion: to *have* a motivating reason is to

have a desire. But what he needs is more than that; it is this: motivating reasons *are* (or, at least, include) desires. The difference between the two conclusions is crucial for Smith. For the former but not the latter is consistent with the claim that desires are states of being motivated, which derive their motivational force from beliefs. To see this we just have to rephrase the first conclusion to claim: to *have* a desire is to be motivated by a belief about how things are and a belief that these things give reasons to act. Since Smith provides no further backing for why we should accept the second conclusion instead of the first, his argument fails.

Finally, the problem of listlessness that pulls us towards moderate judgement internalism can receive two responses. First, it can be argued that the listless agent lacks the relevant normative belief. (Garrard and McNaughton 1998, 52) There are two ways to do this. One is to deny that the concepts the listless agent uses are translatable to the concept of reason. This is an option with moral beliefs since here we need rationalism to establish the connection. But it need not be the case with evaluative concepts: here the connection is often presupposed. Therefore, we should rather take the other route and claim that as a matter of psychological fact listless people, though they make moral or evaluative judgments, don't have normative beliefs. This is not implausible. Seriously depressed people often cannot see that some action, in fact, that anything matters; and it is a reasonable interpretation to say that this is because they see no reason to act.<sup>99</sup>

By construction, however, this is a rather contingent response. It is hard to come up with a general psychological law stating that listless people always lack the requisite normative belief.

---

<sup>99</sup> Dancy says something similar: "People who suffer from accidie are those who just don't care for a while about things which would normally seem to them to be perfectly good reasons for action; this is so whether the reasons are moral reasons or more ordinary ones. Depression can be a cause of accidie. The depressive is not deprived of the *relevant* beliefs by his depression; they just leave him indifferent. He knows that if he doesn't act now he will lose the opportunity he has been working for for two years, but he can't see that this matters." See Dancy (1993), pp. 5, *Italics* is mine. It is not clear, however, if by 'relevant' Dancy means moral, prudential and other beliefs or he means normative beliefs. On the first reading my interpretation in the text stands; on the second it does not. Garrard and McNaughton (1998), pp. 52 gives the first interpretation, while Dancy's later writings seem to suggest that he has the second version in mind.

Perhaps for this reason Dancy favours another kind of response. (Dancy 1993, 24-5) The idea is to show that the CTM is compatible with a moderate version of normative belief internalism. It goes like this. Since belief and desire are taken to have separate existence, we can say that though beliefs are what motivate they do not always do so. Beliefs are *intrinsically* motivating: they motivate in their own right but do so only contingently. And we can explain why this is so by pointing to the motivational defects Stocker and others cite: it is these factors that block the otherwise enduring connection between beliefs and desire/motivation. Again, listlessness comes out as harmless from the point of view of the CTM.

Although this response is more elegant than the first one, it should be noted that it assumes the truth of two theses.<sup>100</sup> First, the motivational defects appealed to in the explanation are only *causes* of the agent's behavior and are not themselves motivating reasons. Second, in the corollary case when the agent is motivated to act, the *absence* of motivational defects is not among the agent's motivating reason(s) to act. (Dancy *ibid.* 25-6) The first thesis I take to be uncontroversial since not even advocates of the HTM regard carelessness, inattention, despair, drunkenness or neurophysiological disorders as motivating reasons (e.g. Smith 1994, Chapter 5).

The second thesis is more controversial and it cannot be defended here properly (though, we should note, its rejection would not lead us straightaway to the HTM). The issue here is how we picture the explanation of action: whether we think that all factors that are necessary for the success of the explanation are part of the explanation, or we think that these factors play different roles and need not be included in the explanation itself. The latter is clearly the position we need (this is parallel to the normative case in Chapter I, note 7). Certain factors - such as the absence of certain disorders serve as background conditions to the explanation of action - enable the

---

<sup>100</sup> We could have considered these claims earlier when dealing with the desire/desire-entailing belief position since there too the truth of both theses is implicitly assumed. Hence what I say here also has an effect on the fate of that position.



explanation so to speak (i.e. the explanation wouldn't take place without them), without themselves being part of that explanation.<sup>101</sup> (Dancy 2004, 46-7)

### **4.3.2. The CTM and new objections**

Although it has advantages over other accounts of motivation, the CTM remains controversial. In the remainder of this chapter, I deal with objections that have not been considered so far. I am going to discuss four charges, starting with the ones that are easier to answer and proceeding to those that pose a more serious challenge to the CTM. To facilitate discussion my presentation takes a question and answer form.

*First question.* It is reasonable to suppose that the following requirement is true: a belief can only have an effect on the agent's behavior if it manifests itself in occurrent thought. And if we insist on keeping desire in the picture we get something like Scanlon's idea of attention-directed desires: desiring something involves having the thought that something is desirable and this thought keeps occurring to the agent from time to time. (Scanlon 1998, 39) However, for a thought to manifest, or as it is sometimes called to 'surface' on the agent's consciousness in this way, it needs the help of a mental state. But the thought itself cannot play this role because it disappears from time to time and once it disappeared it cannot cause later instances of itself. Hence we need the presence of another mental state and, on the functionalist approach at least, this can only be a desire. (Persson 1999, 163-4) Can this be made consistent with the CTM?

---

<sup>101</sup> As noted in the text, I cannot here settle this debate. But, at least in the present case, certain intuitively appealing considerations present themselves. First, as Dancy notes, the analogous case in the theory of causation supports the truth of the second position. To take his example, even if we allow that the causes of my attending a conference in the US would not have been sufficient had the US recently declared war on England, we need not hold that the fact that this had not happened was among the causes of my presence there. See Dancy (1993), pp. 24. Second, if we endorse the first position, it seems that we have to include each and every factor the absence of which may be necessary for the success of the explanation. And if we include beliefs in the picture (which we should), this might result in a never ending list of factors. See Dancy (2004), pp. 47.

*Answer.* I am not sure that it need be; later I will say more about my worries concerning Scanlon's idea. Till then, we can say the following. Desire is still a state of being motivated with an independent existence and distinct psychological make-up. Hence the CTM has no trouble with admitting that it is something over and above the beliefs it helps to surface and stay afloat. At the same time, however, the desire will still be rationally explained by these beliefs, which will still serve as the real motivators. There will be some sort of *interaction* between belief and desire: the former explaining the latter and the latter helping the former stay on the surface of the agent's thought. (Dancy 1999, 216) There is no contradiction here. We should remember that the notion of explanation the CTM employs is not causal but rational: it wants to say that the motivating reasons that explain the agent's desire are constituted by a pair of beliefs. And this role is compatible with the claim that the CTM is not concerned with the formation of desire, or with its persistence. For all what the CTM says it is possible that a desire comes to life without normative beliefs playing any kind of role, or that a desire persists after the beliefs have disappeared.<sup>102</sup> The CTM only claims that we need a pair of beliefs in order to make sense of the agent's motivation and his action. Without the beliefs the agent's action would be unintelligible, i.e. lacking rational explanation even from the agent's point of view. And this role of beliefs in no way contradicts the idea of interaction spelled out above.

*Second question.* On the CTM it is a pair of beliefs that do the job of motivating and one of those beliefs is a belief about reasons. This naturally brings up the issue of weakness of will; it is hard to see how the CTM can make sense of this phenomenon. What happens with the weak-willed

---

<sup>102</sup> This is my answer to James Lenman's objection that the rational formation of desires may not be informed by *anything*, neither belief nor desire. What we in fact do in such cases, says Lenman, is passively discovering what we like and what we are averse to; we are simply so constituted as to respond to certain experiences by forming certain desires and aversions. In this case we can still speak of causation but not about motivation. See Lenman (1996), pp. 295. As my statement in the text shows, I agree with Lenman but I don't think that this phenomenon is incompatible with the CTM.

agent is that he doesn't act even though he judges - let us again suppose: believes - he has sufficient reason to act. And this need not happen because the agent lacks motivation to act; it is enough if his motivation is not enough to produce action. But, it seems, the CTM is incapable of granting anything like this. If one sincerely thinks that he has sufficient reason to act then, since his desire is taken to be rationally explained by a pair of beliefs one of which is a normative belief, he should also have sufficient motivation. Hence the CTM is either false because it cannot account for weakness of will or it must deny that weakness of will exists. This looks plainly obvious, doesn't it?

*Answer.* No, it does not. A general extension of this problem is the phenomenon that I have admitted to be common in the previous chapter: that the motivational strength of the desire need not always track the strength of the perceived reasons for it. The claim now is that the CTM has no resources to make sense of this phenomenon. But this is not the case. We can still say two things. (Dancy 1993, 24, 26-7; 2004, 145-6) First, since belief and desire are separate existences, it might be that the connection between the agent's recognition of the reason and his desire/motivation is severed. It is not severed entirely as in the case of listlessness, but only to a degree: his motivation is reduced but not eliminated completely. Second, we can say that the agent's recognition of the reason is clouded, hence the strength of his desire/motivation that responds to it is reduced. Both considerations can explain weakness of will as well as the more general phenomenon mentioned above. The basic point is that it is not the stringency of the perceived reasons what matters but the clarity of the perception and/or the quality of the connection between the perception and the responding motivation. Moreover, in both cases we can cite familiar factors as being responsible for the observed 'anomaly': physical tiredness, illness, depression, inattention, carelessness and so on.

*Third question.* We have noted that the idea that normative judgments express beliefs and the CTM together entail a moderate form of normative belief internalism. This being so, however, it seems that we have to face counterexamples other than listlessness: the amoralist and the wicked person. The former is someone who, systematically and without suffering from motivational disorders, is indifferent to his moral beliefs. The latter is someone who, systematically and without suffering from motivational disorders, is moved by his moral beliefs but in a perverted way: he is moved to do what is wrong because it is wrong. (Brink 1989, 45-50) Can the CTM allow for the existence of these people?

*Answer.* Yes, it can; in fact, it can do better than other attempts. Normally, the internalist reaction to the amoralist has followed in the footsteps of Richard Hare by claiming that the amoralist does not make moral judgments because he only uses moral concepts in inverted commas. (Hare 1952, 124-5, 164-5; McNaughton 1988, 139-40; Miller 1992, 84; Dancy 1993, 6; Smith 1994, 68) And wickedness is typically accommodated on the ground that such people make mistaken but otherwise proper moral judgments, that is, they are moved by what they mistakenly take to be right. (Anscombe 1957, 75; McNaughton 1988, 140-44; Dancy 1993, 7) However, there are serious doubts about the adequacy and plausibility of both these strategies. (Svavarsdóttir 1999, 176-181, 191-194; Velleman 2000a, 118-9) Therefore it would be better to have a theory that can do away without appeal to inverted commas or mistaken moral judgments.

The CTM is such a theory. First of all, the existence of amoral and wicked people, as it stands, is irrelevant for the success of the CTM. As to the amoralist, recall the point made in section one, the falsity of moral belief internalism implies the falsity of normative internalism only if *rationalism* is shown to be true. And to repeat, since the CTM appeals to beliefs about reasons, its natural extension is normative internalism; hence it is immune to the counterexample

of the amoralist. This is a point acknowledged on both sides. Many view the amoralist as someone who is not moved by his moral beliefs because he sees no reason to act. (Brink 1989, Chapter 3; 1997, 18-21; Garrard and McNaughton 1998, 48; Skorupski 2001; Wallace 2001) As to the wicked person, the CTM can see the wicked in the ‘traditional’ way: he is moved by what he mistakenly takes to be right and thus reason-providing. Or, it can see him in the ‘radical’ way: he is moved by what he correctly takes to be wrong and thus reason-providing. (Velleman 2000a, 121-2) Obviously, the two explanations cannot go together, but it is no task of the CTM to decide which one is right.

Second, even if we disregard the above responses, we can still say that the existence of amoral and wicked people *presupposes* the truth of normative internalism, hence, on the present interpretation, the truth of the CTM. Two considerations support this claim. Take our practices of persuasion. What we typically do in normative discussion is to call our interlocutor’s attention to certain facts about his situation in virtue of which he has reason to act. In other words, what we try to do is to make him see and endorse the truth of some normative proposition. But we wouldn’t do this if we thought that externalism is true and he is only moved to act if he has some independent desire to act on his normative beliefs. (Johnson 2001; Zagzebski 2003, 104) Further, we face the same situation even in the individual psychology of the amoralist or the wicked. We normally find that they are people in whom once the internalist connection functioned properly and the joint only came apart later. (Blackburn 1998, 61-4) In sum, though there can be counterexamples to internalism, they can only exist against a *social* and *individual* background that supports the theory.<sup>103</sup> (Dreier 1990, 11; Blackburn 1998, 61; Lenman 1999, 448)

---

<sup>103</sup> Gibbard remarks that perhaps even this much concession to externalism is not tenable. He thinks that we should retain the core of the inverted commas strategy and claim that for those who use normative concepts without being systematically moved, these concepts are empty. These people, Gibbard claims, cannot have meaningful, fine-grained normative convictions on matters of normative controversy. As a result, they cannot have normative

One might find this response inadequate. What we want to say is that normative belief is *supposed to* motivate: it motivates when the agent does not suffer from motivational disorders *and* when he is not amoral or wicked. But just what do we mean by ‘supposed to motivate’? Here is a suggestion. (Bedke ms) Moral behavior has a proper function, i.e. a function that has helped people to selectively fit, thus proliferate in the course of evolution. It enables and furthers cooperative, mutually beneficial outcomes for moral individuals so long as they stick to their own kind and avoid too much free riding. Moral judgment has a crucial role in this process. It ensures the effective operation of moral interaction by translating moral recognition into behavior; and the only way it can do this is by being intrinsically motivating. What we have here, then, are two nested proper functions: moral interaction brings benefits to people and moral judgment makes moral interaction possible. This is what we mean by the claim that moral judgment, hence normative belief, is supposed to motivate: it is its naturally evolved *purpose* to do so.<sup>104</sup>

*Fourth question.* The HTM offers us a constitutive account of action-explanation, that is, it claims that an event for which a belief-desire explanation cannot be given is not an action at all. To do the same, the CTM must claim that whenever we have an action we have a desire that is rationally explained by the beliefs it appeals to. However, there are counterexamples to this claim, desires that cannot be explained while their satisfaction qualifies as an action. Hence the

---

discussion with others because there is nothing, i.e. nothing significant they can disagree about. See Gibbard (2003), pp. 157-8. I find Gibbard’s point attractive but I think this is because I share his underlying rationale (or, at least, I am neutral on the issue): that normativity is, or at least partly constituted by motivating force (as he earlier put it, normativity is ‘endorsement’). Yet, for the same reason, externalists would reject it. They would say that one can have meaningful normative discussion with others on the basis of one’s normative convictions even if he is using normative concepts in the externalist way. See e.g. Cuneo (2002), pp. 484 for such a response.

<sup>104</sup> Bedke (ms) calls this idea Moral Judgment Purposivism (MJP). It should be noted that he takes the MJP to articulate an *a posteriori/synthetic* necessity as opposed to being a conceptual constraint. However, he also remarks that the MJP can also be seen as a theoretical extension of the kind of conceptual analysis moderate judgment internalism offers.

CTM must commit itself to the ‘deeper truth’ of Humeanism and admit that the only motivating states capable of playing a constitutive role are belief-desire pairs. (Smith 2003, 460-4)

*Answer.* The question is what those counterexamples are. Smith himself refers to the phenomena Stocker (1979) deals with. This gives us two options: one is listlessness, the other is perverse desires. What we associate with listlessness is a state of passivity when the agent is not moved to do anything: this is what depression, neuro-psychological disorders or exhaustion typically results in. Consequently, this is not a case we have to deal with here: there are no desires present on these occasions exactly because of the disorders mentioned. As to perverse desires, these are cases when the agent, due to certain destructive moods and emotions, desires what he himself takes to be bad. (Stocker 1979, 746-9) This means that structurally these cases are no different from the case of the wicked person; in fact, they are instances of *occasional* badness. They can be interpreted in either of the ways mentioned there, though I think, exactly because of the moods and emotions Stocker cites, the second ‘radical’ interpretation is more in order. They thus pose no counterexample to the CTM and its aim to serve as a constitutive account of action-explanation.

There are several other counter-examples, however. In the previous chapter we have dealt with cases when one just ‘feels like’ doing something. Here, I admitted, it would sometimes be superfluous to suppose that people think they have reason to have what they desire. And there are examples other than these that invite the same analysis. In Smith’s case a man walks out of his way to get his daily paper at a newsstand where he can look at himself in the mirror. Here we can plausibly explain the man’s action with the desire to look at himself in the window, even though he would deny that he has such a desire or that he has reason to look at himself in the window. (Smith 1994, 106) Finally, we have the case of spontaneous actions. When I meet a friend on the street and out of habit greet him, I certainly don’t do it because I think I should. Also, when

someone throws a ball at me and being well-trained in handball I catch it, I certainly do not think that I should catch it because I have reason to avoid being hit. In none of these cases do we have the requisite belief, while our behavior, arguably, qualifies as action (though the ball case can also be seen as a mere reflex movement).

Contrary to appearance, these desires need not pose a threat to the CTM. True, we should not suppose that agents in the examples have as their *motivating thought* the belief that they have reason to desire something. But this is not something the CTM requires. To repeat, what the CTM claims is that motivating reasons are constituted by a pair of beliefs that explain the agent's desires, thus ensuring that his motivation is intelligible. It is nowhere claimed that this explanation requires that the agent be aware of these beliefs. Even if we agree that belief must surface on the agent's consciousness, we are not in trouble. For surfacing only requires that some thought emerges from the agent's un- or subconscious; it does not require that the agent be aware of this thought (he may fail to notice it, for instance). (Scanlon 1998, 47; 2002, 341; Persson 1999, 163n; Dancy 1999, 215) The job the CTM requires these beliefs to do is to explain the desire, i.e. the motivatedness in a way that makes that motivation intelligible. And, arguably, they can do this even if the beliefs surface without the agent's being aware of them; in any case, we would need psychological research to prove the opposite.

This, if true, takes care of most of the 'feels like it' cases as well as Smith's example. As I noted in the previous chapter, even our craziest whims allow for an explanation of this sort. And in Smith's case we can say that just as the agent was not aware of his desire to look at himself in the window, so he was not aware of his belief that he has reasons to do so. This leaves us with the case of spontaneous action. Here it would be hard to appeal to unconscious thoughts or beliefs that have surfaced but the agent failed to take note of them. Even then we can say the following. It is possible that the agent has a more general belief, which explains his motivation in the given



case without issuing itself in a present belief. (cf. Bratman 1987, 126 on intentions) For instance, in the ball-game case we can say that the agent has a general belief that he should protect himself from flying objects. His motivation can then be seen as a response to this general idea without presupposing any particular belief being present on the occasion. In fact, we can go further than this. We can refer to Scanlon's idea that in cases of automatic and spontaneous action even if the agent has no first-order normative belief, we can still attribute to him a standing normative judgment that is not present to the agent's consciousness but is only activated on certain occasions. (Scanlon 1998, 54-5)

Finally: urges and cravings. These desires are again familiar for us from the previous chapter; therefore I don't provide a detailed description here. Some of them can be brought under the scope of the CTM. Sometimes we can attribute to the agent a normative belief to the effect that having the object of the desire would put an end to one's suffering. This is what typically happens in the case of the appetites such as hunger or thirst (recall that these desires have a two-tier structure: what we are concerned here with is the first level of basic need, not the second level parasitic upon it).<sup>105</sup> Again, note that there is no claim here about the genesis or even the maintenance of appetites. Obviously, Nagel was right when he claimed that appetites simply assail us and, just as obviously, if one is thirsty or hungry, he remains so regardless what he thinks about the normativity of the matter. Nor are we claiming that the hungry or thirsty person on every occasion has the thought that he has reason to eat or drink because that would alleviate his suffering. This, we saw above, is again not something the CTM needs. What the CTM claims, I have been emphasizing, is that the agent's motivation, hence his action is explainable insofar as he has the requisite beliefs. And though this normally happens in the case of the appetites, it need not happen. Think of the ascetic, for instance, or the person on hunger strike. They certainly

---

<sup>105</sup> János Kis has pushed me hard to confront the case of the appetites.

reject these impulses of the body: in their eyes these desires are alien, and accordingly, they deny them of any rational explanation.

This takes us to those instances of urges and cravings, presumably the majority, where even the above minimal level of desirability is not achieved. Here we can say three things. We can first try to borrow an idea of Scanlon. In characterizing attention-directed desires he claims that the thought of taking something to be a reason these desires involve can run counter to the agent's judgment, which he construes as expressing a proper normative belief. (Scanlon 1998, 41) Such a taking, he says, is more like a tendency to judge that he has reasons, a 'seeming' that need not be endorsed by the agent himself. (Ibid. 41-2, 65) Consequently, we might try to say that what happens in the case of urges is that the agent is drawn to the object of the urge, i.e. it occurs to him in a favourable light *without* the agent judging that it is indeed favourable.

The problem with adopting Scanlon's idea is twofold. On the one hand, it doesn't seem to work in the case of genuine urges and cravings. In those cases, think of Quinn's man or Frankfurt's unwilling addict or even the hunger or thirst of the ascetic or the person on hunger strike, it is hard to believe that the agent has even a tendency to judge that he has reasons to desire the thing in question. These desires simply cannot be brought under *any* evaluative/normative category. At the same time, Scanlon's construction is ambiguous. At several places he admits that on a broad enough understanding of desire, under which he admittedly means the functionalist reading, his 'takings' can be seen as desires. (Ibid. 378 32n; 2002, 337, 339; Copp and Sobel 2002, 254-7) And even if we suppose that these 'seemings' are cognitive, we still have to say what sort of cognitive states they are because, admittedly, they are not beliefs. They can be quasi-cognitive states (Chang 2004, 65), besires (Dancy 2004, 145-6) or what have you; the issue is undecided.

For these reasons I suggest that we set aside Scanlon's idea in the case of genuine urges and cravings and turn to approaches that are already familiar for us from the previous chapter. Tailored to the present context, the first points out that while the agent's urge may not admit of rational explanation, this need not mean that the agent's motivation is unexplainable. We can say that the agent's motivation has two parts: one is taken by the urge; the other comes from his desire to continue his life, which is now riddled with the problems caused by his urge. This second part, unlike the first, is rationally explainable; hence it does not escape the scope of the CTM. The second response goes further than this. It admits that urges and cravings cannot be rationally explained and that no other motivation is present in these cases. But it points out that these 'purely functional states' do not result in behavior that qualifies as action. These events are not under the control of the agent who is a mere bystander to the motivational energy flowing through his body. Although this analysis can be denied, the denial must be made on *independent* grounds, i.e. by presenting a plausible theory of action. Until such an argument is provided the CTM has just as much right to delineate the set of admissible actions as does the HTM. Smith's objection therefore fails.<sup>106</sup>

#### **IV. Summary**

In this chapter I have considered the motivational defense of the Model. I have argued that a crucial premise of the argument, the Humean Theory of Motivation (HTM) is mistaken. To show this I have examined three criteria the HTM requires to be met: the Existence Criterion (EC), the Motivational Criterion (MC), and the Intelligibility Criterion (IC). I have argued that the EC is a

---

<sup>106</sup> I suspect that the idea that behavior generated by these urges and cravings nevertheless qualify as action is based on the claim, considered in Chapter I, that there are two senses of motivating reasons: the agent's reasons and the reason's why one does something. But here we can say the same thing that we say in the text. Until this suggestion is proved to be true on *independent* grounds, the CTM has just as much right to delineate the set of admissible actions as does the HTM. Smith's objection still fails.

plausible requirement, that the IC should be accepted for methodological reasons, whereas the MC should be rejected. In doing so I have presented and defended an alternative theory of motivation, which, to illustrate the contrast with the HTM, I called the Cognitivist Theory of Motivation (CTM).

## CONCLUSION

I have begun this dissertation with the claim that reasons are important both in our everyday life as well as for the solution of certain philosophical problems. This inspired me to take a closer look at the probably most influential theory of practical reason: the Desire-Based Reasons Model or, as I briefly called it, the Model. I also said that when carrying out such an investigation we should combine the two approaches that are typical in philosophical dealings with reasons. The approach that focuses on the logical structure, metaphysics and epistemology of reasons, and the approach that concentrates on specific ideas about what gives us reasons to act. It was in this framework that I formulated my argumentation. My reasoning took the following form and produced the following results.

I have first worked out a framework for the discussion. This was the task of Chapter I. In line with the dominant view in the literature I pictured normative reasons as considerations that speak in favour of action. I then rejected the idea that the considerations referred to are mental states. This and the distinction between two rival understandings of favouring, the reason-relation and the rationality-relation, provided me with enough space in which to situate the Model. I distinguished three versions of the Model among which the first was somewhat standing apart from the second and the third that were in important respects similar to each other. On this basis, I then set out to explore and argue against these versions from different angles.

Chapter II was the first step on this road. Here I attempted to refute the second and third version of the Model through its naturalist meta-ethical background. To this end, following Derek Parfit, I presented an argument, which claimed that naturalism makes reason-claims trivial. My discussion showed, however, that this is only so if we can prove that naturalist reductions are

always eliminative and not vindicatory. Vindication requires, I argued, at least one condition to be met: the reduction proposed should be tolerably revisionist. In the remainder of the chapter I clarified the idea of tolerable revisionism and showed that naturalist advocates of the Model have difficulty in meeting this condition. Hence vindication, I concluded, remains an unfulfilled promise for the Model.

The aim of Chapter III was to show that the best way to reject the second and third version of Model, if one doesn't want to get muddled up in debates about naturalism, is to present and defend the reverse thesis: that desires are based on reasons (first premise), which they transmit but to which they cannot add (second premise). The Model reconstructs the reason-relation in a particular way: it says that the relation must make reference to a privileged subset of the agent's desires. What in opposition I claimed was that this couldn't be right, not if the two premises hold. For they claim that desires merely transmit reasons on which they are based and do not add to them, hence the role of desire in this picture is normatively redundant. This required the defense of the two premises against potential counterexamples and I spent the chapter with working out such a defense. In the course of doing so, I also considered and rejected the up to that point neglected first version of the Model as a possible source of the second premise as well as a source in itself.

This took us to Chapter IV, which considered the motivational defense of the Model. I argued that a crucial premise of the argument, the Humean Theory of Motivation (HTM) is mistaken. To show this I examined three criteria the HTM requires to be met: the Existence Criterion (EC), the Motivational Criterion (MC), and the Intelligibility Criterion (IC). I argued that the EC is a plausible requirement, that the IC should be accepted for methodological reasons, whereas the MC should be rejected. In doing so I presented and defended an alternative theory of motivation, which, to illustrate the contrast with the HTM, I called the Cognitivist Theory of

Motivation (CTM). On this basis I concluded that the last hope for the Model vanished: the investigation concerning the appropriateness of the Model as a theory of reasons finally came to an end.

Where does this leave us? The sides, I believe, are clear. It leaves us, on the one hand, in a relatively stable negative position: the Model is unacceptable and should be rejected. This may – understandably, I should add - leave one with an empty feeling since it does not tell us what alternative to opt for. This is not exactly so. Already in my introduction to the dissertation I signaled that the debate concerning the viability of the Model corresponds to a well-known antagonism in ethical theory: that between value-based and desire-based reasons. So, to this extent, there *is* an alternative, which we get to just by rejecting the Model. Of course, this is still vague but this is a vagueness I am willing to accept. It could not have been the purpose of this dissertation to work out a novel theory of practical reason. There are just too many unanswered questions on the way. To begin with, we have to face the issue of buck-passing I mentioned in Chapter III. That is, we have to decide which is the primitive, reason or value. Then we need to find a proper meta-ethical background for our alternative theory. Many in this tradition favour a non-naturalist account, but there are very severe metaphysical and epistemological worries to face here. And these are just a sample of the issues to be considered. And though they are no doubt important questions, their discussion would amount to the writing of another thesis, in fact, another book. Hence for the moment we have to rest with this conclusion: the Model is a failure and we need to find a better theory. This latter task, however, must wait for another occasion.

## APPENDIX I: REASONS – FACTS OR PROPOSITIONS?

### *I. Reasons: facts or propositions?*

Chapter I saw me arguing against the idea that the things that can be reasons ('q' in the schematic structure of reason given there) are mental states. This leaves us with two options: reasons are either states of affairs or true propositions.<sup>107</sup> The majority opinion in the literature stands more on the side of states of affairs. Dancy writes: "Intuitively it seems to be not so much propositions as states of affairs that are reasons. It is her being ill that gives me reason to send for the doctor, and this is a state of affairs, something that is part of the world, not a proposition." (Dancy 2000b, 114) And Joseph Raz agrees: "Language and our intuitions lend little support to the idea that all reasons are statements. It does not seem natural to say that the statement it will rain is a reason for me to take an umbrella. It is either the fact that it will rain or my belief that it will which would be cited as the reason." (Raz 1975, 17) But there are also exceptions. Scanlon explicitly endorses the propositional reading and David Millar and Timothy Williamson does the same in the case of reasons for belief. (Scanlon 1998, 56-7; Millar 1991, 55, 65; Williamson 2000, 194-200)

Strangely, however, in the literature there is not much discussion of which position is right. Nor have I examined the two proposals in Chapter I; here, in part at least, I would like to make good on this omission. In my brief treatment, I discuss two recent objections by Jonathan Dancy, which I think present the best case against the propositional reading.<sup>108</sup> Due to the nature

---

<sup>107</sup> In this appendix I don't distinguish states of affairs from facts or properties. I am aware that this is metaphysically far fetched, but for my purposes no such distinction is needed.

<sup>108</sup> The passages quoted from Dancy and Raz may suggest a further objection, namely that intuitions and linguistic considerations support the state of affairs account. This isn't necessarily the case. Take cases when the reasons concerned are things that can be expressed by sentences. I have in mind claims like:



of my discussion, I won't claim to have made a decisive case for the propositional account; at best I will draw attention to the complexity of the issues involved.

## ***II. Dancy's objections***

Dancy's first objection is this. Propositions, he claims, are not robust enough normatively. (Dancy 2000b, 115-6) As he puts it: "On either understanding [propositions are either classes of possible worlds or abstract objects modeling the structure of an assertoric sentence], propositions are, as we might say, too thin or insubstantial to be able to make an action wrong. They are the wrong sort of beast." Were reasons to be propositions, he goes on, "they would subvert their own purpose", that is to say, they would cease to be normative. And he concludes, "I offer this argument as conclusive against the suggestion that a thing believed, understood as a proposition, is capable of being a normative reason."

Dancy is one of the originators of the account of normative reason I have presented in Chapter I and employed throughout the dissertation. He holds that what turns a consideration into a reason, i.e. endows it with normativity is that it speaks in favour of action. It is this relation that establishes the link between reason and action and thus grounds reason's capacity to guide conduct. Accordingly, to refute the propositional reading, Dancy has to provide an argument for the view that a proposition cannot speak in favour of anything. But the latter is not an easy task to accomplish. Both Williamson and Millar argue that epistemic reasons are propositions that *evidentially support*, i.e. speak in favour of our beliefs. (Williamson 2000, 194-99; Millar 1991,

- 
- "What he said previously was exactly the reason to run and never look back";
  - "His statements clearly voiced the reasons for refusing conscription".

Since propositions are the kind of things that we say, therefore at least some reasons may turn out to be propositions. As far as linguistic intuitions are concerned, the claim that reasons are states of affairs is not the only position to hold. In general, I think that language and intuitions are considerations that can be used to back either position, and even if they stand more on the side of the state of affairs reading, they are just not decisive.

55, 65) I don't see why this could not be the case with practical reasons as well; and, for one, Dancy provides no argument against the idea. Propositions in this crucial respect are not as 'insubstantial' and 'thin' as he claims them to be.<sup>109</sup>

Dancy's second argument, however, cuts deeper. It goes like this: things that have truth-value are representational; since propositions have truth-value they are representational; reasons are not representational; hence reasons cannot be propositions. (Dancy 2000b, 117) This argument needs a bit more fleshing out. The idea seems to be the following.<sup>110</sup> Representational things are transparent: we always look through them to see what they represent. The same happens with reasons. When we look for reasons we don't stop at propositions, but go directly to the facts they represent. Hence propositions are redundant and therefore fall out during one's quest for reasons. As Dancy eloquently puts it, there is an 'ontological gulf' between things that are capable of being true and things that are capable of being the case. Being instances of the former, propositions cannot serve as normative reasons for anything.

The propositional theorist has two responses open to him. One is to give up the idea of factual propositions. This would be something like a Fregean position, which takes a proposition to be an object of thought that has no truth-maker, no fact independent of the proposition standing in the correspondence relation to it. This does not mean that these propositions cannot be true or false, the idea is only that we can do without a robust truth-maker. Recently, John Skorupski has advanced a theory - he called it 'irrealist cognitivism' - that uses this idea. (Skorupski 2000, 116-

---

<sup>109</sup> To be fair, Dancy has a response open to him. Already in Dancy (2000b), pp. 1 and later in (2003a), pp. 101, 106-7 and in (2004), Chapter 2 he makes a distinction between the favouring relation that connects the reason to the action and the *making-it-the-case* relation that connects the reason to the oughtness of the action. On this ground he further claims, in personal communication, that we need to avoid any serious dislocation between what we say about favourers and what we say about ought-makers. Hence reasons should be things that are suitable for both relations, but propositions are not like that: it is difficult to see how they can make it the case that one ought to act. The reason why I don't mention this objection in the text is that, as I briefly discuss in note 22 of Chapter I, it is not at all obvious that we should make Dancy's distinction.

<sup>110</sup> I owe this suggestion to János Kis.

8; 2002, 133-8; 2006) But this solution also has its price. There is first the question whether all propositions are non-representational or only some of them. Skorupski, for instance, leaves it open whether the things that can be reasons are factual propositions or not, but insists that the favouring relation is the object of non-representational propositions. The question is how one can reconcile these two things without giving up both. Moreover, there is the well-known epistemological challenge to views of this kind. It is hard to see how one can give a proper epistemological account of our knowledge of non-representational propositions. Since on this view there are no truth-makers, it is anything but obvious how we can come to know these propositions.<sup>111</sup>

The second response appeals to our conception of the role reasons play in human deliberation. It seems to me that the correct view on the role of reasons is the following: they are the kinds of things that figure in our practical deliberation, elements that we are responsive to when making decisions. It is obvious that what prompts us to act, or rather, to consider different ways of acting are states of affairs. But *what* we consider, weigh, entertain and face up to in practical reasoning are the propositions that represent these states of affairs. We just don't have a *direct* connection to states of affairs in the world; we can only get acquainted with them through the propositions that represent them. Therefore there is no way we can do without propositions; although propositions are representational they do not fall out in our quest for reasons – for they are exactly the kind of things we are looking for. To this, however, one can say that I confuse two things: the objects of our thoughts (these are states of affairs) and what expresses the content of our thoughts (these are propositions). It is the former and not the latter that we weigh up,

---

<sup>111</sup> Skorupski's response, which centers on Kant's distinction between spontaneity and receptivity, is illuminative but far from complete. See for instance the problems Dancy mentions in Dancy (2000a). In fairness, however, it must be noted that Skorupski has since then developed the idea further in Skorupski (2002) and (2006). My decision to avoid detailed discussion of his position is therefore more a result of his often stressed claim that the things that can be reasons are (or, at least, can be) states of affairs and reported by factual propositions.

consider, and accept or reject in reasoning, hence reasons are states of affairs and not propositions.<sup>112</sup> I don't find this claim obviously true. I think we agree that reasons constitute the content of our reasoning; what we differ on is whether this content is given by the objects or by the content of our thoughts. And there is as much to be said in favour of one reading as of the other.<sup>113</sup>

---

<sup>112</sup> Joseph Raz made me confront this point. The idea also appears in Searle (2001), pp. 36 as well. In addition to the response in the text, one might also try to say that propositions provide not only the contents of beliefs but can also be their objects. This idea appears in Sillins (ms). His argument is this. We can have true or false beliefs; propositions are typically the kinds of things that can be true or false; hence, it seems reasonable to say that a true belief is a belief in a true proposition, whereas a false belief is a belief in a false proposition. Since I find this argument indecisive, I don't mention it in the text.

<sup>113</sup> In addition to the negative defense of the propositional reading presented in the text, a positive argument may suggest itself. It might be claimed that the construal of reasons as states of affairs has a hard time with the individuation of reasons. Take the following example. There are plenty of split personalities in cartoons about superheroes. Just think of characters like Spiderman, Superman or Batman. What can we say about the reasons we have in connection to these people? We may want to claim that the reasons connected to the different personalities locked in the same body are different. The reason to save Superman when he is in trouble is different from the reason to save Clark Kent when he is in trouble. However, another intuition of ours tells that people with split personalities constitute one state of affairs. Superman *is* Clark Kent: he is the same particular with the same, physical and psychological, properties. As a result, on the states of affairs reading we should have to say that the reasons in our examples are the same, which is not the case. The propositional understanding of reasons, on the other hand, is committed to no such view. The reason *that* Superman is in trouble is different from the reason *that* Clark Kent is in trouble. Although these propositions may entail each other, they don't use the same concepts; the same fact is reported by different propositions. For this view of the individuation of propositions see Schick (1991), pp. 72-8. Hence, on the propositional but not on the state of affairs reading, reasons are easily separable. The problem with this argument, a version of which appears in Sillins (ms) as applied to reasons for belief, is that it operates with a very simple-minded account of the individuation of states of affairs. Therefore, in order for the argument to succeed, more refined theories should also be considered and defeated. I thank János Kis and Krister Bykvist for pointing out this complication to me.

## APPENDIX II: KORSGAARD'S CIRCULARITY ARGUMENT

### *I. Korsgaard's argument*

In an influential argument Christine Korsgaard has launched an attack on the non-naturalist or, to use her term, dogmatic rationalist account of practical reason. (Korsgaard 1997, 240-1; 2003, 110; ms a, Lecture 2, 13) Dogmatic rationalists, she says, hold that there are 'eternal normative verities', i.e. irreducible facts about what we have reason to do and it is our knowledge of these facts that we apply in action. But, Korsgaard goes on, the agent who is facing this claim can surely ask: "Why should I care about these facts?" "Why should I apply this knowledge in action?" And there is no answer, she says. This is a problem, she goes on, because it leaves practical principles without justification: even the most uncontroversial ones lose their grounds.

Her primary example is the instrumental principle. (Korsgaard 1997, 241-2) Very roughly (here the details don't matter), the principle says that one has at least *prima facie* reason to take the means to one's ends. But if the fact that this is what we have reason to do is just another 'eternal normative verity' that we apply in action then, by the above logic, one can ask why one should care about that fact. And what can the rationalist say to this? He can try the following route. One, perhaps necessarily, has the end of taking the means to one's ends; one has a reason to do what promotes one's end of taking the means to one's ends; taking the means to *this* end does just that. Obviously, for this argument to work we have to show that we indeed necessarily have the end of taking the means to our desires. But this is only the smaller problem and, as we shall see, it is not impossible to prove it. For the opponent can point out that even if we necessarily have the end of taking the means to our ends, the fact that we have a reason to take

the means to *that* end is just the same old fact, which we have no more reason to care about than before. Consequently, instead of a solution, we get circularity.

Nor is the rationalist better off if he tries to widen the scope of the first premise of his reasoning. For what can he say? There are two options. (Korsgaard 2003, 110-2; ms a, Lecture 2, 14-5) Either he says that the end referred to in the first premise is the end of doing good action; or he says that it is the end of doing what is supported by reasons. But this produces the same difficulty as before. Imagine how the argument would work. It is our, perhaps necessary, end to do good action; taking the means to our ends is itself a means to our end to do what is good; hence we have a reason to take the means to our ends. We thus fall into the same trap: in arguing for it, we presuppose the principle itself. This should not be surprising. After all, our problem arises because we bump into the question of application. But the instrumental principle is the principle of application itself, thus it is no coincidence that we try to justify the principle with itself. Circularity is thus an unavoidable consequence of the rationalist account of the instrumental principle, and this puts the rationalist in deep trouble.

## ***II. Extending the argument***

The first issue we have to deal with when considering Korsgaard's argument is whether it is relevant for us at all. Korsgaard explicitly speaks of dogmatic rationalists as her target and, as I noted, with this term she refers to contemporary non-naturalists and their predecessors such as the intuitionist Samuel Clarke or Richard Price. But I made it clear in Chapter I that non-naturalism does not offer a proper meta-ethical background for the Model. So before we proceed to any kind of substantial analysis of Korsgaard's argument, we have to tackle two questions. First, we have to ask what sort of justification Korsgaard has in mind and, second, whether the demand for

justification understood in this way can be extended to the naturalist account of the instrumental principle as well.

Start with the first question. The issue here is what drives Korsgaard's question of application. I think it is fairly clear that she has in mind a *practical* challenge: any rationalist account of practical reason must show that the property it identifies as an ethical property has a bearing on our deliberation and conduct.<sup>114</sup> Korsgaard herself puts the problem in the following way. She says that when dealing with practical issues what we are dealing with is motivation, but not any kind of motivation. Take the case at hand: the idea that we should take the means to our ends. We are ordinarily motivated to take the means to our ends: the bare co-presence of an end and a suitable instrumental belief is enough to 'effect a motive.' But, she points out, this motive may be the result of mere *causation*: one can simply be so *conditioned* that he always takes the means to his ends. (Korsgaard 1997, 221) What we need, therefore, is that the motivation be the result of the agent's own *recognition* of the appropriate conceptual connection between the belief and the end: that he is moved to act because he thinks he has *reason* to act. (Ibid. 243) In this way, we will have the agent put into the picture: the act will be the result of the agent's own

---

<sup>114</sup> This issue is well-known at least since Aristotle. On the rationalist (more generally, realist) model reasoning ends up with a judgment understood as a belief. However, practical reasoning is practical exactly because it should have a direct bearing on action: it should ultimately take us to action. So the question is how rationalists can bridge this gap between the theoretical conclusion of reasoning and the action it is supposed to lead us to. Although there are many who reject this view of practical reasoning, claiming that it should conclude in judgment not in action, here I don't want to get entangled in this debate. I take the challenge to be genuine and see what comes out of it. I think the same or similar point appears in Nowell-Smith (1954), pp. 41; Hampton (1998), pp. 47; Copp (1990), pp. 246-51; Wedgwood (2002), pp. 143-5; Nagel (1970), pp. 63-5, 143-4; Williams (1981b), pp. 122; Darwall (1990), pp. 260. Railton's second criterion for tolerable revisionism also comes close to this sort of demand and, as I suggest in a footnote in Chapter II, this may actually provide another objection against the second version of the Model, since we may try to run the circularity argument in that context too. See Railton (1990), pp. 173 and Railton (1993b), pp. 324. I should also note, however, that none of the authors mentioned above are clear on what interpretation - of the two I am about to present - they have in mind; in fact, Wedgwood comes up with a third interpretation: he thinks the challenge arises because one can always ask whether he has *most* reason to do what is demanded of him. I see, however, no reason to read Korsgaard in this way, nor do I think there is any need to do so.

mental activity “and not merely the result of the operation of beliefs and desires *in her*.”<sup>115</sup> (Ibid. 221) Korsgaard calls this ‘rational motivation’ and thinks that this is what every theory of practical reason must account for.

A theory can fail to meet this demand in two ways. (Korsgaard 1996b, 12-6, 42, 46-7, 81; 2003, 112; ms a, Lecture 2, 15-6) In the first case, we are concerned with *motivational* issues, accordingly, if we fail here, we will fail to account for the motivation to act on one’s normative judgement. What we investigate here is whether normative judgments necessarily motivate or only occasionally, and whether normative beliefs or only desires are able to move us to act and how the two issues are connected. This is what Korsgaard calls the criterion of explanatory adequacy. In the second case, we are concerned with the *justification* of that motivation, accordingly, if we fail here, we may have the motivation but the judgement involved won’t be a *normative* judgement: it will lack justification in the sense that the agent won’t see reason to act. Consequently, once the agent sees what is behind practical claims according to the given theory, he will refuse to endorse his own motivation. And if he does that, his action, though it might still be explained on the given theory, will be the result of mere causation and not rational motivation. This is what Korsgaard calls the criterion of normative adequacy. A proper theory, Korsgaard claims, must meet both criteria; but she also makes it clear that her primary concern is the second criterion: she calls it the normative question.

This puts our original problem in context. Korsgaard’s question of application makes a practical demand, but now we know that this demand can take forms. Korsgaard herself appears to be clear on which form the circularity argument uses. When she deals with this problem, she

---

<sup>115</sup> This is not the full picture, though. First of all, Korsgaard brings in principles of choice as the basis for our decision of what counts as a reason for us. But she does so in an un-Kantian way: she claims that we identify with these principles (which, according to her, also govern our various practical identities). This is how, she says, the agent appears in her theory. See Korsgaard (1996b), pp. 239-241. Second, it might be that in addition to this recognition and identification with a principle of choice, we also need certain autonomy-making motives to be present. Korsgaard is not clear on this demand: see next footnote on this.



normally describes it as a motivational issue: the agent who faces the truths or facts rationalists propose is just not motivated to act on them. (Korsgaard 1996b, 37-8; 1997, 240-1) The criticism inspired by the other form of the demand she preserves for ‘empiricists’, i.e. naturalists whose primary representative she takes to be Williams. (Ibid. 1997, 219) Here she argues that the empiricist account of the instrumental principle cannot guide action because there is no way one can violate it. (Korsgaard 1997, 223-33; ms a, Lecture 2, 9-10) In absence of error, however, we only have the agent acting on a purely causal basis: there is no rational motivation involved since there is no reason to act. I propose to set aside this argument since it is no concern for us here; I assess it in Appendix III. Our question should be whether the circularity argument and the question of application can only be given a motivational reading, or the alternative reading is also available.

I think it is. To begin with, there are places where Korsgaard herself puts aside the motivational reading. At one point, she says that the dispute between Hume and contemporary rationalists concerning the motivational power of moral perception ‘may just be a standoff’. But, she claims, “if we think about normativity, *rather than motivation*, then we will find that there is something in Hume’s complaint.” And then she goes on to introduce the question of application. (Korsgaard ms a, Lecture 2, 13, Italics are mine) Also, in her first employment of the question of application, she uses the question to refute theories on the ground that they fail to answer the normative question. (Korsgaard 1996b, 28-48) There are, moreover, just too many ways of overcoming the motivational challenge, other than the one Korsgaard herself considers. There are internalist as well as externalist alternatives, the former I have even discussed in Chapter IV.<sup>116</sup>

---

<sup>116</sup> Korsgaard’s position on motivational matters is unclear, at least to me. In Korsgaard (1997), pp. 220 she says that the idea that beliefs can motivate is mysterious but does not dwell on the issue further. Although I cannot argue for this here, I think this is indeed her position. Then the question becomes what she means by the claim that reason alone can motivate, which, as a Kantian, she holds. In Korsgaard (1986) the answer seems to be that there are

Attila Tanyi, THESIS, Page 201

Finally, the question of application appears to be neutral as to the two rival readings. One can just as well ask the question because, though he is moved to act on his knowledge, he sees no reason why to: he is reluctant to endorse the motivation. That is, in accordance with the two points just made, we can grant the rationalist some account of motivation, but still think that agents would refuse to act on this motivation once they see what is behind the claims these theories make on them, i.e. once they see what is supposed to guide their conduct.<sup>117</sup>

This also connects us to our second original inquiry: whether the question of application can be extended to naturalism as well. We just have to look at Korsgaard's more general philosophical enterprise. It is to argue against what she calls substantive realism by employing the question of application. She defines substantive realism by contrasting it with her own procedural realism. (Korsgaard 1996b, 36-7) The former holds that there are procedures for answering normative questions because there are ethical facts or truths that exist independently of these procedures. Whereas the latter claims that there are answers to normative questions because there are procedures for arriving at them: there is no need to suppose the existence of ethical facts

---

motives - a preference for one's greater good and an acceptance of moral principles - which are constitutive of rational agents. But then when she gets to the point of claiming that taking the means to one's ends is constitutive of agency, she rejects the idea that we should understand this as implying that agents have a disposition to act in this way. See Korsgaard (1986), esp. pp. 18, 22; (1997), pp. 245; (1998), pp. 53, 56; cf. Velleman (2000c), pp. 172-3 and Rawls (1980), pp. 525, 528. Alternatively, Korsgaard can hold the position that normative beliefs, by virtue of their content and causal power, can bring desires into being, which then move the agent. Korsgaard associates this idea with empiricists and accepts it as a possible solution. See Korsgaard (1997), pp. 219. But, of course, this does not mean that she holds this position. Nor does she distinguish it, as I think she should, from what in Chapter IV I call the Antecedent Motivational Thesis (AMT), which might indeed be her position as a Kantian (see my discussion in note 14 of Chapter IV on this).

<sup>117</sup> Here is some more support for this point. When one asks the question of application, one is shaken in one's confidence that the recommended course of action is indeed what one ought to do. And, Korsgaard explains, the rationalist can only respond by repeating that indeed, this is what one ought to do. But, she goes on, this is a way of refusing to answer the agent's doubt, since it is just insisting that there are certain truths or facts about reasons, while the agent sees none. See Korsgaard (1996b), pp. 38-41. I don't see why this lack of confidence need be interpreted as a motivational problem; in fact, as I point out in the text, in the just quoted passage and elsewhere nor does Korsgaard see things in this way. Of course, though his confidence is shaken, the agent might still be motivated to act because he has a strong instinct to do what he thinks he ought to do, or because his normative beliefs move him on their own and so on. Then perhaps this motivation just evaporates as the agent judgment ceases to be a normative judgment, or it may persist with the agent refusing to endorse it and act on it. But this is a different story, and not one we are interested in here.

or truths independently of these procedures. Now, it is clear that naturalism and non-naturalism belong to the same substantive realist camp. For both ethics is a theoretical enterprise: to find out and then apply in practice the ethical knowledge we gain in the world. (cf. Korsgaard 1996b, 38, 40, 43-44) Hence they both invite the question of application, and if what I have argued for is accepted, this can take the form of the normative question. The naturalist Model is different from this general picture only in the way it fills in the details: it speaks of facts of desire-satisfaction. Consequently, we just have to substitute ‘desire’ for ‘end’ in Korsgaard’s circularity argument and then follow the same reasoning as before.

### ***III. Responses to the argument***

I suggest therefore that we read the circularity argument as posing a normative, not merely motivational challenge against any kind of substantive realist view, including the naturalist Model. We should first realize that the problem Korsgaard describes, though technically not an infinite regress, is nevertheless something very similar to it: it begins with a question and then, instead of getting an answer, we end up asking the question over and over again. Consequently, the types of responses also follow the strategies one approaches an infinite regress with. There are three: to show that questioning doesn’t start; to accept that it starts but argue that it can be stopped; or to admit both that it starts and that it cannot be stopped but claim that this is not a problem. In what follows I assess all three types of responses. I start with the first, continue with the third and end with the second.

#### **A.2.3.1. First response**

The best representative of the first response is an argument by Peter Railton. He begins with the familiar observation that “a substantial amount of our rational, intentional activity must be

‘automatic’, unmediated by reasoning and recognition.” (Railton 2004, 185) On Railton’s view what happens in such situations is that the agent relies upon his senses, memory or thoughts blindly. He *trusts*, that is to say, attributes default *authority* to his senses, memory or thinking. He expects that things are as he perceives them to be, or that they happened in the way he remembers and thereby also learns important information about his environment or about his own thinking. (Ibid. 187) Although the trust involved is defeasible, it serves Railton’s purpose well enough. For it allows that the agent’s intentional activity may not involve an element of judgement, while taking his activity at the base justified. And this rules out a regress-type of questioning. At the ground level, default trust provides sufficient justification but being non-judgmental and passive, it doesn’t invoke standards of reasoning, which would stand in need of justification. Circularity is avoided.

However, the claim that the need for justification elapses because we have default trust in our senses, memory and thinking is puzzling. Railton himself names the problem. “But how could so passive and non-judgmental an attitude as default trust be the foundation for a form of reason-disciplined agency?” he asks. (Ibid. 191) But he also has a solution. (Ibid. 191-4) His idea is that while both belief and desire are instances of default trust, they are also *normative* attitudes. They are normative because their nature is to realize a certain normative *role* in the individual’s mental architecture: truth-tracking and value-tracking respectively. And they are instances of default trust because they need not involve judgement: it suffices if they constitute an *expectation* that something is the case or that something is desirable. Through belief one learns how things are in one’s environment or how one’s thoughts fit together and through desire one learns about what is valuable and what is not.

The question is whether the account delivers the results Railton expects from it. I don’t think so. Recall first Scanlon’s idea that I have often referred to in this dissertation. He claims

that what sustains automatic intentional activity are certain *standing normative judgments* to the effect that if some putative evidence is not good ground for forming beliefs, or if certain reasons are not good grounds for action, then one does not even unreflectively form beliefs on the basis of such evidence or act on the basis of such reasons. (Scanlon 1998, 24) Since these judgments are standing, they are not present in the agent's consciousness but are only activated on certain occasions. Yet, since they *are* judgments they invoke standards of reasoning, which in turn invite justification. And though the need for justification rarely arises, perhaps it happens only in cases of great distress, this is still enough to open the ground for questioning and circularity.

This idea serves as an ideal supplement to Railton's proposal. We can say that, in virtue of their normative role, desire and belief provide the material for our reasoning: what they mediate serve as 'candidate reasons' for us.<sup>118</sup> We can also grant that these attitudes are instances of default trust: they function automatically without reasoning and judgement. This is not an unusual thought, others, such as Korsgaard or Scanlon, are also willing to grant this role to desires (though their grounds are different). (Korsgaard 1998, 51-4; Scanlon 1998, 65) But, and this is the important bit, what they are directed at is just an apparent reason, which becomes a real reason, i.e. a normative reason only after the agent has decided in its favour. And if Scanlon is right, this decision and the subsequent forming of the belief or action on its basis, is governed by the standing normative judgments he describes. Consequently, our intentional activity, even when it is spontaneous and automatic, is the result of a process that comprises both the aspect of default

---

<sup>118</sup> This resembles a lot to the Davidson/Stampe/Hurley line of thought I presented in Chapter I. But there are two important differences and I try to highlight these already in the text. Yet, for further emphasis, let me repeat them here. First, I read Railton as holding the view that mental states, by virtue of their function, only *mediate* reasons, but are not themselves reasons. This, as I point out in Chapter I, is the natural reading of the tracking function Railton refers to. Second, Railton's speaks of 'function' in his presentation and it is not clear what can be squeezed into this notion. In particular, it is a question whether the counterexamples mentioned in Chapter I with regard to Davidson's theory speak against this function or not.

trust and that of norm-governed decision.<sup>119</sup> (cf. Korsgaard 1996b, 243) We are back where we started: questioning gets off the ground.

### **A.2.3.2. Third response**

Turn now to the third response. It accepts both that questioning gets off the ground as well as that it is circular. But then it adds that no problem arises from this: an infinite chain of questions is not vicious. The idea is well formulated by David Copp. (Copp 1995, 43-4) He distinguishes between the *process* of justification, i.e. the agent's performing the steps of justification and the *status* of being justified, i.e. the claim's having a place in an infinite chain of justifications. Like in geometry: perhaps there is one theorem of geometry only if there is an infinity of theorems, but this is not to say that there is one theorem only if an infinity of theorems have been proven to be such. Copp claims that in the matter at hand what we need is the first reading, and we need it only in a weak sense: the fact that a principle has a place in an infinite chain of justifications is not sufficient to show it *not* to be justified. If this is true, our problem evaporates. A principle can be justified just because it has a place in an infinite chain of justifications; it is not needed that the agent goes through the infinite steps of justification in order to confer justification on the principle. The same is true of the instrumental principle: it can *be* justified without us, the agents justifying it.

Of course, we would then still have to say what that justification consists in. But this is not difficult for we can simply revert to the original realist accounts of the principle. The crucial

---

<sup>119</sup> This also gives us the answer to William FitzPatrick's recent suggestion. He claims that the circularity argument doesn't work because one can account for rational motivation without explicitly employing the principle itself. In a particular case, he says, the rational connection between the principle and behavior is this: one recognizes something as a reason by virtue of its bearing on an act's instrumental relevance to one's end. That is, what one recognizes is that the normative relations the principle is meant to articulate instantiate in the particular case - and this, he says, one can do without direct appeal the principle itself. See FitzPatrick (ms), section three. I agree with this point, but I think it fits well into the framework described in the text: it is exactly this framework that makes possible the spontaneous functioning of the principle.

issue is not this. What carries the argumentative weight in Copp's argument is his claim that in the present case what we need is the first reading. This choice clearly builds on a particular understanding of justification, what we might call third-person justification. It claims that the agent need not have access to the full justification of a given principle; he need not be able to comprehend it. But this is only an *assumption*: Copp does nothing to support his choice, only assumes that he is right. And it is not obvious that he is right. Korsgaard, for instance, makes it clear that the need for justification arises from a first-person standpoint: it is the agent who asks the question about how to regulate his conduct. Hence the answer to the question had better be accessible to him, had better be something he can grasp and appreciate the import of. The third-person standpoint, Korsgaard admits, also appears in a theory of practical reason but it has a different role. (Korsgaard 1996b, 16-7) We saw what this role is. It is the task of motivational explanation, when we show how and why practical claims have psychological effects on the agent.

The question, then, is which of the two readings we should opt for. I don't think there are decisive reasons in favour of either reading. But there are certainly good reasons that support the first-person reading, so all I will do here is to list those reasons. Here they are. First, attributions of justification try to pick out agents as acting conscientiously, i.e. in a responsible, blameless manner with regard to what they should do. Therefore agents should at least be given the opportunity to detect the justification of their actions. Perhaps they won't always seize this opportunity, but if they did they could find the missing justification. (Radzik 1999, 39) Second, justification is a regulative notion. It signifies a property that people should be able to think usefully about in order to decide how to act. (Ibid.) What we expect from a theory of justification are guides for our action and this is only possible if we are capable of detecting what the justification of a given act consists in. This idea is familiar from debates on indirect

consequentialism; it is often called the transparency or publicity condition. (Rawls 1971, 130; Williams 1973, 128; 1985, 101-2) The claim is that a theory that requires widespread ignorance of its account of justification is self-defeating. For its distinctive contribution to ethical theory is exactly this account, and the distinctive interest of such an account is to provide a basis for decision-making. Therefore, when a theory doesn't require people to be aware of its account of justification it abandons the very basis on which its own foundation, as a distinctive ethical theory, is laid.

### **A.2.3.3 Second response**

Our best choice, then, is to stop questioning. This takes us to the second response. Here we find several alternatives. Let us begin with the simplest one. Recall Korsgaard's argument. Her basic problem is that a realist construal of the instrumental principle always leads to the question of application. But this invites the obvious response. Suppose we don't try to justify the principle with itself but with reference to another, more basic norm. What then? Korsgaard thinks that this wouldn't work. Due to the problem of application, she says, such a move would only produce a chain of justification, i.e. an infinite regress of norms. This is why the instrumental principle is so crucial: in their efforts to avoid regress, realists must employ the instrumental principle. (Korsgaard 1997, 242; 2003, 111-2; ms a, Lecture 2, 14-5) We saw how this would work: say that conforming to the more basic norm is an end of the agent, then claim that following the instrumental principle is a means to satisfying this end. But then you realize that you implicitly re-employ the instrumental principle. The problem, to repeat, is that the instrumental principle is the principle in accordance with which we apply truth in practice, hence it cannot be used as its own support.



However, according to some philosophers, this response only begs the question. It squarely denies what they think is possible: that there are principles in the case of which the question of application does not arise. (Parfit 1997, 121-9; ms: earlier draft, Chapter 5) And indeed, Korsgaard makes it clear that on her view no such appeal is acceptable: it is a refusal to answer the ‘normative question’. (Korsgaard 1996b, 34, 39-41; 2003, 112) I don’t want to settle this debate here; instead, let me point out two things. First, on the face of it, the present response is not available to advocates of the Model, since they take the instrumental principle to be fundamental. Moreover, those who make this response are not naturalists but non-naturalists, thus the meta-ethical background they use to stop questioning is again not available to naturalist advocates of the Model. Yet, and this is my second point, naturalists may point out that at least on its second version, the instrumental principle does not exhaust the Model. True, what supports the principle is not another principle but, recall Hubin’s point, it is the claim that desires are ‘brute facts’ like the basic norms in Hans Kelsen’s legal theory. Consequently, the naturalist would just say that an action’s being justified is the very same thing as its subserving certain of the agent’s desires, full stop. Following Darwall, we can call this view *deflationist justification*. (Darwall 1990, 259)

Darwall says that he knows no way of proving deflationism false, even though, as his preceding discussion shows, he has much the same concern and finds the same questions pressing as do Korsgaard. I agree with him, so I suggest that we go on and see if there are other workable and more attractive solutions available. The next attempt starts with a distinction. According to David Velleman, the object of any enterprise is either formal or substantial. (Velleman 2000c, 176) A formal specification gives us the concept of the object of the enterprise: ‘winning’ in the case of a game, for example. A substantive specification, on the other hand, specifies what it is to achieve the formal aim: to run the fastest time, for example. Put in this framework, the formal

object of practical reasoning is to do what one should do, whereas the substantive aim could be anything including, as on the Model, doing what satisfies one's desires. Using this distinction, Ralph Wedgwood has claimed that there is a way to meet Korsgaard's objection. (Wedgwood 2002, 142, 147; 2005, 468) For it is hard to see what sense would be in asking why one should do what in the formal sense he should do: it would be like asking 'Should I do what I should do?' Sure, one can just announce that he sees no point in acting for reasons. But then he opts out of practice altogether and can just as well commit suicide: his life is devoid of all value and is pointless. At the same time, for the others the question of what to do is settled. There is no question of application; their conclusion is regulative of their choice.

Wedgwood's proposal, however, does not necessarily provide us with a solution. First, the instrumental principle appears to articulate a substantive aim, Wedgwood certainly treats it so, so how do we get to it? Wedgwood himself is not concerned with this question. Yet, he is wrestling with another problem and his reasoning may give us a hint. (Ibid. 147-8) In response to Velleman's objection that the notion of a formal object is empty, he points out that the formal reading is compatible with specifying what one in the formal sense should do. That is, it does not deny the existence of 'non-trivial general truths' akin to Velleman's substantive aims. What it claims is that these truths are not given to us in advance of our deliberations about what to do; instead, we have to discover them there.<sup>120</sup> But once we found them, once we know what in the formal sense we should do, we have to act accordingly. (Wedgwood 2005, 468) Yet, this still

---

<sup>120</sup> The contrast here is with 'basic principles of rational choice'. It is interesting that Wedgwood doesn't say much about this, for him crucial category. His main idea is that certain principles are given to us "merely in virtue of our being rational beings" by which he seems to mean that it is constitutive of rational agents that they have a disposition to follow these principles. See Wedgwood (2002), pp. 144, 147. This is a possible interpretation but it is puzzling in light of Wedgwood's own endorsement of Railton's criticism of such constitutive arguments that I will soon present. See Wedgwood (2005), pp. 465. Wedgwood's talk of principles is also strange. For he seems to suggest that these principles are substantial enough - his analogy with the principles of logic and mathematics seems to suggest this at least. But then it is mysterious why he thinks that Velleman's emptiness objection is a serious problem that should be handled through fleshing up these basic principles.

doesn't give us the instrumental principle. There is no assurance that the principle will be among the truths discovered and that it will be the only one. Nor is it clear what the meta-ethical standing of these truths is. It can turn out that what we get in the end is a non-naturalist account. And, according to my reasoning in Chapter I, such a meta-ethical background is not suitable for the Model.

Up to this point, our strategy has been to tackle Korsgaard's argument from the theoretical side by trying to show that certain normative truths can stop the circle of questioning. Let us approach the problem from the other, practical side now. Here is where Korsgaard's own proposal comes into view. It is built up of several steps. First, she suggests that we should approach the problem of justification as a problem of practical problem-solving. We should start with a real practical problem the agent has to solve and then show that the given principle does indeed solve that problem. As she puts it, "If you recognize the problem to be real, to be yours, to be one you have to solve, and the solution to be the only or the best one, then the solution is binding upon you." (Korsgaard 2003, 116) Now, the question is what that problem is. Korsgaard's answer is simple: action. Our plight as self-conscious beings is that "we find ourselves with the necessity of making choices and so in need of reason to act." (Korsgaard 1998, 62) "Human beings", she says, "are *condemned* to choice and action ...[this] is the simple inexorable fact of the human condition." (Korsgaard ms a, Lecture 1, 1-2)

The next step is to clarify what Korsgaard means by this claim and how the instrumental principle fits the picture. The two issues are connected, so I don't separate them either. Korsgaard's views allow for two interpretations. (FitzPatrick 2005, 664-5; ms) On the first reading, certain principles, including the instrumental principle are literally necessary to exercise agency, that is, we need them to be able to act at all. We find two variations here. (FitzPatrick ms) The first starts from the widely accepted idea that in order to remain an agent one must have

ends. Having an end is *constitutive* of being an agent: acting is a teleological enterprise.<sup>121</sup> (Korsgaard 1996b, 122; 1998, 51, 60-2) Then a further thought comes: willing the means is *constitutive* of willing the end. That is, something the realization of which does not at all concern us in our deliberations, cannot qualify as an end for us. It is like walking and putting one foot in front of the other: one cannot walk unless one puts one foot in front of the other. (Korsgaard 1996b, 36; 1997, 249) If we add these two ideas together, we get this: taking the means to our ends is *constitutive* of agency. We cannot act unless we take the means to our ends. And since action, unlike walking, is ‘our plight’, the instrumental principle is justified for us.

The second variation takes the claim about constitution for granted but combines it with a further psychological thesis: practical principles are necessary for the *unification* of agency. In particular, those who don’t follow the instrumental principle will disintegrate as their agency degenerates into a passive arena for the operations of competing desires. (Korsgaard 1997, 247, 254) In her most recent writings Korsgaard puts this idea at the core of her views about justification and normativity. “The necessity of confirming to the principles of practical reason”, she says, “comes down to the necessity of being a unified agent...[which] comes down to the necessity of being an agent...[which in turn] comes down to the necessity of acting...[which] is our plight.” (Korsgaard ms a, Lecture 1, 19, 17) Korsgaard thus gives us a second and perhaps even more pressing reason to conform to the instrumental principle. The principle is not only constitutive of our agency, but is also something we must in the majority of cases follow if we are to maintain our integrity as unified persons. Too many violations of the principle result in disintegration: our agency will fall apart making us incapable to act. Again, the principle is justified for us: questioning is stopped.

---

<sup>121</sup> In my treatment I disregard the problem, familiar for us from several chapters of this dissertation, that certain acts may have nothing to do with the furtherance of the agent’s ends. We need not go into this issue in order to show that Korsgaard’s idea does not help the defense of the Model.

The same problems beset both proposals. There is good reason to think that they are not correct and even if they are, they don't help the defense of the Model. Let us proceed in reverse order. Korsgaard's driving thought is the idea that principles of practical reason serve as solutions to the practical problem of acting. But neither of her proposed solutions employ normative truths and facts; they are left out of the picture altogether. Nor are they needed. If either of the above variations is correct, conformity to the instrumental principle is literally practically necessary. And this is not surprising. Korsgaard intends her account as the constructivist *alternative* to realism: normative concepts name the problem and the principles propose the solution. There is no aim to track normative facts outside the will; instead, practical principles emanate from the will. (Korsgaard 2003, 116) In addition, both variations encounter problems of their own. The appeal to the preservation of agency by avoiding disintegration works only in a general way, pointing to the problem that will plague us if we *regularly* fail to take the means to our ends. It therefore cannot explain why someone should obey the principle in a *particular* case, which is obviously the kind of requirement that the idea of practical justification is premised upon. (FitzPatrick 2005, 674; ms)

The claim that the instrumental principle is constitutive of agency, on the other hand, can be read in two ways. The first conforms to the literal practical necessity account and holds that in order to remain an agent one must *actually* employ the principle. But it is not impossible to imagine cases in which one does not act on the principle; in fact this *has* to be possible since, as Korsgaard herself emphasizes, it must be possible to violate the principle if it is to count as normative. (Korsgaard 1997, 229, 244; ms a, Lecture 2, 9-10) Hence the best solution is to give up the literal practical necessity reading and look for an alternative interpretation. Korsgaard's candidate is this. (Korsgaard 1996b, 36; 1997, 245) We can say that in willing an end the agent is *committed* to taking the means to that end. This is why there is a problem with failing to take the

means to one's ends: it involves a failure to follow through on one's commitments. This solution also preserves the claim of constitution. Although actually acting on the instrumental principle is not constitutive of willing an end, thus of agency, it is nevertheless something the agent *implicitly endorses* while willing the end. (FitzPatrick ms) We cannot just shrug off the principle; it is justified for us.

I can accept this account of the principle. Yet, it is not obvious that it also solves the present problem. The issue here is not that advocates of the Model cannot appeal to this understanding of constitution; as a matter of fact, to the detriment of Korsgaard's constructivist enterprise, they can. (FitzPatrick ms) They can say, think of the first version of the Model, that the instrumental principle is grounded in a fact about the internal relations among the will's operations, namely, that of willing ends and willing means. We can construe this fact as a psychological fact of consistency between these two operations of the will.<sup>122</sup> But even when this is done, there are still two challenges to face. First, reference to this fact re-invites the question: why should I care about this fact? And now we have no literal practical necessity involved either: failing to act on the principle does not put an end to our agency. Second, even if this question receives an answer, the Model still needn't follow.<sup>123</sup> For now we have based justification on a fact that opens the way for other practical principles that are also grounded in facts about

---

<sup>122</sup> Alternatively, we can claim that when this fact is present, another non-natural, irreducible fact also occurs: that the act is rational. See FitzPatrick (ms). But, recall, this is not an option available to us in defending the Model.

<sup>123</sup> Several answers are possible, but two especially stand out (for the critical assessment of Korsgaard's own responses see FitzPatrick (2005), pp. 675-7). FitzPatrick (ms) appears to accept the idea that reference to facts about the operations of the will stops the regress. The driving thought here, I believe, is that these facts are internal to the will, not external. Yet, I think it is not obvious that such facts do indeed stop the regress; further philosophical work would be needed to show that this is so: FitzPatrick, however, provides no such thing. In this regard I find the approach he takes in FitzPatrick (2005), pp. 682-691 more acceptable. The basic idea comes from Korsgaard's formulation of the practical problem-solving framework: that the given principle should be the only *or* the best solution to the agent's practical problem. We can say that even in the absence of a workable literal practical necessity account the instrumental principle is still the best solution to the problem of acting. On Korsgaard's constructivist model it is so by virtue of the fact that it is a principle the agent is committed to in willing an end; on the realist model it is so by virtue of the fact that it is a true principle grounded in facts about commitment. Of course, this brings with it the further task of showing why the principle would be true. Hence, in addition to the second problem mentioned in the text, the defender of the Model has this further task to meet.

consistency.<sup>124</sup> Again, the burden of proof is on the advocate of the Model: he must show that this is not the case.

We have one attempt left. It again comes from Railton. Just like Korsgaard, he argues that the instrumental principle is constitutive of agency. And for the same reasons he also thinks that we must understand constitution in a weaker sense: it does not require that one actually acts on the principle. What it requires, and this is the first difference between him and Korsgaard, is that the one has some *disposition* to follow the principle.<sup>125</sup> (Railton 2003b, 307-13) But the really important difference is that in Railton's view this kind of defense cannot provide justification. His problem is even more radical than mine above. It is not only that we can ask why we should not resist our disposition, but also that in certain situations we can genuinely wonder why we should not eradicate it by putting an end to our agency. (Ibid. 313-5) For example, being a patient with an incurable, painful and costly disease, one can reasonably question the point of staying alive.<sup>126</sup> Or, to take an example from Parfit, when one is attacked by a mob who seeks revenge on his family because he testified against them, he can be tempted by the idea to knock himself senseless temporarily or even permanently. Constitutive arguments, Railton says, have no resources to answer either of these challenges.

At the same time, however, he thinks that this problem and, presumably, mine problem above too can be remedied. For, he says, the agent will ask these questions in such a way that betrays deference to the instrumental principle again. What he will wonder about is whether

---

<sup>124</sup> Indeed, this is what Korsgaard claims within the constraints of constructivism: she thinks that categorical imperatives also follow due to the commitments taken up in the course of our exercise of agency. See Korsgaard (1996b), pp. 120-3. For a critique of her argument see FitzPatrick (2005), pp. 677-81.

<sup>125</sup> This is a restatement of Railton's view in which I follow Wedgwood (2005), pp. 465.

<sup>126</sup> Note that this kind of suicide is different from the suicide I referred to when dealing with Wedgwood's proposal. There we were asked to imagine an agent who, in a hands-up fashion, announces that he is not going follow any principle he is presented with. He is basically saying that he sees no point in acting for a reason; hence he cannot be offered a reason. In contrast with this, the present problem is exactly that the agent is looking for a reason to remain an agent but he sees none. I think Korsgaard is referring to the same sort of difference in Korsgaard (1996b), pp. 243, which makes it even more interesting why she doesn't consider Railton's problem.

crossing the line between agency and non-agency (or whether to act on his disposition or to follow through on his commitment) is the best or only way of getting what he most wants from life. (Ibid. 315; also Rosati 2003, 522) But it is not obvious that the agent *must* make reference to the instrumental principle in order to raise his challenge. There are two options. There may be further principles that are also constitutive of agency and the agent invokes *these* principles in his challenge. We would then have a set of principles a member of which we have to invoke in order to raise a challenge about another member. But, crucially, we would not get the Model since the instrumental principle would no longer be fundamental. More tentatively, it is at least conceivable that the agent makes reference to none of these norms when posing a challenge. (Wedgwood 2005, 466) He may just be genuinely puzzled about how to make up his mind about what to do, and unpersuaded by the proposals philosophers have offered so far.

A further problem looms if we accept Railton's proposal. If someone is puzzled about the instrumental principle, then repeated reference to the principle hardly helps him out. This is Korsgaard's problem again. To take Railton's example, the agent would ask: why should I do what gets me what I most want in life? But Railton thinks that it is exactly the possibility of circularity that shows why the agent cannot ask this question. To explain, he brings an analogy with Lewis Carroll's argument concerning *modus ponens*. (Railton *ibid.* 316-7) Carroll has shown that if one doesn't reason in accordance with *modus ponens* when forming beliefs, then adding *modus ponens* as a premise in his reasoning doesn't help. For to effect a conclusion from the new premises, the agent would have to use *modus ponens*, and this is exactly what he doesn't do. As Railton rightly points out, there is a clear parallel between this argument and the present challenge. If one doesn't reason in accordance with the instrumental principle, then adding the principle as an end or a means in his reasoning doesn't help. For to effect a conclusion from the new premises, he would have to use the instrumental principle, and this is exactly what he



doesn't do. Hence, Railton concludes, the instrumental principle cannot be just another premise in the agent's practical reasoning.

This is puzzling. For the point of Korsgaard's charge was exactly that on a realist construal the instrumental principle will *be* a premise in the agent's reasoning *because of* the push of justification. In fact, she explicitly uses the analogy with Carroll's paradox to illustrate her problem. (Korsgaard 1997, 239-41; ms a, Lecture 2, 16) Hence, from Korsgaard's point of view, Railton merely restates the problem without offering a solution to it. Of course, the analogy with Carroll's problem does suggest possible ways of solving it. Korsgaard, for instance, takes it to show that just as the agent's theoretical reasoning would be 'a mere heap of premises' were she refuse to employ modus ponens, his practical reasoning would also fall apart without the use of the instrumental principle. This then leads directly to her second variation on the idea of literal practical necessity: the claim that we need principles of practical reason in order to unify our agency. (Korsgaard ms a, Lecture 2, 16-7; cf. Blackburn 1995, 709-10) And, at least we cannot rule this out, there can be other lessons one might draw from a parallel with Carroll's paradox.<sup>127</sup> But Railton does no such thing and in the absence of such an attempt it is hard to see what difference his suggestion makes.

---

<sup>127</sup> The best such idea comes from James Dreier. His problem is this. If one believes that a rule requires him to act but is not moved to act, then what is missing must be a desire. But this means that the instrumental principle cannot be just another rule that we need a desire to comply with. And the reason for this is just the parallel with Carroll's paradox: if we require such a desire, then we will never get to the end of the questions. This is significant. For all other rules are applied in action as a result of two things: a desire that has them as its object plus the instrumental principle. Hence, were we to question the principle, thus get entangled in a Carroll-type of regress, our practical reasoning would collapse. Not in the sense as Korsgaard describes but in the sense that nothing would *count as a reason* for us. This is the component missing from Railton's argument. Once it is in place Dreier can conclude that the instrumental principle must have a special status among principles of practical reason. See Dreier (2001), pp. 38-45. But Dreier's argument is premised on two things. First, he takes that a rule can only be applied in action via the instrumental principle. This is just the point made by Korsgaard: that the instrumental principle is the principle of application itself. And, as we saw, there are philosophers who deny this assumption. Second, as Dreier also admits, there may other principles that have the same status as the instrumental principle. They can function both as a supplement as well as an alternative to the instrumental principle.

## APPENDIX III: FAILED ATTEMPTS TO REFUTE THE MODEL

### *I. Five ways to reject the Model*

I mentioned in Chapter III that there are alternative ways to reject the Model by questioning its picture of the reason-relation. The attempts fall into five categories depending on their rationale of rejection. These are the following:

- (1) There are counter-examples to the Model;
- (2) The Model is internally inconsistent;
- (3) The Model is not normative;
- (4) Desires are non-normative mental states;
- (5) The Model has a false picture of practical deliberation;

I also said that none of these attempts are successful. In this appendix my aim is to show why this is so. This is good for us for two reasons. On the one hand, it helps me to explain why we should adopt the idea of reason-based desires as our preferred method of refuting the Model. On the other hand, it shows us important features of the Model thereby giving us a better understanding of it. Since much of what I will say has already appeared in press, I will be brief both in my exposition as well as in my discussion of the attempts.

## ***II. Why the attempts fail***

### **A.3.2.1. First argument**

Advocates of the first approach claim that there are certain cases when it is not the desire that provides us with reasons but something else. Michael Woods has argued that moral and aesthetic considerations provide us with reasons that are not based on our desires, and John Searle has made the same claim with regard to commitments in general. (Woods 1972, 196-7; Searle 2001, 170) However, even if it is successful, this approach will only show that the Model does not hold over the whole range of practical reasons. The result will be a hybrid account on which certain reasons are provided by desires, while others are not. Although it might turn out in the end that this is all we can say about reasons, it would nevertheless be better to begin with a more ambitious refutation and then see how far we can get with it. Moreover, the particular form the objection takes in the hands of Woods and Searle is also problematic. They at no point question the *philosophical picture* underlying the Model - its account of the reason-relation – by giving an alternative to it or showing it to be inconsistent or by some other way. They simply insist that they are right. Consequently, their opponents have a simple response open to them. They can point out that the reasons they refer to are in fact based on desires at least if we employ a broad enough theory of desire. And the result of this debate is an unproductive standoff, not a refutation of the Model.

### **A.3.2.2. Second argument**

The second line of thought appears in the work of Warren Quinn and John Broome. (Quinn 1993, 237; Broome 1997, 136) Their argument is the following. Even an advocate of the Model admits (1) that we have no reason to  $\varphi$  just because we desire to  $\varphi$  since the desire can simply be silly, inconsiderate or alien; he only claims (2) that we have reason to do whatever subserves our desire

to  $\varphi$ ; but (3) we can think of  $\varphi$ -ing as the limiting case of ‘whatever subserves  $\varphi$ -ing’; hence (4) if he still wants to hold (1) he must reject (2): the Model is inconsistent. In Chapter III I briefly noted what my problem is with this argument. Here it is in more detail.

It is clear that the second and third versions of the Model are not subject to the Broome/Quinn argument since they do not accept premise (2). More precisely, they give a role to the premise, which basically articulates the instrumental principle, but restrict its scope. The principle applies only to those desires that are privileged in some sense. Hence the Model is not inconsistent since, though it endorses premise (1), it claims that certain acts are not supported by reasons even if they subserve a desire of the agent. As to the first reading of the Model, here the argument may look more effective in as much this version holds premise (2) in its unrestricted form. This granted there is still no reason to suppose that an advocate of this version of the Model would not be willing to accept premise (1). On the contrary, he can endorse the premise exactly because he can take it to follow from premise (2). This is just what Dancy’s Advice Point tells us. It turns the relation of consistency that stands between desire and action into a claim about reasons. Consequently, here it does not matter what sort of desire we are operating with; no matter how silly, inconsiderate or alien the desire is, if the consistency relation holds, all this does not count.<sup>128</sup>

### **A.3.2.3. Third argument**

The third objection comes from Christine Korsgaard. In her argument against ‘empiricists’, she argues that the instrumental principle (IP) loses normativity if it is taken to be the only principle of practical reason. (Korsgaard 1997, 215, 223; 1996b, 163-4; ms a, Lecture 2, 9-11) Advocates of the Model do just this: they insist that the agent’s ends are not open to rational criticism. This

---

<sup>128</sup> This is, I think, the same point that Dancy in his rejection of the argument makes. See Dancy (2000b), pp. 34.  
Attila Tanyi, THESIS, Page 220

leads them to identify a person's end "with what he *wants most*, and the criterion of what the person wants most appears to be what he actually *does*. The person's ends are taken to be revealed in his conduct." (Korsgaard 1997, 230) As a result, whenever one does something he is taking the means to the end he is *going* to pursue. It doesn't matter if it is not the end that the agent *says* he is pursuing; idleness, shyness or terror, can all be ends even though the agent says that they are only factors that hinder the pursuit of his end. (Ibid. 223) The result is that there is no way to violate the IP. Although the agent can desire non-existent things or act on the basis of false causal judgments, these are more like innocent mistakes than instances of irrationality, But a principle that cannot be violated cannot be normative either. The Model is refuted.

The crux of Korsgaard's argument is her claim that the advocate of the Model is committed to the claim that the agent's ends are revealed in his conduct and that therefore no selection among them is possible.<sup>129</sup> This is a controversial claim. First, Korsgaard herself admits that certain acts are mistaken though done for a revealed end: those that promote desires for non-existent things and those that are carried out on the basis of false causal judgments. She thinks these acts don't count because they are innocent mistakes and not instances of irrationality. At this point, however, we can again invoke the distinction between reason and rationality and the three corresponding versions of the Model. Reasons for action, recall, are not dependent upon the agent's beliefs, only the rationality of his action is. Consequently, Korsgaard's charge invites the following response. If she accepts the distinction, she must also accept that though the invoked acts do not count as irrational, they are still contrary to reason. And since the second version of the Model takes reason to be the primitive, the IP can also be violated. In this case Korsgaard's

---

<sup>129</sup> Since Korsgaard bases her argument on the interpretation of Hume, we have to separate two issues: the truth of her interpretation and the conclusions she derives from it. As to the first, she may even be right. Compare, for instance, Rawls (2000), pp. 49. But her paper is not just an essay on Hume; she also targets those who claim to follow Hume in regarding the IP as the only principle of practical reason. And whether she is right in her judgment about this second group is an open question.

argument is flawed. (Sobel 2001a, 489) Alternatively, she can insist that the other versions of the Model are right in which case, as things stand, her argument follows. But, of course, this needs to be argued for, something Korsgaard does not do.

Second, the claim that the Model has no resources for selecting among the agent's ends can itself be questioned. We saw that advocates of the second and third version of the Model make a distinction between desires that ground reasons and those that do not, and take the IP to apply only to these desires. Korsgaard is aware of this response, but thinks that any such distinction goes beyond the Model. Such a move, she argues, forces advocates of the Model to say that agents ought to pursue these desires over other desires; this is the only way to accord these privileged desires some normative force. (Korsgaard 1997, 230) Korsgaard's complaint seems to be this. If the IP only transmits reasons to the means, then these reasons must come from somewhere. And this 'somewhere', her reasoning suggests, can only be a non-instrumental principle to the effect that one ought to have certain ends, and this, she rightly claims, is inconsistent with the Model. (Hubin 2001, 457-62)

But from Chapter I we know the answer to this suggestion. On the second version, which does indeed have this picture of the IP, desires are taken to be reason-grounding brute facts. The idea comes from Donald Hubin. Pursuing an analogy with Hans Kelsen's legal positivism, he points out that on this version of the Model desires sit at the end of the normative chain: the normativity of actions *just is* their being properly related to certain of the agent's desires. (Ibid. 463-8) Hence there is no need for a separate principle to confer normativity on these desires; we just have to pay attention not to make substantive normative judgments in giving an account of the criterion of privilege employed in our theory. On the third version of the Model, the situation is somewhat different, but here too we don't have to accept Korsgaard's verdict. The account of rationality we build into our account of condition *C* need not go beyond instrumental rationality,

while it can invoke any kind of standard of theoretical rationality. And, again, if we manage to give a naturalist account of rationality, we avoid trouble entirely.

The previous responses save the second and third version of the Model from Korsgaard's objection. Yet, they may leave the objection in place as to the first version, which does not select among the agent's desires. Even so, Korsgaard's conclusion still does not follow. This is because her understanding of the IP is mistaken. (Dancy 2000b, 43-8) On her reading, the principle requires the agent to take efficient means to his ends where these ends are taken to be identical with what he *most* wants. But this is not what the IP tells us to do. What the principle requires from us is to take efficient means to our *individual* ends. Korsgaard's moving thought here, probably, is another objection to the Model, namely that it doesn't include a principle for adjudicating among conflicting ends. In itself this is a legitimate objection to make; but Korsgaard links it to the IP – and *this* is not legitimate. The IP has nothing to say about what one wants most; it must be another principle, if any, that does this work (more about this in a minute). Although there may be problems with the inclusion of such a principle, this doesn't touch upon the real issue, which is that the IP is not a principle of adjudication.

The IP, then, applies to individual ends. It can do so in three ways. First, it may be that people should take the means to each of their ends. But this is absurd. For it to be different from the other two interpretations it would have to mean that one should, some time during one's life, take the means to each of one's ends. Clearly, however, we can *have* ends that we will never pursue in the sense of taking steps to achieve them. Yet, they are clearly our ends. Second, the principle may imply that we should take the means to an end if all the other ends are fulfilled and the opportunity presents itself. This is indeed a principle that we cannot violate; if we do so, then we must admit that we don't in fact have the end. However, recall my definition of the IP in Chapter I, there is a third alternative. It says that for each end, if we do in fact take steps to

achieve that end, we should adopt the most effective means available. This principle *is* floatable. It can be the case that I do not take the acknowledged most effective means to *one* of my ends. Of course, at the same time I might take the most effective means to another end of mine, and then we have to decide if I am rational or not. But this only shows that the IP must be accompanied by some principle of adjudication. Otherwise the IP, standing alone unaccompanied by other principles, is floatable and, in this minimal sense, normative.<sup>130</sup>

#### **A.3.2.4. Fourth argument**

There is, however, another line of thought in Korsgaard's writings, the fourth attempt on our list, which promises a different objection to the Model.<sup>131</sup> The idea is to contrast desires with volitions by applying an analogy with beliefs and perceptions. "To believe something", Korsgaard claims, "is not to be in a certain mental state, but to make a certain commitment." (Korsgaard 1997, 248) But beliefs do not come from nothing. They are the results of the mind's reflection on perceptual states on the basis of rational principles. These perceptual states do not involve normative commitments: they are only items given to the mind to be worked upon by rational principles. (Korsgaard 1998, 12) And, Korsgaard claims, the situation is the same with desires and volitions. (Korsgaard 1997, 248n) Desires are inclinations: they are the agent's awareness that something

---

<sup>130</sup> I take this third point to be consistent with Lavin (2004)'s picture of what he calls the 'error constraint', i.e. the idea that the possibility of error is a necessary condition of normativity. According to Lavin there are three interpretations of the constraint. On the *logical* reading an agent is subject to a principle if there is some kind of action such that if the agent did it she would thereby violate the principle. On the *imperative* interpretation we add facts about the agent: besides the existence of principle-violating actions, it must also be possible for the agent to carry out these actions. Finally, on the *strong* reading of the imperative interpretation, a further constraint is added. It must not only be possible for the agent to violate the principle, but she must do it because of genuine practical error such as terror or idleness. See Lavin (2004), pp. 426-7, 436. Lavin argues that Korsgaard has in mind the strong imperative reading: this is why she says what she says. Put in this context, my point is that we can agree with him on this and still claim that the IP does not violate the error constraint.

<sup>131</sup> There is considerable ambiguity about Korsgaard's position. In other works she appears to deny the view she endorses here. See Korsgaard (1996a), pp. 289-291; (1996b), pp. 239. The idea, however, clearly appears in Barbara Herman's work. See Herman (1993), pp. 198-201 and (1996), pp. 44. And we can also find it in Sidgwick (1907), pp. 37 and Searle (2001), pp. 173, 176.



attracts him or tempts him to action. Since these ‘incentives’ are entirely devoid of rational commitment, desires are like perceptual states: they are grounded in our sensuous nature in the face of which we are passive. (Korsgaard 1998, 51-2) Volitions, in contrast, are like beliefs in that they are the results of the agent’s decision, on the basis of rational principles, to act upon a certain desire. Consequently, since desires do not involve normative commitments; they are not reasons. (Korsgaard 1997, 234)

The challenge for Korsgaard is to come up with a plausible candidate for the state that she now calls desire. But it is difficult to see how she could do this given her own Wittgensteinian commitments.<sup>132</sup> (Hurley 2001, 8-10) There are three candidates. One compares desires to sensations understood as states that are prior both to conceptualization as well as to reason. However, postulating the existence of such a state runs into all the objections against the private given. More importantly, it is also against Korsgaard’s endorsement of Wittgenstein’s idea that, as she puts it, “you cannot peer inwardly at an essentially private and incommunicable sensation.” (Korsgaard 1996b, 137) The second candidate reduces a bit the extremity of the first by taking the requisite state to be prior only to reason not to conceptualization. But this move would be inconsistent with another of Korsgaard’s Wittgensteinian commitment, namely the idea that to employ concepts just is to take up a position in the space of reasons. (Ibid.) Finally, desires might be taken to be bare dispositions such as the desires we considered in Chapter I and Chapter III under the heading of urges. However, while these dispositions do exist and lack normative commitment, it is hard to believe that our ordinary desires are the results of the mind’s rational

---

<sup>132</sup> For completeness’ sake we should note that Korsgaard makes two further suggestions that I do not consider in the text. One is that desires do not involve normative commitment because they are states, whereas commitments are not. The other is that desires are not arrived at after reflection, hence they cannot involve commitments. For this interpretation see Hurley (2001), pp. 5. I do not deal with these ideas because I think it is easy to reject them. Clearly, one can be in a state while also being committed. Hurley gives the example of matrimony, which is a state that essentially involves a commitment. See Hurley (2001), pp. 6. And from Chapter I (and elsewhere in the dissertation) we know that lack of reflection does not imply lack of normative commitment. In fact, even if it does this still doesn’t mean that the desire, *while* the agent has it, involves such commitments.

activity on such dispositions. These states, or at least some of them, are more like failed *attempts* to become states that involve normative commitment than states prior to them. Or not even that, as I suggested in Chapter I.

#### **A.3.2.5. Fifth argument**

The fifth objection claims that the Model gives us a distorted picture of practical reasoning. The charge comes in two stages. First, it is often pointed out that agents can and do deliberate about what they should desire *without* relying on other desires of theirs. They ask not what they should desire given what they already desire, but what they should desire *tout court*. (Korsgaard 1996b, 241; Nagel 1997, 110; Raz 1999a, 50n) Second, even if the latter charge turns out to be unfounded, we still have the already mentioned problem that the Model cannot tell us what to do when the agent's desires conflict. More precisely, the Model directs the agent to sit back and watch which of his desires come out as motivationally strongest. But, the claim goes, this is not reasoning but acting. (Davidson 1980b, 33; Schueler 1995, 100; Raz 1999a, 51) We need a principle of adjudication, it is claimed, but the Model has none.

Advocates of the Model disagree. They think that both charges spring from two fundamental mistakes about the process of practical deliberation. (Pettit and Smith 1990; Blackburn 1998, 250-261) The opposition thinks that because people deliberate about their desires, therefore there must be a standpoint independent of those desires. But what happens in deliberation is that the agent is engaged with the *world*: he deliberates about the *object* of the desire through the lenses of his other desires. Desires are not in the *foreground* of deliberation: we do not deliberate *about* desires; instead, desires are in the *background* of deliberation: we deliberate *from* our desires. The opposition also makes a mistake similar to when one conflates the absence of an experience of causation with the experience of the absence of causation. They

think that because people in deliberation can be aware not only of facts about the world but also of facts about their conative nature, they must be standing at some vantage-point somewhere above these facts. But this is an illusion. What this experience proves is not the existence of such a vantage-point but the existence of a *blindspot* or aspect of ourselves that cannot be seen, by ourselves, while we deliberate.

Once we make explicit these mistakes, we can answer both charges. People cannot deliberate about their ends unconditionally because to do so they would need a standpoint independent of those ends and there is no such standpoint (or at least there is as much reason to believe this, than to believe the opposite). Nor is it a problem for advocates of the Model to admit that conflicts among desires are resolved on the basis of their motivational strength. Since people deliberate about aspects of the world through their desires, they do make choices and decisions. Their role is not reduced to that of the passive bystander: they are *actively* engaged with the world, though, admittedly, not in the way the opposition requires them to be. Moreover, we should note that this objection is unfounded for another reason, which is unrelated to the above understanding of deliberation. (Sobel 2001a, 486) We saw that the second and third versions of the Model give normative priority to certain desires. This means that when these desires conflict with other non-privileged desires the conflict won't be decided merely on the ground of motivational strength but according to the criterion of privilege employed.

## BIBLIOGRAPHY

Anscombe, G.E.M. (1957), *Intention*, Oxford: Blackwells

Audi, Robert (1986), 'Acting for Reasons' in *Philosophical Review* 95:4 (1986), pp. 511-546

Ayer, A. J. (1936), *Language, Truth and Logic*, London: Victor Gollancz, 2<sup>nd</sup> edition

- (1954), 'Freedom and Necessity' in Gary Watson (ed.): *Free Will*, Oxford: The University Press, 1982, pp. 15-23

Baier, Kurt (1958), *The Moral Point of View: A Rational Basis of Ethics*, Ithaca, N.Y.: Cornell University Press

- (1978), 'The Social Source of Reason (APA presidential address)' in *Proceedings and Addresses of the American Philosophical Association*, 52 (1978), pp. 707-733

Bedke, Matt (ms), 'Two Internalisms Concerning Moral Motivation'

Bennett, Jonathan, (1995), *The Act Itself*, Oxford: Clarendon Press

Berridge, Kent C. (2004), 'Motivation Concepts in Behavioral Neuroscience', in *Physiology & Behavior* 81 (2), pp. 179-209

Bittner, Rüdiger (2001), *Doing Things for Reasons*, Oxford: The University Press

Blackburn, Simon (1995), 'Practical Tortoise Raising' in *Mind*, 104: 416 (1995), pp. 695-711

- (1998), *Ruling Passions: A Theory of Practical Reasoning*, Oxford: Clarendon Press

Bognár, Greg (ms), 'Welfare Judgments and Risk'

Bond, E.J. (1983), *Reason and Value*, Cambridge: The University Press

Boyd, Richard N. (1988), 'How to Be a Moral Realist' in Geoffrey Sayre-McCord (ed.), *Essays in Moral Realism*, Ithaca: Cornell University Press, pp. 49-78

Brandt, Richard (1979), *A Theory of the Right and the Good*, Oxford: The University Press

Bratman, Michael (1987), *Intention, Plans, and Practical Reason*, Cambridge, Mass.: Harvard University Press

Brink, David O. (1989), *Moral Realism and the Foundations of Ethics*, Cambridge: The University Press

- (1997), 'Moral Motivation' in. *Ethics* 108:1 (1997), pp. 4-32

- (2001), 'Realism, Naturalism, and Moral Semantics' in. *Social Philosophy and Policy* 18:3 (2001), pp. 154-76

Broad, C. D. (1930), *Five Types of Ethical Theory*, London: Routledge and Kegan Paul

Broome, John (1997), 'Reason and Motivation' in. *Proceedings of the Aristotelian Society*, Supplementary Volume, 1997, pp. 130-146

- (1999), 'Normative Requirements' in. *Ratio* 12:4 (1999), pp. 398-419

- (2004), 'Reasons' in. R. Jay Wallace, Michael Smith, Samuel Scheffler, and Philip Pettit (eds.), *Reason and Value: Essays on the Moral Philosophy of Joseph Raz*, Oxford: The University Press, pp. 28-55

- (2005), 'Does Rationality Give Us Reasons?' in. *Philosophical Issues*, Normativity, 15 (2005), pp. 321-337

- (ms), 'Wide or Narrow Scope? Reply to Kolodny'

Chang, Ruth (2002), 'The Possibility of Parity' in. *Ethics* 112:4 (2002), pp. 659-88

- (2004), 'Can Desires Provide Reasons for Action?' in. R. Jay Wallace, Michael Smith, Samuel Scheffler, Philip Pettit (eds.), *Reasons and Value: Essays on the Moral Philosophy of Joseph Raz*, Oxford: The University Press, pp. 56-90

Chisholm, Roderick M. (1966), 'Freedom in Action' in. Keith Lehrer (ed.), *Freedom and Determinism*, Random House, New York, pp. 11-45

Christiano, Thomas (1992), 'Sidgwick on Desire, Pleasure, and the Good' in. *Essays on Henry Sidgwick*, Bart Schultz, ed., Cambridge: The University Press, pp. 261-278

Chuard, Philippe and Friedrich, Daniel (ms), 'The Vagaries of Defeat'

Cohon, Rachel (1988), 'Hume and Humeanism in Ethics' in. *Pacific Philosophical Quarterly*, 69 (1988), pp. 99-116

Collins, John (1988), 'Belief, Desire, and Revision' in. *Mind* 97:387 (1988), pp. 333-342

Copp, David and Sobel, David (2001), 'Against Direction of Fit Accounts of Belief and Desire' in. *Analysis* 61:1 (2001), pp. 44-53

- (2002), 'Desires, Motives, and Reasons: Scanlon's Rationalistic Moral Psychology' in. *Social Theory and Practice* 28: 2 (2002), pp. 243-76

Copp, David (1990), 'Explanation and Justification in Ethics' in. *Ethics* 100:2 (1990), pp. 237-258

- (1995), *Morality, Normativity, and Society*, New York: Oxford University Press

- (2000), 'Milk, Honey, and the Good Life on Moral Twin Earth' in. *Synthese* 124:1 (2000), pp. 113-137

- (2001), 'Realist-Expressivism: A Neglected Option For Moral Realism' in. *Social Philosophy and Policy*, 18: 2 (2001), pp. 1-43

- (2003), 'Why Naturalism?' in. *Ethical Theory and Moral Practice* 6:2 (2003), pp. 179-200

Crisp, Roger (2000), 'Particularizing Particularism' in. Brad Hooker, Margaret Little (eds.), *Moral Particularism*, Oxford: The University Press, pp. 23-48

Cuneo, Terence (2002), 'Reconciling Realism With Humeanism' in. *Australasian Journal of Philosophy* 80:4 (2002), pp. 465-486

Dancy, Jonathan (1993), *Moral Reasons*, Oxford: Blackwell

- (1999), 'Motivation, Dispositions and Aims' in. *Theoria* 99: 2-3 (1999), pp. 212-224
- (2000a), 'Introduction' in. Dancy (ed.): *Normativity*, Oxford: Blackwell
- (2000b), *Practical Reality*, Oxford: The University Press
- (2003a), 'What Do Reasons Do?' in. *Southern Journal of Philosophy*, Supplementary Volume 61 (2003), pp. 95-113
- (2003b), 'Replies' in. *Philosophy and Phenomenological Research*, 67:2 (2003), pp. 468-490
- (2004), *Ethics Without Principles*, Oxford: The University Press
- (forthcoming a), 'Defending the Right', *Journal of Moral Philosophy*
- (forthcoming b), 'Reasons and Rationality' in. John Skorupski, Simon Robertson, and Jens Timmerman (ed.), *Spheres of Reason*

Daniels, Norman (1983), 'Can Cognitive Psychotherapy Reconcile Reason and Desire?' in. *Ethics* 93:4 (1983), pp. 772-785

Darwall, Stephen, Gibbard, Allan and Railton, Peter (1997), 'Toward *Fin de siècle* Ethics: Some Trends' in. S. Darwall, P. Railton, A. Gibbard (eds.): *Moral Discourse and Practice, Some Philosophical Approaches*, Oxford: The University Press, pp. 3-47

Darwall, Stephen L (1983), *Impartial Morality*, Ithaca, N.Y.: Cornell University Press

- (1990), 'Autonomist Internalism and the Justification of Morals' in. *Noûs* 24:2 (1990), pp. 257-68
- (1992), 'Internalism and Agency' in. *Philosophical Perspectives, Ethics*, 6 (1992), pp. 155-174
- (1997), 'Reasons, Motives and the Demands of Morality: An Introduction' in. S. Darwall, P. Railton, A. Gibbard (eds.): *Moral Discourse and Practice, Some Philosophical Approaches*, Oxford: The University Press, pp. 305-313

- (2001), 'Because I Want It' in. *Social Philosophy and Policy*, 18:2 (2001), pp. 129-154
- (2003), 'Desires, Reasons, and Causes' in. *Philosophy and Phenomenological Research*, 67:2 (2003), pp. 437-443

Davis, Lawrence H. (1979), *Theory of Action*, Englewood Cliffs, New Jersey: Prentice Hall

Davidson, Donald (1980a), 'Actions, Reasons, and Causes' in. *Essays on Actions and Events*, Oxford: The University Press, pp. 3-21

- (1980b), 'How Is Weakness of the Will Possible' in. in. *Essays on Actions and Events*, pp. 21-43
- (1980c), 'Freedom to Act' in. *Essays on Actions and Events*, pp. 63-83
- (1980d) 'Intending' in. *Essays on Actions and Events*, pp. 83-103
- (1980e), 'Mental Events' in. *Essays on Actions and Events*, pp. 207-227

Dreier, James (1990), 'Internalism and Speaker Relativism' in. *Ethics* 101:1 (1990), pp. 6-26

- (2001), 'Humean Doubts about Categorical Imperatives' in. Millgram, Elijah (ed.): *Varieties of Practical Reasoning*, Cambridge, Mass.: Bradford Book, MIT Press, pp. 27-47
- (2002), 'The Expressivist Circle: Invoking Norms in the Explanation of Normative Judgment' in. *Philosophy and Phenomenological Research*, 65:1 (2002), pp. 137-143

Elstein, Daniel (ms), 'The Open Question Argument and Cornell Realism'

Enoch, David (2005), 'Why Idealize?' in. *Ethics* 115:4 (2005), pp. 759-787

Firth, Roderick (1952), 'Ethical Absolutism and the Ideal Observer' in. *Philosophy and Phenomenological Research* 12:3 (1952), pp. 317-345

FitzPatrick, William J. (2004), 'Reasons, Value, and Particular Agents: Normative Relevance without Motivational Internalism' in. *Mind* 113:451 (2004), pp. 285-318



- (2005), 'The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity' in. *Ethics* 115:4 (2005), pp. 651-691
  - (ms), 'Normative Realism, Kantian Constructivism and the Instrumental Principle'
- Frankfurt, Harry (1971), 'Freedom of the Will and the Concept of the Person' in. Gary Watson (ed.): *Free Will*, Oxford: The University Press, pp. 81-95
- Gampel, Eric H. (1997), 'Ethics, Reference, and Natural Kinds' in. *Philosophical Papers* 26:2 (1997), pp. 147-163
- Garrard, Eve and McNaughton, David (1998), 'Mapping Moral Motivation' in. *Ethical Theory and Moral Practice* 1:1 (1998), pp. 45-59
- Gert, Bernard (1998), *Morality: Its Nature and Justification*, New York: Oxford University Press
- Gert, Joshua (2003), 'Requiring and Justifying: Two Dimensions of Normative Strength' in. *Erkenntnis* 59:1 (2003), pp. 5-36
- Gibbard, Allan (1990), *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Cambridge, Mass.: Harvard University Press
- (2003), *Thinking How to Live*, Cambridge, Mass.: Harvard University Press
- Greenspan, Patricia R. (1975), 'Conditional Oughts and Hypothetical Imperatives' in. *The Journal of Philosophy* 72:10 (1975), pp. 259-276
- Grice, Paul (2001), *Aspects of Reason*, Richard Wagner (ed.), Oxford: Clarendon Press
- Gulliver P.H. (1973), 'Negotiations as a Mode of Dispute Settlement: Towards a General Model' in. *Law and Society Review* 7:4 (1973), pp. 667-692
- Hampton, Jean (1998), *The Authority of Reason*, New York: Cambridge University Press
- Haidt, Jonathan (2001), 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment' in. *Psychological Review* 108:4 (2001), pp. 814-834
- Hare, R. M. (1952), *The Language of Morals*, Oxford: Clarendon Press

- (1963), *Freedom and Reason*, Oxford: Clarendon Press
  - (1971), 'Wanting: Some Pitfalls' in. Robert Binkley, Richard Brunaugh, and Ausonio Marras (eds.): *Agent, Action, and Reason*, Toronto: The University Press
  - (1981), *Moral Thinking*, Oxford: Clarendon Press
- Harrison, Jonathan (1976), *Hume's Moral Epistemology*, Oxford: Clarendon Press
- Hastie, R. and Pennington, N. (1988), 'Explanation-Based Decision-Making: Effects of Memory Structure on Judgment' in. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14:3 (1988), pp. 521-533
- Hatzimoysis, Anthony (2002), 'Analytical Descriptivism Revisited' in. *Ratio* 15:1 (2002), pp. 10-22
- Heuer, Ulrike (2004), 'Reasons for Actions and Desires' in. *Philosophical Studies* 121:1 (2004), pp. 43-63
- Hill, Thomas E. (1992), 'The Hypothetical Imperative' in. *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca, NJ: Cornell University Press, pp. 17-37
- Jacobson-Horowitz, Hilla (2006), 'Motivational Cognitivism and the Argument From Direction of Fit' in. *Philosophical Studies* 127:3 (2006), pp. 561-580
- Herman, Barbara (1993), 'Agency, Attachment, and Difference' in. *The Practice of Moral Judgment*, Cambridge, Mass.: Harvard University Press, pp. 184-208
- (1996), 'Making Room for Character' in. *Aristotle, Kant and the Stoics*, ed. Whiting and Engstrom, Oxford: The University Press, pp. 36-60
- Horgan, Terence and Timmons, Mark (1990-1991), 'New Wave Moral Realism Meets Moral Twin Earth' in. *Journal of Philosophical Research* 16 (1990-1991), pp. 447-465
- (1992a), 'Troubles on Moral Twin Earth: Moral Queerness Revived' in. *Synthese* 92:2 (1992), pp. 221-260

- (1992b), 'Troubles for New Wave Moral Semantics: the 'Open Question Argument' Revived' in. *Philosophical Papers* 21 (1992), pp. 153-175

Hubin, Donald C. (1996), 'Hypothetical Motivation' in. *Noûs* 30:1 (1996), pp.31-54

- (1999), 'What's Special about Humeanism' in. *Noûs* 33:1 (1999), pp. 30-45

- (2001), 'The Groundless Normativity of Instrumental Rationality' in. *The Journal of Philosophy*, 98: 9 (2001), pp. 445-468

- (2003), 'Desires, Whims, and Values' in. *The Journal of Ethics* 7:3 (2003), pp. 315-335

Hurley, Paul (2001), 'A Kantian Rationale for Desire-Based Justification' in. *Philosophers' Imprint* 1:1 (2001), pp. 1-16,

Jackson, Frank (1998), *From Metaphysics to Ethics*, Oxford: The University Press

James, Henry (1881/1997), *The Portrait of a Lady*, London: Penguin Books

Johnson, Robert N. (2001), 'Commentary on Svavarsdóttir's 'Moral Cognitivism and Moral Motivation'', in. *Brown electronic Article Review Service*, Jamie Dreier and David Estlund, editors, URL = <http://www.brown.edu/Departments/Philosophy/bears/homepage.html>, Posted August 2001

Kagan, Shelly (1992), 'The Limits of Well-being' in. *The Good Life and the Human Good*, Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul, eds., Cambridge: The University Press

Katz, Leonard D. (1986), *Hedonism as Metaphysics of Mind and Value*, Ph.D. dissertation, Princeton University

- (2005), 'Opioid Bliss as the Felt Hedonic Core of Mammalian Prosociality – and of Consummatory Pleasure More Generally?' in. *Behavioral and Brain Sciences* 28:3 (2005), pp. 356

- (2006), 'Pleasure', *Stanford Encyclopedia of Philosophy*, (Summer 2006 Edition), Edward N. Zalta, ed., URL = <http://plato.stanford.edu/entries/pleasure/>

- (forthcoming), 'Hedonic Reasons as Ultimately Justifying and the Relevance of Neuroscience' in. Walter Sinnott-Armstrong, ed., *Moral Psychology*, Volume 3: *The Neuroscience of Morality*, Cambridge, Mass.: The MIT Press
  
- Kim, Jaegwon (1999), *Mind in a Physical World: an Essay on the Mind-Body Problem and Mental Causation*, Bradford Book, Cambridge, Mass.: MIT Press
  
- Kolodni, Niko (2005), 'Why Be Rational?' in. *Mind* 114:455 (2005), pp. 509-563
  
- Korsgaard, Christine M. (1986), 'Skepticism about Practical Reason' in. *Journal of Philosophy*, 83:1 (1986), pp. 5-25
  
- (1996a) 'The Reasons We Can Share: An Attack On the Distinction Between Agent-Neutral and Agent-Relative Values' in. *Creating the Kingdom of Ends*, Cambridge: The University Press, pp. 275-311
  
- (1996b), *The Sources of Normativity*, Cambridge: The University Press
  
- (1997), 'The Normativity of Instrumental Reason' in. Cullity, Gareth and Gaut, Berry (eds.): *Ethics and Practical Reason*, Oxford: The University Press, pp. 215-254
  
- (1998), 'Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind in. *Ethics* 109:1 (1998), pp. 49-66
  
- (2003), 'Realism and Constructivism in Twentieth Century Moral Philosophy' in. *The Journal of Philosophical Research*, Charlottesville, Virginia: The Philosophy Documentation Center, 2003, pp. 99-122
  
- (ms a), *Self-Constitution: Action, Identity and Integrity*, Locke Lectures, URL = <http://www.people.fas.harvard.edu/~korsgaard>
  
- (ms b), 'Normativity, Necessity, and the Synthetic A Priori: A Response to Derek Parfit', URL = <http://www.people.fas.harvard.edu/~korsgaard>

Lavin, Douglas (2004), 'Practical Reason and the Possibility of Error' in. *Ethics* 114:3 (2004), pp. 424-458

Lenman, James (1996), 'Belief, Desire and Motivation: An Essay in Quasi-Hydraulics' in. *American Philosophical Quarterly*, 33:3 (1996), pp. 291-301

- (1999), 'The Externalist and the Amoralist' in. *Philosophia* 27:3-4 (1999), pp. 441-457

Lewis, David (1986), *On the Plurality of Words*, Oxford: Blackwell

- (1988), 'Desire as Belief' in. *Mind* 97:387 (1988), pp. 323-32

- (1989), 'Dispositional Theories of Value' in. *Proceedings of the Aristotelian Society*, Supplementary, 63 (1989), pp. 129-132

Little, Margaret Olivia (1997), 'Virtue as Knowledge: Objections from the Philosophy of Mind' in. *Noûs* 31:1 (1997), pp. 57-79

Locke, Don (1974), 'Reasons, Wants and Causes' in. *American Philosophical Quarterly*, 11:3 (1974), pp. 169-179

- (1982), 'Beliefs, Desires and Reasons for Action' in. *American Philosophical Quarterly*, 19:3 (1982), pp. 241-249

Loeb, Don (1995), 'Full-Information Theories of Individual Good' in. *Social Theory and Practice* 21:1 (1995), pp. 1-30

Macdonald, Cynthia (1998), 'Tropes and Other Things' in. S. Laurence and C. Macdonald (eds.), *Contemporary Readings in the Foundations of Metaphysics*, Oxford: Basil Blackwell, pp. 329-350

McDowell, John (1978), 'Are Moral Requirements Hypothetical Imperatives?' in. *Proceedings of the Aristotelian Society*, Supplementary Volume, 53 (1979), pp. 13-29

- (1979), 'Virtue and Reason' in. *Monist* 62:3 (1979), pp. 331-350

- (1995), 'Might There Be External Reasons?' in J. E. J. Altham and H. Ross (eds.), *World, Mind and Ethics*, Cambridge: The University Press, pp. 68-85
- McKeever, Sean and Ridge, Michael (2005), 'What Does Holism Have To Do With Moral Particularism?' in *Ratio* 18:1 (2005), pp. 93-103
- McNaughton, David and Rawling, Piers (2003), 'Descriptivism, Normativity and the Metaphysics of Reasons' in *Proceedings of the Aristotelian Society*, Supplementary Volume 77 (2003), pp. 23-45
- McNaughton, David (1988), *Moral Vision: An Introduction to Ethics*, Oxford: Basil Blackwell
- Mead, George Herbert (1934), *Mind, Self, Society: From the Standpoint of a Social Behaviorist*, Charles w. Morris (ed.) Chicago, Illinois: The University of Chicago Press
- Mele, Alfred (2003), *Motivation and Agency*, New York: Oxford University Press
- Mill, John Stuart (1963), *Miscellaneous Writings*, edited by John M. Robson, Toronto: The University Press
- Millar, Allan (1991), *Reasons and Experience*, Oxford: Clarendon Press
- Miller, Richard (1979), 'Reason and Commitment in the Social Sciences' in *Philosophy and Public Affairs* 8:3 (1979), pp. 241-266
- (1985), 'Ways of Moral Learning' in *Philosophical Review* 94:4 (1985), pp. 507-56
- (1992), *Moral Differences: Truth, Justice and Conscience in a World of Conflict*, Princeton: The University Press
- Millgram, Elijah (1996), 'Williams' Argument Against External Reasons' in *Noûs* 30:2 (1996), pp. 197-220
- Milligan, D.E. (1974), 'Reasons as Explanations' in *Mind*, 83:330 (1974), pp. 180-193

- Millikan, Ruth Garret (1995), 'Pushmi-Pullyu Representation' in. James E. Tomberlin (ed.), *Philosophical Perspectives: Connectionism and Philosophical Psychology*, Atascadero, Cal.: Ridgeview Publishing Co., 9 (1995), pp. 185-200
- Moore, G. E. (1903), *Principia Ethica*, Cambridge: The University Press
- Nagel, Thomas (1970), *The Possibility of Altruism*, Princeton: The University Press
- (1986), *The View from Nowhere*, New York: Oxford University Press
  - (1997), *The Last Word*, New York: Oxford University Press
- Nisbett, R. E. And Wilson, T. D. (1977), 'Telling More Than We Can Do: Verbal Reports on Mental Processes' in. *Psychological Review*, 84:3 (1977), pp. 231-259
- Noggle, Robert (1999), 'Integrity, Self, and the Desire-Based Accounts of the Good' in. *Philosophical Studies*, 96:3 (1999), pp. 301-328
- Nowell-Smith, Patrick (1954), *Ethics*, London: Penguin
- Parfit, Derek (1984), *Reasons and Persons*, Oxford: The University Press
- (1997), 'Reasons and Motivation' in. *Proceedings of the Aristotelian Society*, Supplementary Volume 71 (1997), pp. 99-130
  - (2001), 'Rationality and Reasons' in. Dan Egonsson, Björn Petersson, Jonas Josefsson, Toni Rønno-Rasmussen (eds.): *Exploring Practical Philosophy: From Action to Values*, Aldershot: Ashgate, pp. 17-41
  - (ms), *Climbing the Mountain*, earlier draft: *Rediscovering Reasons*
- Persson, Ingmar (1999), 'Dancy on Desire and the Internalism of Reasons' in. *Theoria* 99:2-3 (1999), pp. 156-170
- Pettit, Philip (1987), 'Humeans, Anti-Humeans, and Motivation' in. *Mind* 97:388 (1987), pp. 590-593

Pettit, Philip and Smith, Michael (1990), 'Backgrounding Desire' in. *Philosophical Review* 99:4 (1990), pp. 565-592

Phillips D. Z. (1977), 'In Search of the Moral 'Must': Mrs. Foot's Fugitive Thought' in. *The Philosophical Quarterly* 27:107 (1977), pp. 140-157

Platts, Mark (1979), *Ways of Meaning*, London: Routledge and Kegan Paul

- (1981), 'Moral Reality and the End of Desire' in. Mark Platts (ed.), *Reference, Truth, and Reality*, London: Routledge and Kegan Paul, pp. 69-82

- (1991), *Moral Realities: an Essay in Philosophical Psychology*, London: Routledge and Kegan Paul

Price, Huwe (1989), 'Defending Desire-as-Belief' in. *Mind* 98:389 (1989), pp. 119-127

Quinn, Warren (1993), 'Putting Rationality in Its Place' in. *Morality and Action*, New York: Cambridge University Press, pp.228-255

Rabinowicz, Wlodek and Rønnow-Rasmussen, Toni (2004), 'The Strike of the Demon: On Fitting Pro-Attitudes and Value' in. *Ethics* 114:3 (2004), pp. 391-424

Radcliffe, Elizabeth (1999), 'Hume on the Generation of Motives: Why Beliefs Alone Never Motivate' in. *Hume Studies* 25:1-2 (1999), pp. 101-122

Radzik, Linda (1999), 'A Normative Regress Problem' in. *American Philosophical Quarterly*, 36:1 (1999), pp. 35-47

Railton, Peter (1990), 'Naturalism and Prescriptivity' in. *Social Philosophy and Policy*, 7:1 (1990), pp. 151-174

- (1993a), 'What the Non-Cognitivist Help us to See the Naturalist Must Help us to Explain' in. John Haldane and Crispin Wright (eds.): *Reality, Representation, and Projection*, Oxford: The University Press, pp. 279-300



- (1993b), 'Reply to David Wiggins' in. *Reality, Representation, and Projection*, pp. 315-328
- (1997), 'Moral Realism' in. S. Darwall, P. Railton, A. Gibbard (eds.): *Moral Discourse and Practice, Some Philosophical Approaches*, Oxford: The University Press, pp. 137-163,
- (2003a), 'Facts and Values' in. *Facts, Values, and Norms: Essays Towards a Morality of Consequence*, New York: Cambridge University Press, pp. 43-69
- (2003b), 'On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action' in. *Facts, Values, and Norms*, pp. 293-322
- (2004), 'How to Engage Reason: The Problem of Regress' in. R. Jay Wallace, Michael Smith, Samuel Scheffler, and Philip Pettit (eds.), *Reasons and Value: Essays on the Moral Philosophy of Joseph Raz*, Oxford: The University Press, pp. 176-201

Rapoport, J. L. (1989), *The Boy Who Couldn't Stop Washing: The Experience and Treatment of Obsessive-Compulsive Disorder*, New York: Penguin Books

Rawls, John (1971), *A Theory of Justice*, New York: Oxford University Press

- (1980), 'Kantian Constructivism in Moral Theory', Dewey Lectures, in. *Journal of Philosophy* 77:9 (1980), pp. 515-572
- (2000), *Lectures on the History of Moral Philosophy*, Barbara Herman (ed.), Cambridge, Mass.: Harvard University Press

Raz, Joseph (1975), *Practical Reason and Norms*, Princeton, N.J.: The University Press

- (1999a), 'Incommensurability and Agency' in. *Engaging Reason*, Oxford: The University Press, pp. 46-67
- (1999b), 'The Truth in Particularism' in. *Engaging Reason*, pp. 218-246
- (1999c), 'On the Moral Point of View' in. *Engaging Reason*, pp. 247-272

Ridge, Michael (2003), 'Moral Non-Naturalism' in. *Stanford Encyclopedia of Philosophy*, (2003 Winter Edition), Edward N. Zalta, ed., URL = <http://plato.stanford.edu/entries/moral-naturalism>

Rosati, Connie S. (1995a), 'Persons, Perspectives and Full Information Accounts of the Good' in. *Ethics* 105:2 (1995), pp. 296-325

- (1995b), 'Naturalism, Normativity, and the Open Question Argument' in. *Noûs* 29:1 (1995), pp. 46-70

- (1996), 'Internalism and the Good for a Person' in. *Ethics* 106:2 (1996), pp. 297-326

- (2000), 'Brandt's Notion of Therapeutic Agency' in. *Ethics* 110:4 (2000), pp. 780-811

- (2003), 'Agency and the Open Question Argument' in. *Ethics* 113:3 (2003), pp. 490-527

Ross, Peter W. (2002), 'Explaining Motivated Desires' in. *Topoi* 21:1-2 (2002), pp. 199-207

Scanlon, Thomas M. (1998), *What We Owe to Each Other*, Cambridge, Mass.: Belknap Press

- (2002), 'Replies' in. *Social Theory and Practice*, 28:2 (2002), pp. 337-358

- (2004), 'Reasons: A Puzzling Duality?' in. R. Jay Wallace, Philip Pettit, Samuel Scheffler (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, Oxford: Clarendon Press, pp. 231-246

- (ms), 'Structural Irrationality' in. G. Brennan, R. Goodin, and M. Smith (eds.), a volume of essays in honour of Philip Pettit

Schiffer, Stephen (1976), 'A Paradox of Desire', in. *American Philosophical Quarterly*, 13:3 (1976), pp. 195-203

Schick, Frederic (1991), *Understanding Action: An Essay on Reasons*, Cambridge: The University Press

Schroeder, Timothy (2004), *Three Faces of Desire*, New York: Oxford University Press

Schueler, G. F. (1995), *Desire: Its Role in Practical Reason and the Explanation of Action*, A Bradford Book, Cambridge, Mass.: MIT Press

Searle, John R. (1983), *Intentionality*, Cambridge: The University Press

- (2001), *Rationality in Action*, Cambridge, Mass.: Bradford Book, MIT Press

Shafer-Landau, Russ (2003), *Moral Realism: A Defense*, Oxford: Clarendon Press

Shaver, Robert (2006), 'Sidgwick on Moral Motivation' in. *Philosophers' Imprint*, 6:1 (2006), pp. 1-14, URL = [www.philosophersimprint.org/006001](http://www.philosophersimprint.org/006001)

Sidgwick, Henry (1907), *The Methods of Ethics*, Cambridge: The University Press, 7th ed.

Sillins, Nicholas (ms), 'Are Reasons for Belief Facts?'

Simpson, Evan (1999), 'Between Internalism and Externalism in Ethics' in. *Philosophical Quarterly* 49:195 (1999), pp. 201-214

Skorupski, John (2000), 'Irrealist Cognitivism' in. Jonathan Dancy (ed.): *Normativity*, Oxford: Blackwell, pp. 116-139

- (2001), 'Commentary on Svavarsdóttir's 'Moral Cognitivism and Moral Motivation'', in. *Brown electronic Article Review Service*, Jamie Dreier and David Estlund, editors URL = <http://www.brown.edu/Departments/Philosophy/bears/homepage.html>, Posted August 2001

- (2002), 'The Ontology of Reasons' in. *Topoi* 21:1-2 (2002), pp. 113-124

- (2006), 'Propositions about Reasons' in. *European Journal of Philosophy*, 14:1 (2006), pp. 26-48

- (ms a), *Reasons*

- (ms b), 'The Concept of a Person's Good', URL = <http://www.st-andrews.ac.uk/academic/philosophy/STAFF/jmswip.htm>

Smith, Michael (1987), 'The Humean Theory of Motivation' in. *Mind* 96:381 (1987), 36-61

- (1988a), 'Reason and Value' in. *Proceedings of the Aristotelian Society* 88 (1988), pp. 243-258
  - (1988b), 'On Humeans, Anti-Humeans, and Motivation: A Reply to Pettit' in. *Mind* 97:388 (1988), pp. 589-595
  - (1994), *The Moral Problem*, Oxford: Blackwell
  - (1995), 'Reply to Ingmar Persson's Critical Notice of *The Moral Problem*', in. *Theoria* 61:2 (1995), pp. 159-181
  - (1997), 'A Theory of Freedom and Responsibility' in. Cullity, Gareth and Gaut, Berry (eds.): *Ethics and Practical Reason*, pp. 293-321, Oxford: The University Press
  - (1999), 'The Definition of 'Moral'' in. D. Jamieson (ed.), *Singer and His Critics*, Oxford: Blackwell
  - (2003), 'Humeanism, Psychologism, and the Normative Story' in. *Philosophy and Phenomenological Research*, 67:2 (2003), pp. 460-467
  - (2004a), 'The Incoherence Argument: Reply to Shafer-Landau' in. Michael Smith: *Ethics and the A Priori*, Oxford: The University Press, pp. 43-55
  - (2004b), 'Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion' in. *Ethics and the A Priori*, pp. 114-135
- Sobel, David (1994), 'Full Information Accounts of Well-Being' in. *Ethics* 104:4 (1994), pp. 784-110
- (1999), 'Pleasure as a Mental State' in. *Utilitas*, 11:2 (1999), pp. 230-4
  - (2001a), 'Subjective Accounts of Reasons for Action' in. *Ethics* 111:3 (2001), pp. 461-92
  - (2001b), 'Explanation, Internalism, and Reasons for Action' in. *Social Philosophy and Policy* 18:2 (2001), pp. 218-236

- (2005), 'Pain for Objectivists: the Case of Matters of Mere Taste' in. *Ethical Theory and Moral Practice*, 8:4 (2005), pp. 437-457

Stampe Dennis W. (1987), 'The Authority of Desire' in. *Philosophical Review*, 96:3 (1987), pp. 341-81

Stevenson, Charles L. (1944), *Ethics and Language*, New Haven, Yale University Press

- (1963), 'The Emotive Meaning of Ethical Terms' in. *Facts and Values*, Westport, Conn.: Greenwood, 1963, pp. 10-32

Stocker, Michael (1979), 'Desiring the Bad: An Essay in Moral Psychology' in. *The Journal of Philosophy*, 77:12 (1979), pp. 738-753

- (1981), 'Values and Purposes: the Limits of Teleology and the Ends of Friendship' in. *Journal of Philosophy*, 78:12 (1981), pp. 747-765

Stratton-Lake, Philip (2000), 'Expression, Description, and Normativity: Simon Blackburn's *Ruling Passions*' in. *Res Publica* 6:1 (2000), pp. 117-125

Stroud, Barry (1977), *Hume*, London: Routledge and Kegan Paul

Sturgeon, Nicholas (1985a), 'Moral Explanations' in. David Copp and David Zimmerman (eds.), *Morality, Reason, and Truth*, Totowa: Rowman and Allanheld, pp. 181-228

- (1985b), 'Gibbard on Moral Judgment and Norms' in. *Ethics* 96:1 (1985), pp. 22-33

- (1986a), 'What Difference Does It Make Whether Moral Realism Is True?' in. *Proceedings of the 1985 Spindel Conference*, ed. N. Gillespie, *Southern Journal of Philosophy*, Supplement 24 (1986), pp. 115-142

- (1986b), 'Harman on Moral Explanations of Natural Facts' in. *Proceedings of the 1985 Spindel Conference*, ed. N. Gillespie, *Southern Journal of Philosophy*, Supplement 24 (1986), pp. 69-78

Sumner, L. Wayne (1996), *Welfare, Happiness, and Ethics*, Oxford: Clarendon Press

Svavarsdóttir, Sigrun (1999), 'Moral Cognitivism and Moral Motivation' in. *Philosophical Review* 108:2 (1999), pp. 161-219

Trigg, Roger (1970), *Pain and Emotion*, Oxford: The University Press

Velleman, J. David (1988), 'Brandt's Definition of "Good"' in. *The Philosophical Review*, 97:3 (1988), pp. 353-371

- (2000a), 'The Guise of the Good' in. *The Possibility of Practical Reason*, New York: Oxford University Press, pp. 99-122

- (2000b), 'What Happens When Someone Acts?' in. *The Possibility of Practical Reason*, pp. 122-143

- (2000c), 'The Possibility of Practical Reason' in. *The Possibility of Practical Reason*, pp. 170-199

Wallace, R. Jay (1990), 'How To Argue About Practical Reason' in. *Mind* 99:395 (1990), pp. 355-380

- (2001), 'Commentary on Svavarsdóttir's 'Moral Cognitivism and Moral Motivation'', in. *Brown electronic Article Review Service*, James Dreier and David Estlund, editors, URL = <http://www.brown.edu/Departments/Philosophy/bears/homepage.html>, Posted August 2001

- (2003), 'Explanation, Deliberation, and Reasons' in. *Philosophy and Phenomenological Research*, 67:2 (2003), pp. 429-435

Watson, Gary (1975), 'Free Agency' in. *Journal of Philosophy* 72:8 (1975), pp. 205-220

Wedgwood, Ralph (2002), 'Practical Reasoning as Figuring Out What is Best: Against Constructivism' in. *Topoi* 21:1-2 (2002), pp. 139-152

- (2005), 'Railton on Normativity' in. *Philosophical Studies* 126:3 (2005), pp. 463-479

Wiggins, David (1990), 'Moral Cognitivism, Moral Relativism, and Motivating Moral Beliefs' in. *Proceedings of the Aristotelian Society* 91 (1990), pp. 61-86

- (1993), 'Cognitivism, Naturalism, and Normativity: A Reply to Peter Railton' in. John Haldane and Crispin Wright (eds.): *Reality, Representation, and Projection*, Oxford: The University Press, pp. 301-313

Williams, Bernard (1973), 'A Critique of Utilitarianism' in. Smart, J. J. and Williams, Bernard: *Utilitarianism, For and Against*, Cambridge: The University Press, pp. 77-151

- (1981a), 'Internal and External Reasons' in. *Moral Luck*, Cambridge: The University Press, pp. 101-114, 1981
- (1981b), 'Ought and Moral Obligation' in. *Moral Luck*, pp. 114-124
- (1981c), 'Practical Necessity' in. *Moral Luck*, pp. 124-31
- (1985), *Ethics and the Limits of Philosophy*, Cambridge: The University Press
- (1995a), 'Internal Reasons and the Obscurity of Blame', in. *Making Sense of Humanity*, Oxford: The University Press, pp. 35-45
- (1995b), 'Moral Incapacity' in. *Making Sense of Humanity*, pp. 46-56
- (1995c), 'Replies', in. Altham, J. E. J. and Harrison, Ross (eds.): *World, Mind and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, Cambridge: The University Press, pp. 185-225
- (2001), 'Postscript: Some Further Notes on Internal and External Reasons' in. Millgram, Elijah (ed.): *Varieties of Practical Reasoning*, Cambridge, Mass.: Bradford Book, MIT Press, pp. 91-7

Williamson, Timothy (2000), *Knowledge and Its Limits*, Oxford: The University Press

Woods, Michael (1972), 'Reasons for Action and Desires' in. *Proceedings of the Aristotelian Society*, Supplementary Volume 46 (1972), pp. 182-201

Zagzebski, Linda (2003), 'Emotion and Moral Judgment' in. *Philosophy and Phenomenological Research* 66:1 (2003), pp. 104-124

Zangwill, Nick (2003), 'Externalist Moral Motivation' in. *American Philosophical Quarterly* 40:2 (2003), pp. 143-154

Zimmerman, David (2003), 'Why Richard Brandt Does Not Need Cognitive Psychotherapy, and Other Glad News about Idealized Preference Theories in Meta-Ethics' in. *The Journal of Value Inquiry* 37:3 (2003), pp. 373-394