University of Sussex

A University of Sussex PhD thesis

Available online via Sussex Research Online:

http://sro.sussex.ac.uk/

Reason in Action:
A Realist Account

Ezra Benjamin Cohen

PhD Philosophy
University of Sussex
August 2017

Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:………………………………………

University of Sussex
Ezra Benjamin Cohen
PhD Philosophy

Reason in Action:
A Realist Account

Summary

This thesis argues against the Humean theory of practical reasons, criticising its foundations in philosophical and moral psychology. It develops a realist account of value-based reasons, underpinned by a distinctive cognitivist moral psychology, and a non-causalist account of the rational explanation of action.

Contemporary Humeans reject Hume's own theory of thought, but this leaves the Humean theory of practical reasons without justification for a conception of desire as non-cognitive and not open to fundamental rational evaluation. Two possible strategies for filling this justificatory gap are (i) an appeal to grammatical considerations about the attribution of desires and their content, or (ii) an appeal to distinctions in respect of direction of fit. I argue that neither is successful.

Kant's moral psychology provides the key to an alternative account, but is unsatisfactory due to its acceptance of a theory of thought which is relevantly similar to Hume's, and of non-compulsory rationalist presuppositions. Separated from these aspects, Kant's insights open a path to developing a conception of desire as essentially rationally evaluable. I argue that, in addition to such a conception, we should accept an account of rational attitudes as constitutively normative. On the basis of these two views, I argue that desire is a kind of evaluative belief.

An independently plausible account of reasons takes them to be evaluative facts, and this neatly connects to the normative philosophical psychology. I consider the implications of such a view for the rational explanation of action, arguing that while causal theories of action and action-explanation are unacceptable, the normative philosophical psychology allows the development of non-causal alternatives to them. The non-causal account of action and action-explanation leaves space for an explanatory role for reasons themselves, beyond that provided by merely psychological explanation, as well as an explanatory role for an agent's character and emotions.

[T]he temperate man desires the things he ought, as he ought, and when he ought; and this is what reason directs.

~ Aristotle[1]

The Philosopher's business here is with the springs of action insofar as we can clarify these for ourselves as reasons for acting, and with the nature of desire.

~ David Wiggins[2]

---

[1] Aristotle, *Nicomachean Ethics*, 1119b13-18, as translated in M. F. Burnyeat, 'Aristotle on Learning to Be Good', p. 80

[2] David Wiggins, 'Weakness of Will, Commensurability, and the Objects of Deliberation and Desire', p. 262

Contents

Part II: Reasons, Actions, and Explanation

4: Values, Reasons, and Actions

5: The Explanation of Action

Acknowledgements

When I arrived at Sussex as an undergraduate, I never dreamt that I would still be here, almost exactly ten years on, submitting a PhD thesis. It would not be *entirely* unfair to lay the blame for it at the feet of Murali Ramachandran, who first suggested that I consider it – unforgivably prematurely – after my first term.

In the intervening decade, I have had the great fortune of being in a department as good as anyone has a right to hope for, and have benefitted and learned something (philosophically or otherwise) from pretty much every member of faculty. Above all, I am grateful to Michael Morris, who has been my ideal, but very, very, (sometimes unbelievably) real supervisor. It is impossible to convey the profound influence Michael has had on my thinking; while I know that whatever merit there is in this thesis is largely due to him, I have no doubt that his contribution is greater than I realise. Michael has been an incredible teacher, support, and, I would like to think, friend. He has always had much more faith in me, and confidence that it would all come together in the end, than I did. I hope I have not made too much of a hash of it, and that he will not be too disappointed that I have not disagreed with him more.

For company, chats over lunch, being prepared to talk philosophy more than any non-philosopher could understand the desire to do, giving feedback and asking good questions at presentations, and generally for teaching me a lot, my thanks go to Richard Weir, Elaine Finbarr O'Connell, Andre Almeida, Gabriel Martin, Andy Rebera, Tom Beament, Alice MacNair, Tom Swaine-Jameson, Chris Ferguson, Alex Elliott, David Martinez Rojas, and Patrick Levy. Jacob Berkson, Tim Carter, and Robb Dunphy deserve special mention and thanks for their many, many generous hours of listening to, reading and critiquing my work. Thanks also to my parents and my brother for their encouragement, and for repeatedly asking me to explain what I was working on.

Most importantly, my deepest thanks and love go to Alice; she has read and edited pretty much everything I've ever written, and any semblance of literacy I have is due to her. For supporting and tolerating me through the whole process, and most of all for spending her life with me, I am forever in her debt.

For Ida, who ate it

Part I

Philosophical and Moral Psychology

Chapter 1

Introduction:

Excavating the Humean Theory of Reasons

How much of Hume's theory should we keep, how much should we reject?

~ Michael Smith[1]

1.1 Introduction

This thesis is about the role of reason in action. It argues that our rational faculty plays a fundamental role in our ability to act intentionally for reasons. The account of agency which I shall defend encompasses a robust realism concerning value-based practical reasons, a form of cognitivism about normative or, as I will usually prefer to say, evaluative judgement, and an internalist view of the motivational implications of evaluative judgements. Moreover, its being, in my view, impossible to say much that is satisfying or illuminating about reasons or about motivation without also considering what is motivated and explained, and what explanatory character such explanation has, this thesis also develops a distinctive non-causalist account of the nature and explanation of action.

Anyone reading David Hume's account of the relation between reason and morality in his *Treatise of Human Nature* must be struck by the general movement of his discussion from his philosophical psychology to the nature of morality and moral judgement. There are good reasons why it should take such a shape, the most important of which is that moral considerations are supposed to be *practical*. This notion can be understood in two ways, either as the notion that moral considerations are, or provide us with, *reasons* for acting in accordance with them, or as the notion that they are, or provide us with, *motivations* for acting in accordance with them. In fact, I think that both these ways of understanding the practicality of moral considerations are correct, and not really independent. What unites them is that moral considerations can be

---

[1] Michael Smith, *The Moral Problem*, p. 91

explanatory of what someone does, and we can explain someone's actions by saying what their reasons were for acting, and by saying what motivated them to act. This is not to set up two different rival explanations, for the notion of a reason for acting is unintelligible except as the notion of the sort of thing whose recognition could motivate one to act, and, of course, thereby explain one's action. The sense in which moral considerations are, or provide us with, motivations for acting, then, should be understood simply as their *being* the sort of things which are, or provide us with, reasons which could motivate us to act in accordance with them. To borrow a phrase from Thomas Nagel, 'To accept a reason for doing something is to accept a reason for *doing* it, not merely for *believing* that one should do it.'[2]

If we are to make sense of the practicality of morality, we need first to have some account of what is involved in motivation, and that means we need, at least in outline, a philosophical and moral psychology. Without saying what the relation is between morality and psychology, it would remain obscure how morality could be practical, or even what it would *be* for it to be practical. So a satisfying account of the nature of morality requires a prior philosophical and moral psychology, which means the discussion of these subjects must exhibit just the kind of movement found in Hume.

As we shall see in §1.2, in Hume's hands, philosophical and moral psychology is shaped around a distinction between two sorts of psychological states: exercises of reason or the understanding, i.e. beliefs or judgements, and the passions, centrally desires and aversions. The resulting view takes motivation to be entirely a matter of the operation of states on the passionate side of this distinction. On the basis of this, morality – which is practical – is assimilated to the passions and Hume consequently denies it to be 'deriv'd from reason'.[3] In contrast to Hume, Nagel, in *The Possibility of Altruism*, argues for a form of moral rationalism, which sees both prudential and moral requirements as derivable from the impersonal structures of thought that are involved in being a unified rational agent at all. But the movement of thought is just like that in Hume, and Nagel even goes so far as to declare: 'I conceive ethics as a branch of psychology. My claims concern its foundation, or ultimate motivational basis.'[4] Hume and Nagel end up with radically opposed accounts of the nature of morality, but both

---

[2] Thomas Nagel, *The Possibility of Altruism*, p. 64
[3] David Hume, *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*, III.i.1, p. 455
[4] Nagel, *The Possibility of Altruism*, p. 3

think of its nature as following, in a more or less straightforward way, from their respective philosophical and moral psychologies.

It should come as no surprise that when we think of moral considerations as being essentially *practical* considerations, we find ourselves thinking of them as being, or providing us with, reasons for action. On a familiar and compelling conception of the notion of a reason, reasons just are the kinds of things which favour or justify what they are reasons for;[5] moral considerations are one central and obvious, perhaps even overriding, sort of consideration about what would favour or justify actions (or states of affairs, or character etc.), and reasons for action are the kinds of things in terms of which we might explain someone's action. That an understanding of specifically moral considerations (being practical, and so being, or providing us with, reasons to act) requires a prior account of philosophical and moral psychology, therefore, should not be seen as something peculiar to the moral domain. An understanding of reasons to act in general requires this. Indeed, it is a mark of the progression of philosophical discussions of normative or evaluative issues over the last half-century that there has been a shift of focus from the nature of moral considerations and moral motivation in particular, to a focus on the nature of reasons for action and motivation in general.[6] Disagreements between non-cognitivists in the tradition deriving from Hume, on the one hand, and realist or rationalist cognitivists of various sorts, on the other, are now played out both within and without the moral domain, and concern reasons for action, belief, and feeling in general.[7]

The account of reasons that I shall defend is a realist one, and so, while I agree with Hume and Nagel that the starting point should be in philosophical and moral psychology, I cannot accept that, to update Nagel's phrase, the theory of reasons is just a branch of psychology. The role of an account of philosophical and moral psychology, in my view, is not to provide the foundations of an account of reasons by itself, but to give an independently plausible account of our motivational and more broadly psychological constitutions, in the context of which a realist account of reasons for action, fully complemented with their practical and explanatory significance, is

---

[5] See in particular T. M. Scanlon, *What We Owe To Each Other*, ch. 1.

[6] This shift in focus has also brought the spotlight onto reasons for belief or emotion. See e.g. T. M. Scanlon, *Being Realistic About Reasons*; John Skorupski, *The Domain of Reasons*.

[7] For this kind of generalised non-cognitivism, see e.g. Alan Gibbard, *Wise Choices, Apt Feelings*; for the generalised realist view, see e.g. Jonathan Dancy, *Practical Reality*, and Scanlon, *Being Realistic About Reasons*, and for the generalised rationalist view, see e.g. Christine M. Korsgaard, *The Sources of Normativity*, and Skorupski, *The Domain of Reasons*.

intelligible, defensible and, I hope, attractive. Jonathan Dancy puts the point nicely when he claims that 'Hume was right to rest his non-cognitivism on his theory of motivation, and […] any effective opposition has to do the same, i.e. to find a rival theory of motivation that will re-establish the possibility that […] reasons are both cognitive and able to motivate in their own right.'[8]

Part I of this thesis is accordingly concerned with issues in philosophical and moral psychology, in particular with the task of undermining Humean and Kantian accounts, and developing an alternative which keeps their best features but allows a realist account of reasons. In Part II, we will proceed to develop a realist account of reasons, an account of actions, and an account of the explanation of action which shows how facts about our psychology, our reasons, and our actions are all related.

In my view, the major obstacle to a realist account of reasons is a cast of thought which is the philosophical legacy of Hume.[9] Hume's argument that the practicality of morality requires it to be based on the passions rather than on reason, as well as certain other arguments against morality being a matter of conformity or contrariety with reason, constitute the sharp point of the Humean sword, and it is these which have been inherited and reformulated by contemporary Humeans. Not even the most orthodox contemporary Humean, however, accepts Hume's own antiquated philosophical psychology. This raises some pressing questions. Just how far, and in what respects, do contemporary Humeans depart from Hume's own philosophical psychology? Which aspects of Hume's psychology *should* be rejected, and are these the same as those which the Humeans actually do reject? More importantly, what are the consequences of rejecting those aspects of Hume's theory? More important still: how do the Humean arguments for non-cognitivism about reasons for action in general fare once the sharp point has been, as it were, snapped from the shaft, and the arguments are examined in isolation from their original context in Hume's own philosophical and moral psychology?

In this introductory chapter, I propose to excavate contemporary Humean theories, beginning with an examination of the fundamental aspects of Hume's own account (§1.2), before tracing the major respects in which contemporary versions of Humeanism, recognising where Hume went wrong, take a different view (§1.3-§1.5).

---

[8] Jonathan Dancy, *Moral Reasons*, p. *x*
[9] Although many strands in that cast of thought are found, in a less crystalised from, in Early Modern and Enlightenment philosophy more generally.

What I will argue (§1.6) is that the resulting Humean view tries to maintain the spirit of Hume's account, but the motivation and justification for that account is lost in the process of correcting for Hume's mistakes. In the etymologically literal sense, the Humean theory will have been excavated, that is, hollowed out. I will conclude in §1.7 with an outline of Chapters 2 to 5. Chapter 2 considers some strategies to re-motivate the Humean theory; Chapter 3 develops an alternative philosophical and moral psychology suited to underpin a realist account of reasons; Chapter 4 offers an account of reasons and suggests the framework for an understanding of the nature of actions, and Chapter 5 brings these together with the psychology of Chapter 3 in order to provide an account of the explanation of action.

1.2 Hume's Critique of the Practical Role of Reason

According to Hume's famous pronouncement, 'Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.'[10] Hume's bold and controversial statement is meant to be a corrective to what he takes to be the mistaken view that the faculty of reason may provide the ground of our actions.[11] Against this view, Hume argues that it follows from his philosophical psychology that it is impossible for our actions to be grounded in reason.

Hume's philosophical psychology is based around a distinction he makes between two parts of human nature: our rational and our passionate nature. According to his theory, all experience and activity of the mind consists in different kinds of *perceptions*, amongst which there is a hierarchy of distinctions. The first is between *impressions* and *ideas*, both of which may be either simple or complex. Impressions 'enter [the mind] with most force and violence', while ideas are 'the faint images of these in thinking and reasoning'.[12] Upon receiving any simple impression, the mind produces a copy of it as a simple idea which resembles and represents it, and all simple ideas in this way have their origin in the impressions which they represent.[13] Hume distinguishes between 'primary ideas', which represent impressions, and 'secondary ideas, which are the images of the primary'[14]. Impressions similarly admit of a

---

[10] Hume, *Treatise*, II.iii.3, p. 415
[11] My use of 'ground' here, rather than 'cause', which would be more natural with regard to Hume, is supposed to provide a neutral formulation on issues which will be germane later.
[12] Ibid. I.i.1, p. 1
[13] Ibid. p. 4
[14] Ibid. p. 6

primary/secondary distinction; primary impressions are 'those of SENSATION' and
secondary impressions are 'those of REFLEXION'.[15] But while secondary ideas are
representations of primary ideas, secondary impressions of reflexion are not
representations of primary or original impressions of sensation. Rather:

> [They are] derived in a great measure from our ideas, and that in the following order.
> An impression first strikes upon the senses, and makes us perceive heat or cold, thirst or
> hunger, pleasure or pain of some kind or other. Of this impression there is a copy taken
> by the mind, which remains after the impression ceases; and this we call an idea. This
> idea of pleasure or pain, when it returns upon the soul, produces the new impressions of
> desire and aversion, hope and fear, which may properly be called impressions of
> reflexion, because derived from it.[16]

What this shows is that the crucial difference between impressions and ideas, for
Hume, cannot really be that impressions are more forceful, vivacious, and violent, while
ideas are calmer and fainter, as he officially claims. Nor can it be that ideas are
'derived' while impressions are not, for secondary impressions as well as secondary
ideas are derived from primary ideas. The crucial difference that Hume needs to
distinguish between them is just that ideas resemble or *represent* what they are derived
from; as we would now put it, they have *representational content*, while impressions do
not.[17]

In Book II of the *Treatise,* Hume gives as paradigms of secondary impressions
'the passions, and other emotions resembling them',[18] and it is the passions, in
particular desire and aversion, to which Hume gives central place in his moral
psychology.[19] As he says in the opening of *Treatise* II.iii.3, whilst it is usual in
philosophy 'to talk of the combat of passion and reason, [and] to give preference to
reason, and to assert that men are only so far virtuous as they conform themselves to its
dictates', he instead 'shall endeavour to prove *first,* that reason alone can never be a

---

[15] Ibid. I.i.2, p. 7; in II.i.1 he calls them *original* and *secondary* impressions.

[16] Ibid. I.i.2, pp. 7-8

[17] See e.g. ibid. I.i.1, p. 4, or I.iii.14, p. 157, 'Ideas always represent their objects or impressions', and
also II.iii.3, p. 415 where Hume asserts this difference. I return to this below. (In other passages Hume
seems to suggest that impressions may be representational, e.g. ibid. I.iv.5, p. 233. This, I think, is an
inconsistency in Hume, and in conflict with the criterion he needs for his distinction.) Cf. Elizabeth S.
Radcliffe, 'Hume's Psychology of the Passions: The Literature and Future Directions', §3.1.

[18] Hume, *Treatise*, II.i.1, p. 275

[19] In this thesis, I shall mostly speak of desire, but what I say should be taken also to apply, *mutatis
mutandis*, to aversions. We could think of them as desires of opposite polarities, as it were, or 'flavours'
of desire, as Neil Sinhababu puts it ('The Humean Theory of Motivation Reformulated and Defended', p.
468).

motive to any action of the will; and *secondly,* that it can never oppose passion in the direction of the will.'[20]

Let us look at the first part of Hume's endeavour – his attempt to show that reason alone can never motivate action. Passions, according to Hume, are produced by the expectation of pleasure or pain, and are forces which impel us to action. As he says, ''Tis obvious, that when we have the prospect of pain or pleasure from any object, we feel a consequent emotion of aversion or propensity, and are carry'd to avoid or embrace what will give us this uneasiness or satisfaction.'[21] By contrast, for Hume, reason is not a force that impels us to action, but rather it informs our emotions (passions) which in turn impel us to act. As he defines it, reason, or as he now calls it 'the understanding,' is concerned with 'the discovery of truth and falshood',[22] operating in two ways: first as it judges from demonstration, i.e. from deductive, *a priori* proof, and second as it judges from probability, i.e. from empirical, *a posteriori* discovery of the relations of objects, particularly in the form of causal reasoning. Here we have two basic conceptions of the nature of passions and of reason which figure in Hume's argument:

> Passion: Passions are non-representational *forceful impulses*, which cause actions.

> Reason: Reason is the power of *judging* how things are and how they are related, either by *a priori* demonstration, or *a posteriori* discovery.

Hume thinks that it will be just obvious that *a priori* demonstrative knowledge could not cause our actions, but could only influence them insofar as it informs our reasoning about causal relations.[23] As for the role of the understanding, as it supplies empirical knowledge of causal relations, Hume also denies that this causes our actions, instead claiming that it influences us in virtue of extending and directing our desires and aversions.[24] As we saw, passions – secondary impressions – are supposed to be

---

[20] Hume, *Treatise*, II.iii.3, p. 413
[21] Ibid. p. 414
[22] Ibid. III.i.1, p. 458
[23] Ibid. II.iii.3, p. 413-4
[24] Cf. Christine M. Korsgaard, 'Skepticism about practical reason', pp. 313-4

produced in response to our ideas of pleasure or pain, and are directed at the sorts of things or events which produce our original sensations in *expectation* of their producing similar sensations. But causal reasoning allows us to gain knowledge of which events are the causal antecedents of pleasure and pain, and the ways in which we may affect things so as to cause the events which produce them, and this process of reasoning extends our desires and aversions to those things which are causally related to such sensations.[25] As John Broome emphasises, though, for Hume:

> Reason has an influence on what we do only because it can guide our beliefs. Our beliefs in turn affect the desires we have, and hence what we do. Beyond that, reason can do nothing. In particular, though our beliefs affect our desires, *reason plays no part in mediating this effect*. Reason does not guide our desires.[26]

This is Hume's main psychological argument that it is impossible for reason to provide the ground of our actions, which we may reconstruct thus:

(HP1) Something can be the ground of an action only if it is its cause.

(HP2) Something can be the cause of an action only if it is a *forceful impulse*.

(HP3) Reason is only the power of judging how things are, and how they are related.

So,

(HP4) Nothing belonging to the power of reason is a forceful impulse.

So,

(HP5) Nothing belonging to the power of reason is the cause of any action.

So,

(HP6) Nothing belonging to the power of reason is the ground of any action.

So,

(HP7) Reason is not a power that can provide the ground of any action.

In the second part of his endeavour, Hume argues that, in addition to being incapable of providing the ground of any action, it is impossible for reason to oppose any action, or for any motivating passion to be contrary to it. Reason cannot oppose an action, according to Hume, because to do so it would have to meet the same conditions

---

[25] Hume, *Treatise*, II.iii.3, p. 414
[26] John Broome, 'Reasons and Motivation', p. 131, emphasis added.

on being capable of providing the ground of actions, conditions he has already argued it *fails* to meet: 'Nothing can oppose or retard the impulse of passion, but a contrary impulse; and if this contrary impulse ever arises from reason, that latter faculty must have an original influence on the will, and must be able to cause, as well as hinder any act of volition.'[27]

It is impossible for a passion to be contrary to reason, he argues, because this would require reason and passion either to be directly contradictory, e.g. for a judgment of reason to have the representational content that *p*, while a passion had the content that *not p*, or, at least, for them to have contents which were inconsistent because they imply such a contradiction. But we saw earlier that Hume's theory requires distinguishing between secondary ideas and secondary impressions – passions – on the basis of the former being representational while the latter are not. As he says:

> A passion is an original existence […] and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possest with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high.[28]

If passions necessarily lack any representational content, then *a fortiori* they cannot have a content which is contrary to, or inconsistent with, that of a judgment. For Hume, then, a passion cannot, strictly speaking, be unreasonable.[29] As Broome observes, though, 'The Humean theory of reason is surely implausible. If you want to stay alive, and you believe that jumping [from a burning building into a river] is the only way to do so, and the result is that you want to remain on the windowsill and sing, this want is surely irrational. Reason surely tells you at least that.'[30]

Setting aside this implausibility for the moment, the important point for now is just that reason, according to Hume's argument, can neither provide the ground for, nor oppose, any action. We should note, however, that the requirements he sets for something to be capable of being the ground of an action, or of opposing it, precisely match his definition of the nature of a passion – to be the ground of an action a thing

---

[27] Hume, *Treatise*, II.iii.3, p. 415

[28] Ibid. p. 415. Note Hume's denial here that being thirsty is a passion.

[29] Hume does allow that, loosely speaking, we may regard passions as unreasonable in two derivative ways: when they are based on a false judgment that something exists which in fact does not, and when the choice of means (i.e. the object to which causal reasoning extends the passion) is inadequate to bring about the end, for instance when it is based on false causal reasoning. But, he stresses, 'even then 'tis not the passion, properly speaking, which is unreasonable, but the judgment' (ibid. p. 416).

[30] Broome, 'Reasons and Motivation', p. 132

must be a forceful impulse which causes it, and passions are defined as forceful impulses which may cause actions.

Having claimed to establish that reason is incapable of providing the ground of any action, and that such a ground is always constituted by a passion, Hume goes on to argue that morality is not based in reason either, using arguments that almost exactly mirror those we have just considered. Just as he says that when we experience a passion, we 'are carry'd to avoid or embrace what will give us this uneasiness or satisfaction,'[31] so he argues that 'morals [i.e. moral judgements] have an influence on the actions and affections'.[32] He supports this by the empirical observation 'that men are often govern'd by their duties, and are deter'd from some actions by the opinion of injustice, and impell'd to others by that of obligation,' and by appealing to our efforts to instil morality and formulate moral principles: 'If morality had naturally no influence on human passions and actions, 'twere in vain to take such pains to inculcate it; and nothing wou'd be more fruitless than that multitude of rules and precepts, with which all moralists abound.'[33] Since moral judgements influence people's actions, and nothing belonging to reason can do so, moral judgements cannot, he claims, belong to the power of reason and so cannot really be judgements at all. Hume's moral argument may be reconstructed thus:

> (HM1) Moral 'judgements' are the grounds of some actions.
> (HM2) Nothing belonging to the power of reason is the ground of any action
> (from HP6).
> So,
> (HM3) Moral 'judgements' do not belong to reason.
> So,
> (HM4) Moral 'judgements' are not really judgements at all (from Reason).[34]

Nor, he argues, can morals be contrary to reason, for 'our passions, volitions, and actions [are] original facts and realities, compleat in themselves',[35] i.e. they are non-representational or original existences. As in his earlier argument, for something to be

---

[31] Hume, *Treatise*, II.iii.3, p. 414
[32] Ibid. III.i.1, p. 457
[33] Ibid.
[34] Compare Derek Parfit's reconstruction of this argument in 'Reasons and Motivation', p. 106
[35] Hume, *Treatise*, III.i.1, p. 458

contrary to reason it would have to have a contrary representational content to that of a judgment. Since, according to Hume, neither actions nor their subjective grounds – passions or moral 'judgements' – can be contrary to reason in this way, but either *can* be immoral, it follows that to be moral or immoral is not a matter of being in agreement or disagreement with reason.

According to Hume:

> [These arguments prove] *directly*, that actions do not derive their merit from a conformity to reason, nor their blame from a contrariety to it; and [prove] the same truth more *indirectly*, by shewing us, that as reason can never immediately prevent or produce any action by contradicting or approving of it, it cannot be the source of the distinction betwixt moral good and evil, which are found to have that influence.'[36]

Furthermore, the arguments purport to show that what seem to be moral judgments are not really judgments at all; judgments belong to reason, have representational content, and are truth-apt, but since apparent moral judgments influence actions, they cannot really belong to reason, nor really be representational. As Hume says of our volitions, ''Tis impossible, therefore, they can be pronounced either true or false, and be either contrary or conformable to reason.'[37]

1.3 Desires, Phenomenology, and Representational Content

As we would expect, Hume officially gives an empiricist account of the epistemology of desire; our desires are known to us by their phenomenology, by their force and vivacity. However, his account of the metaphysics of desire is also apparently phenomenological; he claims that desires simply *are* forceful impulses whose impressions on the mind have a distinctive phenomenology, through which we know them. Hume makes this explicit: 'Each of the passions and operations of the mind has a particular feeling [....] This particular feeling *constitutes the very nature of the passion*; and therefore needs not be accounted for.'[38] This is implied by his insistence that they are non-representational 'original existences', and it is because they are such forces that Hume thinks they are suited to be the causes of actions.

Hume's account of desires as contentless mental states, and as mere phenomenological impulses, however, fundamentally misrepresents the nature of desire.

---

[36] Ibid.
[37] Ibid.
[38] Ibid. III.iii.1, p. 590, emphasis added. Cf. Barry Stroud, *Hume*, pp. 159, 163-4.

Desires are not purely phenomenological, like sensations, but are psychological attitudes with content; one does not merely have a desire, which is felt, but, for example, a desire *to listen to some music*. The phrase '*to listen to some music*' here gives the content of the desire, and desires *do* make reference to something else, e.g. to the activity of listening, and to music.[39] Contemporary Humeans therefore reject Hume's account of desires as original existences.

For example, Michael Smith argues that a phenomenological conception of desire is thoroughly unsatisfactory. According to Smith, there are two possible versions of such a conception, a weaker and a stronger. Hume's view, as Smith recognises, is a version of the strong phenomenological conception of desire, according to which 'desires are, like sensations, simply and essentially states that have a certain phenomenological content';[40] that is to say that one has some desire just in case one is subject to its distinctive phenomenological feel, and that is all there is to it. As with sensation, Smith suggests, this might be thought to imply an epistemology of desire according to which one has a desire if and only if one believes that one does, on the grounds that, necessarily, one believes that one is subject to some phenomenological feeling just in case one feels it, and on the assumption that one's feelings are transparent to one.[41] But, he claims, this view is false, and it is a condition of adequacy on an epistemology of desire that it be fallibilist – that it allows that it is possible to be wrong about whether one has some desire or not.[42] If Smith is right, the strong phenomenological conception of desire, a version of which Hume seems to hold, faces severe epistemological difficulties.[43]

But there is a more fundamental problem that arises from treating desires on the model of a kind of feeling at all. Smith argues that were desires to be treated on such a model, the logical form of ascriptions of desire and of sensation would be the same.[44]

---

[39] I will use the term 'attitude' in a sense that entails that it has a content, and 'state' when this is not entailed.

[40] Michael Smith, 'The Humean Theory of Motivation', p. 45

[41] Ibid. pp. 45-6

[42] Ibid. p. 46

[43] Hume comes close to conceding that passions do not always have a distinctive phenomenology, saying that 'there are certain calm desires and tendencies, which, tho' they be real passions, produce *little* emotion in the mind, and are *more* known by their effects than by the immediate feeling or sensation.' (*Treatise*, II.iii.3, p. 417, emphasis added). Against this, and in contradiction to a strict reading of Hume, Stroud claims that according to Hume, 'We do not simply *feel* "calm passions"; their existence and efficacy is not discovered by direct inspection' (*Hume*, pp. 163-4).

[44] It is reasonable to ask why sameness of epistemology should imply sameness of form of ascriptions. Here I will simply accept Smiths claim, since, first, it is hard to see how a divergence might be motivated

But, he says, 'ascriptions of desires, unlike ascriptions of sensations, may be given in the form "A desires that p", where "p" is a sentence. Thus, whereas A's desire to φ may be ascribed to A in the form *A desires that he φs*, A's pain cannot be ascribed to A in the form *A pains that p*.'[45] The point, for Smith, is that it is a second condition of adequacy on a conception of desire that it accommodate desires having propositional content. If Smith is right, it is not merely that Hume denies that desires have content, but that his conception of desire is actually incompatible with their having content.[46]

According to the weak phenomenological conception of desire, meanwhile, desires essentially have a phenomenology, but also have propositional content. But Smith argues that it is not even a *necessary* condition of having a desire that one believes one has it, and this would equally undermine the weak conception.[47] Further, Smith claims that it is anyway objectionable because, in virtue of being a phenomenological conception at all, it 'in no way contributes to our understanding of what a desire as a state with propositional content is, for [it] cannot explain how it is that desires have propositional content'.[48] Of course, Smith realises that there is nothing objectionable about admitting that it is possible for desires to have a distinctive phenomenology, in fact it is clear that *some* desires actually do. But even this would not help with providing an adequate epistemology, for at best it could provide knowledge that one had a desire with a certain phenomenology, but could not provide knowledge of what the content of the desire was, and so not of *what* desire one had.[49] Furthermore, it would still be of no use in explaining the nature of desires such that it would be intelligible that they have content at all.

Hume's account of desires as original existences, then, must be, and is, rejected by contemporary Humeans. But there is a further controversy over the proper specification of their content. We saw that Smith takes it that the form of a desire ascription is '*A* desires that *p*', where '*p*' marks the place of a propositional 'that'-clause, and is substitutable for a complete sentence which gives the content of the

---

by the phenomenological conception itself, and second, because this conception would not be helpful in explaining why they have a form which takes a content clause.

[45] Smith, 'The Humean Theory of Motivation', p. 47

[46] I argued in §1.2 that it is essential for Hume's distinction between secondary ideas and secondary impressions – passions – that passions do not have content.

[47] Ibid. pp. 46-8.

[48] Ibid. p. 48

[49] Ibid. pp. 47-8

desire. Let us call this the 'propositional view' of the content of desire.[50] The propositional view has come under attack in recent years.[51] The attack notes that the propositional view depends upon a mistaken view of the grammar of desire attributions. To begin with, it is pointed out that the most natural way to ascribe a desire to someone is by a sentence of the form '*A* desires to φ', a sentence which takes an infinitival phrase to give the content of the desire, rather than a propositional 'that'-clause.

The opponents of the propositional view point out that even if one resists this more natural infinitival construction, and instead insists on focusing on a construction which takes a 'that'-clause (perhaps because one takes oneself to have grounds for thinking that the latter gives the "true form", and is more metaphysically revealing),[52] the grammar of such a construction does not take a complete sentence. As Maria Alvarez notes, a desire ascription with a propositional 'that'-clause is ungrammatical: we get, e.g., '*A* desires that she goes to the cinema.' To get a grammatical construction, we need to complete the 'that'-clause with a subjunctive: '*A* desires that she *go* to the cinema.' But, since the phrase 'she go to the cinema' is not a grammatical complete sentence, it is unsuitable, she claims, to express a proposition; the 'that'-clause in a desire ascription, therefore, is not a propositional clause.[53]

We have, then, three proposed ways to specify the content of a desire: (1) the propositional view, which is strictly ungrammatical, (2) what we may call the 'infinitival view', and (3) the 'subjunctive view'. The latter two are natural allies, since they are united in their rejection of the propositional view.[54] I will return to these conceptions of how to specify the content of desires in Chapter 2. For now, the important point is that each of these views rejects Hume's conception of desire as lacking representational content.

---

[50] The propositional view is closely related to the widespread view (deriving from Bertrand Russell, 'The Philosophy of Logical Atomism', p. 218, and *An Inquiry into Meaning and Truth*, p. 65) that since desires are psychological attitudes, they are therefore propositional attitudes.

[51] Hanoch Ben-Yami ('Against Characterising Mental States As Propositional Attitudes') denies not just that desires, but that psychological attitudes generally, are propositional attitudes.

[52] For criticism of this idea see ibid. pp. 86-8

[53] Maria Alvarez, *Kinds of Reasons*, p. 67 (the example is hers). Cf. John Hyman, *Action, Knowledge, and Will*, pp. 107-10

[54] These have affinities with John Bricke's interpretation of Hume according to which, while desires have representational content, they are not representations, because they cannot be true or false (*Mind and Morality: An Examination of Hume's Moral Psychology*, p. 25). The question then will be what justifies the view that desires are not true or false.

1.4 Desires, Motivation, and Reasons

Hume offers a theory of motivation, rather than a theory of reasons; it is a theory of the psychological processes which are or would be involved in motivating one to act, not a theory of what would, or does, favour or justify actions. Strictly speaking, according to Hume, there can be no practical reasons at all,[55] for practical reasons would be something in terms of which actions might be rationally evaluated. Since actions do not themselves have representational content, it follows, on Hume's view, that they are not rationally evaluable. Similarly, it is plausible to think that if there were practical reasons, they would bear some close relationship to reason, and they would be the grounds of actions that one does for those reasons. Since reason, according to Hume, is unable to play any role in grounding actions, this would seem to rule out there being practical reasons.

But even contemporary Humeans take Hume's view here to be far too strong; actions are rationally evaluable, and there can be numerous sorts of practical reasons, for instance ethical, prudential, economic, aesthetic, political, or pedagogic. We can explain someone's doing something, and often, when it was something they did intentionally, we are able to explain their doing it by appeal to the very considerations in the light of which they did it. It is natural to put these points by saying that we can give reasons for why someone did something, and very often these are the very reasons for which they did it – what we mention are genuine practical reasons. Similarly, it is plausible to take deliberation to be, among other things, a process of considering what reasons one has for, or against, doing one thing or another, and giving or seeking advice to be practices which contribute to people's knowing what reasons they have, and, hopefully, acting accordingly. Contemporary Humean views, however, while accepting that there may be reasons for action, take it that Hume was nevertheless right to stress the fundamental role of desire in the theory of motivation. The lesson to take from Hume, they think, is that desire must be similarly fundamental in an account of practical reasons, because of certain putative *a priori* constraints on them. As Warren Quinn says of morality in particular, contemporary 'subjectivist' Humeans 'have extended Hume's idea that morality produces motives only through its noncognitive content to the idea

---

[55] Broome, 'Reasons and Motivation', p. 132, makes this point, although he puts it in slightly different terms and focuses on the relation between reason and desire. See also Christine M. Korsgaard, 'The Normativity of Instrumental Reason', p. 221-2, and Nagel, *The Possibility of Altruism*, p. 64.

that it produces reasons only in the same way.'[56] But Quinn's point is generalisable to all spheres of human agency.

The feature which contemporary Humeans take to be most important to preserve from Hume is what I will refer to as the Exclusive Efficacy of Desire thesis:

> (EED) Desires, and only desires, are motivationally efficacious psychological states or attitudes.[57]

This is supplemented by what I shall call the Explanatory Condition:

> (EC) R is a reason for A to φ only if (i) A could be motivated to φ for R, and (ii) were A to φ for R, R would figure in a true explanation of A's φ-ing.[58]

As Thomas Scanlon explains:

> Desires are commonly understood in philosophical discussion to be psychological states which play two fundamental roles. On the one hand, they are supposed to be motivationally efficacious: desires are usually, or perhaps always, what move us to act. On the other hand, they are supposed to be normatively significant: when someone has a reason […] to do something this is generally, perhaps even always, true *because* doing this would promote the fulfillment [*sic*] of some desire which the agent has.[59]

In the first half of this passage Scanlon summarises the importance of (EED). The second part, though, concerns what Humeans take to be the fundamental role of desire in the account of practical reasons, which is supposed to explain how (EC) could be met. We will consider this second point presently.

It is important to note, however, that there is an everyday thick sense of 'desire' in which it is all too common for people to be motivated to do things which they have, in this sense, no desire to do. So, for example, it is very natural to say that someone was motivated to go to the dentist, and indeed that they acted on this motivation, despite not

---

[56] Warren Quinn, 'Putting rationality in its place', p. 28
[57] See e.g. Smith, 'The Humean Theory of Motivation', esp. pp. 52-5; Sinhababu, 'The Humean Theory of Motivation Reformulated and Defended', pp. 465-6, 468-9, 488; and Donald C. Hubin, 'What's Special About Humeanism?', pp. 38-40
[58] The *locus classicus* is Bernard Williams, 'Internal and external reasons', p. 102.
[59] Scanlon, *What We Owe To Each Other*, p. 37.

in the least wanting to.[60] On the other hand there is an attenuated sense of 'desire' which is also common in everyday usage, but ubiquitous in philosophical usage, according to which it is *always* true that if someone does something, they want, in *this* sense, to do it.

In this attenuated sense in which one may want or desire something, it is natural to take it that there are various sorts of attitudes which are motivationally efficacious, for which 'want' or 'desire' may be used as umbrella terms. So, for example, Bernard Williams says that what he calls an agent's subjective motivational set, or simply *S*, 'can contain such things as dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent', but he also says that the term 'desire' 'can be used, formally, for all elements in *S*.'[61] 'Desire' here is effectively stipulated to refer to whatever kind of thing might be motivationally efficacious. Similarly, Davidson uses the term 'pro-attitude' in such a stipulative sense to include:

> desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values *in so far as these can be interpreted as attitudes of an agent directed toward actions* of a certain kind. The word 'attitude' does yeoman service here, for it must cover not only permanent character traits that show themselves in a lifetime of behaviour, like love of children or a taste for loud company, but also the most passing fancy that prompts a unique action, like a sudden desire to touch a woman's elbow.[62]

When the Humean position maintains the view that desires, and only desires, are motivationally efficacious, 'desire' must be taken in the second, attenuated, sense. (I will continue to use the term 'desire' rather than the uglier 'pro-attitude'.) Similarly, it must be in this sense that the Humean view that all practical reasons are grounded in the desires of agents must be understood.[63]

The easiest way to accommodate (EED) and (EC) is to construe practical reasons as constituted, at least in part, by desires in this sense. Simple Humean accounts

---

[60] The point is not affected by the observation that there is likely something *else* which they want, and which going to the dentist contributes to, say, having healthy teeth. The point is that, in this sense of desire, it is just not true that they require a desire to φ in order to be motivated to φ.

[61] Williams, 'Internal and external reasons', p. 105

[62] Donald Davidson, 'Actions, Reasons, and Causes', p. 4, emphasis added.

[63] A potentially non-Humean interpretation of the attenuated sense of 'desire' seems to be offered by Nagel, when he claims 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' (*The Possibility of Altruism*, p. 29). For an extended consideration of these two uses of 'desire', see G. F. Schuler, *Desire: Its Role in Practical Reason and the Explanation of Action*.

of reasons of this kind take reasons to be constituted by a belief together with a desire.[64] The belief-desire pairs which constitute reasons, of course, would have to be related in quite specific ways. On a typical version of this theory they must be related such that, roughly, there is a desire that *p* together with a means-end belief that φ-ing would bring it about that *p*, and such a belief-desire pair would constitute a reason for φ-ing. The description of the content of the desire matches the description of the end in the content of the belief, while the description of the means in the content of the belief matches the description under which the action for which it is a reason would count as intentional when performed for that reason.[65] The belief, in other words, connects the action at one end, with the desire at the other. This provides a construal of what it would mean for reason to direct or guide the motivating impulse of a desire. But even on the belief-desire theory of reasons, statements of reasons may in fact only mention one or the other, since the context may make the fuller statements mentioning both, for which they are elliptical, unnecessarily otiose, and because a desire might only operate in the 'background', rather than appearing explicitly in the foreground of deliberation.[66]

The simple belief-desire account of reasons, however, is too crude. As Williams argues, this account takes it that any desire an agent has partly constitutes or gives rise to a reason for them to act. But this is highly implausible when the desire in question is a desire for what is falsely believed to be a means to satisfy a prior desire.[67] Because this view does not discriminate between those desires which are based on true beliefs from those based on false beliefs, both would equally give rise to reasons for an agent to do something. Williams gives the example of someone who wants to drink *that* liquid, believing it to be gin, when in fact it is petrol. It would be odd, he claims, to say that they have a reason to drink it, since there is nothing which really justifies or favours

---

[64] See e.g. Davidson, 'Actions, Reasons and Causes', esp. pp. 3-5, and also Smith, *The Moral Problem*, ch. 4 & 5. Smith, however, only adopts such a belief-desire theory of *motivating* reasons, while claiming that *normative* reasons are facts, thus imposing a fundamental *ontological* distinction between the sorts of things that could be motivating reasons and normative reasons. Following, among others, Bernard Williams ('Internal reasons and the obscurity of blame', p. 39, and 'Postscript: Some Further Notes on Internal and External Reasons', p. 93), Dancy (*Practical Reality*, pp. 2, 6, 103) and Alvarez (*Kinds of Reasons*, pp. 32-3), I reject an ontological distinction between motivating and normative reasons, instead taking it that reasons in general are facts, which can be characterised as motivating, normative, or explanatory depending, roughly, on the context and our explanatory interests.

[65] This needn't depend on the common assumption that intentional actions are actions performed for reasons, but only the weaker view that actions performed for reasons are intentional. The stronger assumption is rejected by, e.g., Robert Brandom, 'Actions, Norms, and Practical Reasoning', esp. p. 478; and Alvarez, *Kinds of Reasons*, p. 103.

[66] See Davidson, 'Actions, Reasons and Causes', pp. 6-7; Philip Pettit, 'Humeans, Anti-Humeans, and Motivation', p. 530, and Philip Pettit and Michael Smith, 'Backgrounding Desire'.

[67] Compare the loose sense in which Hume agrees that desires may be contrary to reason discussed in fn. 29 above.

drinking it. Yet in spite of this we can perfectly well explain their drinking it by reference to their desire and false belief. To think that our being able to explain this warrants taking them to have had a reason to drink the petrol, though, would imply that our conception of reasons is 'only concerned with explanation, and not at all with the agent's rationality,' when it *should* be concerned with their rationality.[68] Williams makes it plausible that while a concern with explanation need not involve any normative notion such as justification, such a notion is essential to a concern with rationality, and so to rational explanation and reasons.[69] The following, therefore, is taken as a third *a priori* condition on reasons:

> (RN)   Reasons are essentially normative.

But while this means that the simple belief-desire theory of reasons is inadequate, Williams argues in an influential series of papers that (EED), (EC) and (RN) support the closely related view that all reasons must be 'internal' rather than 'external'.[70] In the next section, we will look more closely at this view.

1.5 Internal Reasons, External Reasons, and Value

An internal reason, in Williams' sense, is one which is conceptually dependent on the agent's having some desire to which φ-ing is relevant, while an external reason would be one that was independent of any such relation to the agent's actual desires. Williams' thesis that all reasons are internal, then, is the view that all reasons for an agent *A* to φ are constrained by their actual desires. The particular constraint that an agent's desires place on the reasons they have, according to Williams, is that the agent could come to accept that they had those reasons through a process of 'sound deliberation' – i.e. rational deliberation where erroneous beliefs have been corrected and the agent is fully informed – starting from their actual desires.

The interpretation of Williams' argument that all reasons are internal is controversial, and I will not take a stand here on exactly what argument should be

---

[68] Williams, 'Internal and external reasons', p. 102-3
[69] Cf. Williams, 'Internal reasons and the obscurity of blame', p. 36
[70] Bernard Williams, 'Internal and external reasons'; 'Internal reasons and the obscurity of blame'; 'Replies', esp. pp. 186-194, and 'Postscript', pp. 91-7. Strictly, his claim is that the they support the view that for a reason statement of the form '*A* has a reason to φ' or 'There is a reason for *A* to φ' to be true, it must be interpreted as having an 'internal', as opposed to 'external', sense.

attributed to Williams himself.[71] What is clear is that the position he reaches takes it to be essential to reasons for someone to φ that they could figure in true explanations of their φ-ing, and that they have normative force.[72] Williams' argument has most commonly been interpreted as an Humean argument;[73] Stephen Finlay reconstructs the argument on this interpretation thus:

> SA1. R is a reason for A to φ only if A would, if ideally rational, be motivated to φ by believing R;
> SA2. A would, if ideally rational, be motivated to φ by believing R only if A has a desire-like state that could bring A to be motivated to φ by believing R;
> SA3. If R is an external reason for A to φ, then A does not have a desire-like state that could bring A to be motivated to φ by believing R;
> SA4. Therefore there are no external reasons.[74]

The position that is assumed in this argument is broadly Humean in character. It sees motivation as necessarily provided by, and only by, one's desires,[75] and it takes a procedural and instrumental conception of practical reason such that, as Finlay puts it, it 'merely facilitates flow of motivational force from pre-existing desires, channeling it from one object to another.'[76] Both of these features of the position are recognisably descended from Hume's philosophical psychology. But now, if (EC) requires an appeal to what motivates the agent, but rational deliberation could not produce a new motivation without it being derived from a prior desire, then the hypothesis that something could be a reason for an agent independently of any desires they actually have would rule out its satisfying (EC).

There are two things going on here. First, there is an assumption that the

---

[71] For a good critical overview of the vast literature on Williams' position see Stephen Finlay, 'The Obscurity of Internal Reasons'. For an extensive taxonomy and critical discussion of different possible positions in the internalism/externalism debate, see Parfit, 'Reasons and Motivation'.

[72] Williams, 'Internal and external reasons', pp. 102-3, and 'Internal reasons and the obscurity of blame', p. 36

[73] Although Bernard Williams' comment in *Ethics and the Limits of Philosophy*, p. 246 fn. 19, that 'I do not believe that there can be an absolutely "external" reason for action, one that does not speak to any motivation the agent already has (as I have stressed, Kant did not think so either)', suggests that he thought that the internal reasons theory is neutral between Humeanism and Kantianism.

[74] Finlay, 'The Obscurity of Internal Reasons', pp. 3-4. This reconstruction (which does not give Finlay's own interpretation) fails to account for the difference, roughly, between a reason that one is unmotivated by, and a reason which is *independent* of such motivation.

[75] Cf. Williams, 'Internal and external reasons', p. 107. My discussion ignores the distinction that Dancy (*Practical Reality*, pp. 17-9) draws between internalism about reasons and desire-based accounts of reasons.

[76] Finlay, 'The Obscurity of Internal Reasons', p. 4. Applying the same idea to rationalising psychological explanations, R. Jay Wallace, 'How to Argue about Practical Reason', p. 370, calls this the 'desire in-desire out principle'. Cf. Williams, 'Internal and external reasons', p. 109, and Sinhababu, 'The Humean Theory of Motivation Reformulated and Defended', pp. 465-6. See Parfit, 'Reasons and Motivation', for discussion of the distinction between procedural and substantive conceptions of rationality.

fundamental basis of motivation is a non-cognitive ingredient, so that even though the desires that one acts on might be in some minimal sense rationally grounded, such rational grounding is ultimately based on desires which are beyond the scope of rational evaluation, but rather are simply given.[77] Second, there is the idea that in order for a reason *R* to be normative, i.e. to go beyond claims about what an agent is presently disposed to do, such that it may play 'an important part in discussions about what people should become disposed to do,'[78] it must stand in Williams' deliberative relation to one's actual desires. The argument therefore assumes that whatever normativity reasons have is due to their being instrumentally valuable under conditions of sound deliberation.[79]

Some philosophers interpreting Williams' argument along these lines have accused him of begging the question against the external reasons theorist either by assuming a partisan theory of motivation, or of practical reason, or both.[80] But Williams may also make another question-begging assumption, ruling out in advance the possibility of a substantively realist account of value. This would explain Williams' complaint that it is obscure what the external reasons theorist could possibly *mean* by claiming that someone had an external reason to do something, such that an external reasons statement could meet two conditions: (1) that an agent who is not already motivated to act for this reason gains such a motivation when they accept the external reasons statement, and (2) that this new motivation that they gain is rationally arrived at, via a sound deliberative route from their prior motivations. The only possible answer, he thinks, will be unhelpful:

> *What* is it that one comes to believe when he comes to believe that there is a reason for him to φ, if it is not the proposition, or something that entails the proposition, that if he deliberated rationally, he would be motivated to act appropriately? We were asking how any true proposition could have that content; it cannot help, in answering that, to appeal to a supposed desire which is activated by a belief which has that very content.[81]

But suppose that doing something was *itself* in some way good or valuable, or

---

[77] Cf. Korsgaard, 'Skepticism about practical reason', p. 314

[78] Williams, 'Internal reasons and the obscurity of blame', p. 36

[79] Thus Williams, 'Postscript', pp. 92-3, asserts that an act's 'making normative sense to him implies that it made normative sense in terms of his S.' 'Instrumental' here would have to be construed in the widest possible way, so as to include constitutive means, and to accommodate Williams' rich conception of deliberation in, e.g., 'Internal and external reasons', pp. 104-5; see also Hubin, 'What's Special About Humeanism?', p. 32; and Korsgaard, 'The Normativity of Instrumental Reason', pp. 215-6.

[80] For a list of some examples see Finlay, 'The Obscurity of Internal Reasons', p. 4

[81] Williams, 'Internal and external reasons', pp. 109-10

was instrumentally good or valuable for something else (unrelated to any actual desires of the agent) which was itself good or valuable, and that this value was metaphysically real.[82] A *realist* external reasons theorist might think that this entailed that there was a reason for them to do it; moreover, they might think that this reason could explain their doing it if they believed or knew that doing it would have this value, and that having such a belief would be rational, for instance, if they could *tell* that doing it would have this value. So there would be no mystery about the content of the belief one would need to acquire in order to be motivated to act on an external reason, and if doing things could *really* be valuable in this way, this could evidently meet the conditions that Williams says the external reasons theorist needs: 'that the agent should acquire the motivation *because* he comes to believe the reason statement, and that he should do the latter, moreover, because, in some way, he is considering the matter aright.'[83]

If evaluative realism were seen by Williams as a possibility at all, this kind of position would seem to give a quite natural response to his arguments against external reasons. But if Williams excludes the possibility of such a position in advance, this is another way in which his argument would beg the question.

Williams denies that even the applicability of thick ethical concepts – concepts such as *courage*, *cruelty*, or *recklessness* – would provide agents with external reasons, saying that the true application of such a concept does not mean 'that another agent who does not use the concept has a reason to avoid or pursue certain courses of action in virtue of that concept's application.'[84] Seeing evaluative realism as a possibility would allow us to avoid obscurity when we say that claiming that someone has a reason to behave otherwise goes beyond the description of someone as, for instance, acting inconsiderately.[85] We would be claiming that there is some value which provides normative force to the claim that they should act otherwise. Moreover, the fact that someone is acting inconsiderately would show that they have a flawed perspective on

---

[82] In saying that something is itself good or valuable, I mean to deny that it is merely instrumentally good or valuable, but to remain neutral on whether it is intrinsically or extrinsically good. Cf. Christine M. Korsgaard, 'Two Distinctions in Goodness'. I think that Korsgaard's two distinctions are inadequate, and that we must make a third distinction between primary or original value and secondary or derived value. Jonathan Dancy, *Ethics Without Principles*, ch. 9, makes a similar suggestion.

[83] Williams, 'Internal and external reasons', pp. 108-9. Something like the position I have described here seems to be what John McDowell advocates in 'Might There be External Reasons?', although one might, as I do, have doubts about the extent to which McDowell's is a realist position. See also Joseph Raz, 'Agency, Reason, and the Good'.

[84] Williams, 'Internal reasons and the obscurity of blame', p. 37-38. I discuss thick concepts in more detail in Chapter 4.

[85] Cf. Parfit, 'Reasons and Motivation', p. 109

what is (really) of value, whether because they have a false conception of what is valuable, or, for example, they do not realise that what they are doing is inconsiderate,[86] or both.[87] If this were right, there would be some value which would ground the normative force which Williams rightly thinks is essential to reasons, without this being grounded in the agent's prior desires.[88]

So far, I have said very little about how it could be possible for external reasons to meet (EC). Much of the force of Williams' argument against external reasons comes from the thought that doing so would be problematic for external reasons since they would, by definition, be independent of agents' existing motivations, but for an agent to φ for a reason, they would have to be motivated to φ in some way that was appropriately related to that reason. As he says:

> Internalist theory explains how it is that the agent's accepting the truth of 'There is a reason for you to φ' could lead to his so acting, and the reason would thus explain the action. It is obvious on the internalist view how this works. But suppose we take the externalist view, and so accept that it can be true that A has a reason to φ without there being any shadow or trace of that presently in his [subjective motivational set]. What is it the agent comes to believe when he comes to believe he has a reason to φ? If he becomes persuaded of this supposedly external truth, so that the reason does then enter his [subjective motivational set], what is it that he has come to believe? This question presents a challenge to the externalist theorist.[89]

Williams is *not* claiming here that external reasons are problematic because it would be impossible to be motivated to act for them, on the grounds that if one had such a motivation then it would turn out to have been an internal reason all along. An external reason to φ would not be a reason which had no related motivation, but rather a reason which was *independent* of any such motivation.[90] It might still be that one was

---

[86] This should not be thought to entail that this person would then have an internal reason to be considerate. It might be that they have no prior desire which would ground, through a sound deliberative route, an internal reason to be considerate, whether or not they would be motivated to behave otherwise were they to realise that how they are behaving is inconsiderate; cf. Williams, 'Internal and external reasons', pp. 107-8

[87] Cf. David Wiggins, 'Deliberation and Practical Reason', p. 281; Rosalind Hursthouse, 'Virtue Theory and Abortion', pp. 227-8; and Aristotle, *The Nicomachean Ethics*, 1142b15-20. Another possibility would be that they were weak-willed.

[88] Scanlon, *Being Realistic About Reasons*, pp. 11-3, suggests that Williams assumes that for a consideration to be a reason for someone, it must be something to which they were already committed to seeing as a reason, such that rejecting it as a reason for them amounted to their being self-contradictory or irrational. This, Scanlon suggests, is why his discussion imagines contexts in which the constraints on what counts as a reason are revealed by what one person could say to another to *convince* them that something is a reason for them. But as he says, 'what it takes for a claim to be correct need not be the same as what it takes for the claim to be one that one's opponent in argument cannot consistently deny.'

[89] Williams, 'Internal reasons and the obscurity of blame', p. 39

[90] Williams, 'Internal and external reasons', p. 101; cf. Finlay, 'The Obscurity of Internal Reasons', p. 17

motivated in such a way that the supposed external reason was *also* internal. If we accepted that it was possible for someone to act for an external reason, so that in such circumstances the reason would explain their action, they would have an appropriate motivation (or, it would become part of their motivational set).[91] But Williams cannot see how, starting from a position in which one did not have such a motivation, but in which there was an external reason for one to φ, gaining a belief that one has an external reason to φ could both be *true* and endow one with such a motivation.[92] Williams thinks that the content of this belief must render it intelligible that one acquires the motivation, and provide an understanding of *what that motivation is*. Internal reasons, he supposes, have an advantage here, since their being, or becoming, an element in an agent's motivational set fits the familiar Humean view about how desire extends itself through instrumental reasoning. If it were obvious how this worked, as Williams takes it to be, then it would be obvious how an internal reason could explain one's action.

In Chapter 2, however, I will argue that the way that Humeans typically think about the motivational efficacy of desire is untenable, so internalism is in no better position to accommodate the motivational and explanatory capacity of reasons, in terms of their relation to an agent's existing desires, than is externalism, without requiring such a relation. In Chapter 3, I will begin to argue that the best way of understanding the nature of desire actually provides an explanation of how *both* internal and external reasons could meet the explanatory condition; my account, though, will not fully emerge until Chapter 5.

1.6 Efficacy, Explanation, and Arbitrariness

We have seen that contemporary Humeans rightly take (EC) as an *a priori* constraint on reasons, and, because they maintain (EED), they also take it that reasons must be constrained by the actual desires that agents have. In other words, they take it that:

> (RDB) Reasons are *desire-based*.

I indicated already that I take (EC) to be a genuine constraint, which is neutral between Humean and anti-Humean theories. I also take it that (RDB) is a necessary feature of an Humean theory of reasons, which is supposed to rest jointly on (EC),

---

[91] Williams, 'Internal and external reasons', pp. 107-8
[92] Cf. Finlay, 'The Obscurity of Internal Reasons', pp. 17-8

(EED), and a further necessary feature of an Humean theory. This further necessary feature of an Humean theory is that it conceives of desires as non-cognitive psychological attitudes which are not, fundamentally, rationally-evaluable, thereby upholding a fundamental contrast between desire, and cognitive and rationally evaluable psychological attitudes such as belief and knowledge. We can formulate this necessary principle of an Humean theory, that desire is fundamentally arbitrary, thus:

(FAD) Fundamentally, at least, desire is not rationally evaluable.

In Chapter 2, I will argue that (FAD), and therefore also (RDB), are baseless.

As we saw in §1.3, the attenuated sense of 'desire' allows that it may be stipulated to include whatever is in fact motivationally efficacious, and (EED), therefore, is also neutral between Humean and anti-Humean theories; in particular, it is neutral with respect to (RDB). The significance of this is that (RDB) rests in part on the Humeans' account of desire itself, that is, on what constraints their theory puts on what sorts of things could be motivationally efficacious, and so what is included under the attenuated sense of 'desire'. (FAD) provides an abstract and general characterisation of these constraints, and is therefore the most fundamental principle of the Humean theory.

It is of particular importance to the Humean theory that cognitive, rationally-evaluable psychological attitudes like belief or knowledge, are not allowed to count as desires, in the attenuated sense, since if they did there would be no fundamental contrast between belief and desire such as Hume takes there to be, and no space for a contemporary analogue of this which takes desire, but *not belief*, to be fundamental in an account of reasons. It is just this that (FAD) ensures, since to count rationally evaluable attitudes such as belief as desires would flout this principle. The acceptance of (FAD), in other words, is what results in a distinctively Humean interpretation of (EED).

Furthermore, if Humeans were to allow that desires were rationally evaluable, they would face a trilemma. Either they would have to maintain (i) that rational evaluation of desires is ultimately grounded in reasons which are not desire-based, (ii) that every desire is rationally evaluable, but only in terms of reasons which are themselves desire-based – reasons that are based on desires other than the one that is evaluable in terms of them, or (iii) that the rational evaluation of desires is infinitely regressive. Hume himself rejects (iii) out of hand, and he is surely right that it is not

plausible.[93] Accepting (ii) would mean implausibly taking it that the rational evaluation of desire is, as it were, lifted up by its own bootstraps. No reason for any desire would be independent of some other desire, and reasons for the other desires might eventually be based on the first desire. Accepting (i) would rule out any kind of unified account of the nature of reasons. While this is not a decisive objection, and it is often taken to be plausible that, e.g., practical and epistemic reasons are fundamentally different in kind, this view could not even allow that reasons for desire were themselves practical reasons, which is a more dubious position.[94] The trilemma is not decisive, but certainly indicates the *prima facie* implausibility of a Humean view which allowed that *every* desire that someone had was rationally evaluable.[95]

In Hume, the grounds for accepting (FAD) and (EED), and so for thinking that desires, but not beliefs, are motivating psychological states is that he thinks of them as original existences, and therefore as mere impulses which, as it were, simply push one around. Only an original existence, Hume claims, could be 'active' in this way, and beliefs are not original existences. In §1.3, however, we found that Hume's view about the nature of desires is false; desires, as well as beliefs, have content, and they are not mere impulses. If this is right, though, then the first necessary condition for an Humean theory of *reasons*, that it take reasons to be desire-based, has lost its footing. Hume's grounds for holding that desire, *but not belief,* is fundamental to motivation cannot be the Humean's grounds for that view, and so cannot be the Humean's grounds for thinking that they are fundamental to reasons either.

As for (FAD) itself, in Hume's view this rests on the account of desires as original existences, together with the view that only what is representational admits of rational evaluation.[96] Not only did we see in §1.3 that desires have content, but we also saw in §1.4 that, in admitting that there can be practical reasons at all, contemporary Humeans are committed to the view that at least one sort of thing – action – is rationally

[93] David Hume, *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, p. 293

[94] Cf. Nagel, *The Possibility of Altruism,* 'if the desire is a motivated one, the explanation of it will be the same as the explanation of his pursuit' (p. 29), and 'The reasons for a practical judgement are reasons for *doing* or *wanting* something, and not just reasons for believing something' (p. 63). Wallace ('How to Argue about Practical Reason', p. 374) uses the point that epistemic reasons are not plausibly thought of as desire-based to press an argument against the Humean theory as a general theory of reasons.

[95] See Mark Schroeder, *Slaves of the Passions*, esp. ch. 10, for a valiant attempt to defend such a view. For reasons closely related to this, it is unclear that Schroeder's position is best characterised, as he characterises it himself, as an Humean view; see David Enoch, 'On Mark Schroeder's Hypotheticalism: A Critical Notice of *Slaves of the Passions*', pp. 438-40, 444-5.

[96] Hume, *Treatise*, III.i.1, p. 458

evaluable, despite not itself being representational. It is just that, for the Humean, once the rational evaluation of action reaches whatever prior desires an agent in fact happened to have, there is no possibility of any more fundamental rational evaluation. But if Hume's grounds for disallowing the legitimacy of rational evaluation at that point are unavailable to the Humean, what justification can they offer for (FAD)?

We have, then, two problems which contemporary Humeans must address:

(P1) What could ground the Humeans' acceptance of (EED), interpreted so as to support (RDB)?

(P2) What could ground the Humeans' acceptance of (FAD)?

So long as 'desire' in (EED) and in (FAD) is interpreted in the same way, a satisfactory answer to (P2) would also provide an answer to (P1). For if the Humeans could justify their acceptance of (FAD), then (FAD) would, as we saw, justify an interpretation of (EED) which, together with (EC), would support (RDB). But there could, perhaps, be a weaker Humean strategy to support (RDB), which did not go via (FAD). Pursuing such a weaker strategy, the Humean would seek to justify an account of desire which ruled out an interpretation of (EED) which permitted attitudes such as belief to count as desires, in the relevant attenuated sense. On its own, such a strategy would leave the Humean facing the trilemma raised earlier in this section, but, as I conceded, the trilemma is not decisive. Alternatively, such a strategy might be combined with an effort to justify (FAD), thereby yielding a stronger version of the Humean theory. In Chapter 2 we shall consider some plausible and prominent ways of carrying out the weaker and the stronger strategies for answering (P1) and/or (P2). I will argue that they are unsuccessful.

1.7 Conclusion: Summary and Preview

In considering how contemporary Humeans, in correcting for Hume's mistakes, move a fair distance from Hume's philosophical and moral psychology, we have found that the core features of the Humean theory have lost their original motivation and justification which they had in the context of Hume's own theory. The major steps in this journey away from Hume are the contemporary Humeans' rejection first of his account of

desires as original existences, then of his denial that there are practical reasons at all, then of the thesis that practical reasons are constituted by our desires (and beliefs). They arrive at the view that reasons, since they are normative and subject to (RN), are facts, but are nevertheless desire-based, since something is a reason only in virtue of standing in an appropriate relation to the arbitrary desires that an agent actually happens to have. The core Humean theses that desires are arbitrary (FAD), and that reasons are desire-based (RDB), however, appear as unjustified and unmotivated dogmas.

In Chapter 2, we shall consider two kinds of strategy for refilling the hollow shell of the Humean theory, providing new motivation and justification to sustain the Humean theory of reasons, and so responding to (P1) and (P2). In the first half of the chapter, we will consider three arguments, all related to the anti-propositional views of the content of desire which we encountered at the end of §1.3, for the view that desires are non-cognitive attitudes. If sound, these arguments would show that attitudes like belief could not count as desires, but I shall argue that they are in fact unsound. Furthermore, I shall propose an explanation for why the contents of desires are typically specified with an infinitival phrase, or else a subjunctive clause, which does not require that they are non-cognitive, although I also suggest that it is theoretically expedient, and philosophically harmless, to treat desires as if they took a propositional content-clause.

In the second half of the chapter, I will consider a second, stronger strategy for answering (P1) and (P2), by appealing to a supposed distinction in the 'direction of fit' of desire and of belief. I argue that neither of the main Humean accounts of direction of fit are coherent, and cannot bear the justificatory weight that the Humean requires such an account to carry. (FAD) and (RDB), therefore, remain unmotivated and unjustified.

Chapter 3 begins the task of developing a philosophical and moral psychology suitable to ground a realist account of practical reasons. It begins with an examination of Kant's moral psychology, arguing that it is in relevant respects similar to Hume's, and that it is due to his acceptance of characteristically Humean presuppositions that Kant is forced into a rationalist framework to avoid Hume's conclusions. But what Kant is trying to do can also be achieved by denying both the Humean and the rationalist presuppositions that Kant accepts.

The most important feature of Kant's view to take forward is his insistence that there can be fundamentally rationally evaluable incentives for action. This need not be construed, as Kant construes it, as being a matter of its being possible for incentives to be evaluable in terms of the structure of reason. We can instead take the rational

evaluation of motivating attitudes to be based in objective facts, and this permits us to treat desires generally as rationally evaluable. On the basis of this, I argue for normativist conceptions of belief and desire, and argue that it is an implication of such conceptions that desires are cognitive attitudes. In fact, desires are a certain kind of evaluative belief.

Chapter 4 offers an account of reasons and an account of actions. It begins by arguing that what I take to be the most important objections to a realist account of value are grounded in the very conception of desire which Part I argued the Humeans do not succeed in justifying, and which has been rejected and supplanted by the alternative proposed in Chapter 3. Chapter 4 goes on to argue that an account of reasons, since they must be potentially explanatory, should see them as standing in a certain relation, though not a dependence relation, to desires. I offer a proposal for what sorts of things reasons should be taken to be which fits this proposed conception of the relation between reasons and desires. In order to complete an account of reasons which accommodates their being potentially explanatory, we also need an account of the relation between psychological attitudes and actions, and of action itself. I argue that the standard picture of agency, which incorporates a causal theory of action, is unacceptable, and suggest that we accept a framework for an alternative Anscombean conception of actions found in the work of Anton Ford.

Chapter 5 argues that we should also reject a causal conception of action explanation, and that the influential argument given by Davidson for a causal conception does not provide strong grounds to resist this, but rather imposes a condition of adequacy on the non-causal conception we should give. Additionally, it argues that a version of the view that explanation in terms of knowledge is transparent to explanation by the facts allows reasons the explanatory capacity we require of them, but only if a causal conception of psychological explanation is rejected, and explanation in terms of knowledge and explanation in terms of the facts are not taken to be equivalent. It proceeds to develop a non-causal account of rational explanation capable of answering Davidson's challenge, which spells out the explanatory roles of psychological attitudes and of reasons. The proposal raises questions about the role of an agent's character in the explanation of action, and I argue that it plays second-order and adverbial explanatory roles. Furthermore, I explain how this allows us to accommodate phenomena such as *akrasia* and *accidie*, which are often thought to be especially problematic for anti-Humean views, and seeing how the account deals with these

phenomena clarifies and highlights the distinctiveness of the non-causal conception of rational explanation that I have defended.

Chapter 2

Humean Desire?

Two Unsuccessful Justificatory Strategies

But it is a curious fact that the proponents of the contemporary Humean programme [...] having abandoned the empiricist theory of thought that underwrites [it] provide precious little by way of motivation for the view. One is left wondering whether the Humean view should have been given up with the theory that inspired it, an antiquated relic.

~ Fraser MacBride[1]

2.1 Introduction

Recall the following principles from Chapter 1:

(EED) Desires, and only desires, are motivationally efficacious psychological states or attitudes.

(RDB) Reasons are *desire-based*.

(FAD) Fundamentally, at least, desire is not rationally evaluable.

In §1.6, I argued that contemporary Humeans, having abandoned Hume's theory of thought, face two problems:

(P1) What could ground the Humeans' acceptance of (EED), interpreted so as to support (RDB)?

(P2) What could ground the Humeans' acceptance of (FAD)?

In this chapter, I consider two strategies for answering these problems. The first, weaker strategy, attempts to justify an account of desire which would justify an interpretation of (EED) which ruled out attitudes such as belief counting as desires, in the relevant attenuated sense. Without the addition of an answer to (P2) this weaker

---

[1] Fraser MacBride, 'Lewis's Animadversions on the Truthmaker Principle', p. 127

strategy would face the (non-decisive) trilemma which I raised in §1.6, but it might be combined with such an answer if one were forthcoming. A natural thought which spurs this strategy is that there seems to be an obvious fundamental difference between desires and beliefs, namely that desires are non-cognitive attitudes while beliefs are cognitive. What might seem to make this obvious is that the content of beliefs is specified with a propositional clause which is *truth-apt*, while the content of desires is specified with an infinitival phrase or a subjunctive clause, which are not truth-apt (§1.3). But just what the connection is supposed to be between these grammatical considerations and the supposed contrast between the cognitive and non-cognitive natures of beliefs and desires is unclear, and needs to be supported by argument if it is to justify the required interpretation of (EED).

In §2.2, I consider a direct argument from these grammatical features of desire attributions to the view that desires are non-cognitive. If successful, such an argument would rule out cognitive attitudes like belief being counted as desires, in the attenuated sense, thereby supporting the distinctive Humean interpretation of (EED). In §2.3, I consider another argument, found in Alvarez's work, for the view that at least *some* desires are non-cognitive, and which relates this conclusion to the grammatical features of desire attributions. In §2.4, I consider a third argument, given by John Hyman, which attempts to explain the grammatical features of desire attributions by conceiving of desire as a special kind of non-cognitive disposition. None of these arguments, I suggest, are sound. Furthermore, we can explain the grammatical features of desire attributions without committing ourselves to a conception of desire as a non-cognitive attitude, as I show in §2.4. In §2.5, I consider the relatively standard practice of portraying the form of desire attributions as if they took a propositional content-clause.

In the second half of the chapter, I consider the most promising way of carrying out the second, stronger strategy for answering both (P1) and (P2). This appeals to the prominent and familiar idea that there is a distinction in respect of the direction of fit of belief and desire. In §2.6, I consider Smith's teleological argument, which forges a connection between the supposed world-to-mind direction of fit of desire and desire being fundamental to motivation and reasons in the way that Humeans require. In §2.7, I consider the origins of the supposed distinction in respect of direction of fit, which is formulated metaphorically and seems to essentially involve a normative dimension. In §2.8, I consider the first of the two main Humean approaches to giving a non-metaphorical account of the distinction: the functionalist approach, and in §2.9, I

consider the second Humean account of the distinction: the higher-order attitude approach. Neither approach, I argue, is viable. This second strategy for answering (P1) and (P2) is, therefore, unsuccessful, leaving us in a position to reject the Humean conception of desire which underpins the view that practical reasons are desire-based.

2.2 Grammar and the Non-Cognitive Account of Desire

In §1.6 we saw that one strategy for grounding the Humean interpretation of (EED) would be to claim that desire is a non-cognitive attitude, and therefore cognitive attitudes like belief or knowledge could not count as desires, in the attenuated sense. In this section, we shall consider a direct argument from the grammar of the specifications of the contents of desire, to the view that they are non-cognitive attitudes.

The following argument may seem to be directly provided by the grammatical features of desire attributions that we encountered in §1.3:

(G1)   The content of a desire is not specifiable with a complete sentence.

(G2)   Only complete sentences are truth-apt.

(G3)   The content of any cognitive attitude is specifiable with a truth-apt sentence.

So,

(G4)   The content of a desire is not the content of a cognitive attitude.

So,

(G5)   Desires are non-cognitive attitudes.

The significant and controversial premise is (G3), for it is (G3) which mediates between the trivial grammatical points and the supposed metaphysical consequences. Notice that (G3) claims that it is a necessary condition for any cognitive attitude that its content is *specifiable* (not *specified*) with a truth-apt sentence. The similar alternative:

(G3*) The content of any cognitive attitude is specified with a truth-apt sentence

is too strong and obviously false. For example, I might say to someone 'I believe you to be an honourable person'. Here, the attitude that is self-ascribed is explicitly a belief,

yet the content of the belief is not specified with a propositional clause, but rather an infinitival phrase. But that hardly shows that this belief is a non-cognitive attitude. Someone who accepts the (G) argument will claim that this is because the belief in question is paraphrasable as 'I believe that you are an honourable person'. This requires thinking that the propositional formulation is somehow more fundamental, and so revealing of the metaphysical nature of belief, but it is in general unclear what justifies thinking that one of two formulations which paraphrase each other is more fundamental in this way.[2] It is not inconceivable that a language might license only, or in the main, the infinitival construction, but this might plausibly be a mere quirk of the language's grammar, rather than being of deep metaphysical significance.[3] What this shows, I think, is that the tight link between the grammatical possibilities and the nature of an attitude asserted by (G3) is neither obvious nor uncontroversial, and at the very least should not be accepted without argument.

(G3) is broadly consonant with Hume's definition of reason. According to Hume, we saw, reason is the power which is concerned with the discovery of truth and falsity, and cognitive attitudes, like judgements or beliefs, belong to reason, precisely because they are exercises of a psychological faculty in pursuit of that concern. But it is not clear that the conclusion (G5) should follow, given other Humean commitments, even if (G3) is accepted. For it seems that, given these other commitments, (G3) casts doubt on (G1).

To see this, consider that according to Hume, what looks like a moral judgment, the content of which is specifiable by use of a complete sentence, is really a disguised passion or desire. One might think that, on the Humean analysis, its being possible to make a desire attribution by attributing what looks like a moral belief shows that it is possible to specify the content of the desire with a complete sentence. It would not strictly follow from this that the content of desire was truth-apt, and so that desire was cognitive, even on the terms of the (G) argument, since (G2) takes the possibility of expression by a complete sentence to be necessary, not sufficient, for truth-aptness. But it is hard to see what grounds one could have to deny that what looked like moral beliefs really were cognitive attitudes with truth-apt contents, apart from a prior commitment to Hume's non-cognitivist analysis of moral judgements. But Hume's non-cognitivist analysis of moral judgments, we saw in §1.2, depends on his psychological argument

---

[2] Ben-Yami, 'Against Characterising Mental States As Propositional Attitudes', p. 86
[3] In Latin it is standard to give belief attributions with an infinitival clause specifying the belief's content.

(HP) against reason providing the ground of any action, so his non-cognitivist analysis of moral judgments is not available to support such a denial. It is, therefore, hard to see why desires being expressible as moral judgements should not show that (G1) is false, unless one begs the question by assuming that moral judgements are also non-cognitive.[4]

What we need to consider, therefore, in evaluating (G3), is possible reasons why attributions of attitudes using certain attitudinal verbs do not admit content-clauses which take complete sentences. If desires did not grammatically admit propositional content-clauses because of the nature of their content, or because of their (non-cognitive) nature, this might also support a non-cognitivist analysis of moral judgment. If, on the other hand, there was a good explanation for why any attitude could not be specified with a construction that took a propositional content-clause, where that explanation had nothing to do with the attitude in question being non-cognitive (e.g. because it was merely to do with the grammar of the particular attitudinal verb used), we would have grounds to reject (G3); we would have severed the proposed link between the grammar and the nature of the attitude. In order to justify (G3), however, we would have to show that, for *each* attitude which is such that it is ungrammatical to specify its content using a complete sentence, this is to be explained by the non-cognitive nature of the attitude. But then we would have independent grounds for thinking that desire in particular is non-cognitive, and the (G) argument would be superfluous.

Two recent works, Alvarez's *Kinds of Reasons*, and Hyman's *Action, Knowledge, and Will*, contain discussions of the grammar of desire specifications not being such as to take propositional content-clauses. Both associate this with accounts of desire as a non-cognitive attitude, so it is here that I will look for arguments in favour of such a view.

2.3 Alvarez on Cognitive Attitudes, Rational Capacities, and Animal Desire

Alvarez gives two arguments which purport to show that desires are not propositional attitudes, but it seems implicit in what she says that she thinks that an attitude can be cognitive only if it is a propositional attitude. Her first argument is quite

---

[4] In Chapter 3, I shall argue along just these lines that desires should be understood as cognitive attitudes, in fact as a kind of evaluative belief.

straightforward. She claims that, because desire specifications have a grammar which requires either an infinitival phrase or subjunctive clause to give their content, the content-giving phrases seem not to express propositions. This is because propositions are essentially truth-evaluable, and infinitival or subjunctive phrases are not, since they are not complete sentences. Since they do not seem to express propositions, desires cannot be propositional attitudes, and therefore cannot be cognitive attitudes.[5] This argument, however, is uncompelling, since nothing is said to support the alleged connection between the grammar and the nature of the associated attitude.

Alvarez's second argument is more substantial, but, I shall argue, ultimately unsuccessful. What she claims is that certain sorts of desires, in particular those that she calls 'bodily appetites,' are possessable both by people and by some non-human animals.[6] There are two significant characteristics which she thinks such desires have which are particularly important for our purposes: one does not have reasons for having such desires (although they are non-rationally explicable), and one does not need the capacity to reason in order to have them.[7] This makes these sorts of desires 'non-rational' desires, but they are not completely beyond the scope of reason. Rather, just like Hume, she thinks that we can reason *about* them, both about how to satisfy them, and about whether to satisfy them. Strictly though, the desires in themselves are non-rational and arbitrary.[8] With these materials, we can reconstruct her second argument for the view that at least *some* desires are non-cognitive:

(A1)   An attitude is cognitive only if it has propositional content.

(A2)   Having an attitude with propositional content requires a capacity to reason.

(A3)   Some sorts of desire may be had both by people and some non-human animals.

(A4)   Non-human animals lack the capacity to reason.

So,

(A5)   Some desires do not have propositional content.

So,

(A6)   Some desires are non-cognitive attitudes.

[5] Alvarez, *Kinds of Reasons*, p. 67
[6] Ibid. p. 73
[7] Ibid. pp. 67, 73-4, 78-9
[8] Ibid. p. 79

This argument is intended to support the view that some desires are non-cognitive without relying on the grammatical data. In particular, (A1) does not link the cognitive or non-cognitive nature of an attitude to grammatical considerations, but to the already determined nature of its content. The rest of the argument then provides non-grammatical grounds for taking the content of some desires to be non-propositional. If successful, that conclusion could then be used to explain the grammatical characteristics of desire specifications. Because it does not rely on the grammatical data, it seems a stronger argument, but its scope is more limited. Alvarez's first argument was supposed to show that desire in general is non-cognitive, but this argument can only purport to show that a much more limited range of desires are non-cognitive, for just the same reason that it is a stronger argument – it does not rely on the grammatical data.

The trouble, though, is that (A2)-(A4) are in tension with each other. Let us first ask *why* (A3) seems plausible. Presumably, the answer must be that the attribution of certain desires to non-human animals plays a role in explaining certain intentional behaviour that they engage in. But in order to play such a role, it is necessary that we also attribute certain beliefs to them, and it is the combination of the animal having certain desires *and* beliefs which provides for the possibility of the relevant explanations of its behaviour. But that just means that the motivation for (A3) equally motivates:

> (A3*) Some sorts of belief may be had both by people and some non-human animals.

and also:

> (A3**)(A3) is true if and only if (A3*) is true.

But with the addition of (A3*), we can construct the following parallel argument:

(A1)   An attitude is cognitive only if it has propositional content.

(A2)   Having an attitude with propositional content requires a capacity to reason.

(A3*) Some sorts of belief may be had both by people and some non-human animals.

(A4)   Non-human animals lack the capacity to reason.

So,

(A5*) Some beliefs do not have propositional content.

So,

(A6*) Some beliefs are non-cognitive attitudes.

But (A6*) is clearly false, so, unless we reject (A1), (A2) or (A4), we have to reject (A5*) and (A3*). But if we reject (A3*), then by (A3**) we have to reject (A3).

Interestingly, the general thought that having certain sorts of attitudes requires the capacity to reason is one that finds its most forceful contemporary expression in the work of John McDowell. McDowell's claim is that making judgements or having beliefs, as well as having genuine perceptual experience, requires possessing conceptual capacities, and possessing conceptual capacities is essential for having the capacity to reason. But non-human animals, he thinks, do not possess such capacities, and so it must be impossible for them to make judgements, have beliefs, or have genuine perceptual experiences. In his terminology animals have a mere sensitivity to an environment, rather than experience of a world, and presumably they at best have quasi-beliefs and make quasi-judgements.[9] So too, on a McDowellian view, one might allow that non-human animals have quasi-desires. But the point is that precisely because they are quasi-desires which do not require possession of conceptual capacities, such sorts of quasi-desires are *not* shared by people: people have the genuine article, because they *do* have conceptual capacities.[10] That is to say that the thought underlying (A2) and (A4) is precisely such as to motivate the rejection of (A3).[11]

There is a potential third argument to be found in Alvarez's discussion. Here is the relevant passage:

---

[9] John McDowell, *Mind and World*

[10] One might reply that although people have genuine desires and beliefs, they may *also* have quasi-desires and beliefs. The response is impotent, though, since it would do nothing to show that any genuine desires were non-cognitive attitudes, which is the point at issue.

[11] I am inclined to accept that animals can have cognitive attitudes, but I take no official stand here on which of (A1), (A2), and (A4), should be rejected.

> In fact, what is distinctive about desires is […] that if one has a desire one has an inclination to *act* in the way that one believes will satisfy the desire [….] In other words, what is distinctive about the concept of desire is its conceptual relation to *agency*, for an important criterion for the attribution of a desire to a person is whether that person is inclined to act so as to satisfy the relevant desire.[12]

There is a long tradition, going back at least to Hume, of thinking that for an attitude to have such a conceptual relation to agency, it must be a non-cognitive attitude. So in making this claim, Alvarez might seem to be giving an argument which implicitly appeals to this tradition in support of desires being non-cognitive. Such an argument, however, would beg the question. I do not think, though, that such an interpretation of Alvarez herself is required, or even most plausible. Whatever the case may be about Alvarez's own view, I am not concerned here with challenging the view that desire has a conceptual relation to agency. Rather, I am concerned with challenging the view that desire is non-cognitive, so that cognitive attitudes like belief and knowledge could not be desires, in the attenuated sense. It would be perfectly consistent to maintain both that desire has a conceptual relation to agency, and that desire is cognitive and rationally evaluable. In fact, this is just what I will claim in Chapter 3. I will, however, return to this claim of Alvarez's presently, for I think it contributes to an alternative explanation of why desires take infinitival or subjunctive content specifications. First, though, I will consider Hyman's discussion of the grammar of desire specifications.

2.4 Hyman on Dispositions and Infinitives

Officially, Hyman acknowledges that he takes for granted a version of the view that desires are non-cognitive, saying that he 'shall assume that a simple and conventional conception of desire is approximately right. A desire is a disposition, in the broad sense of the term that corresponds to Aristotle's *hexis*.'[13] However, he immediately goes on to give a series of quick arguments to the effect that the assumption that desires are dispositions provides good explanations of phenomena that we anyway need to explain, thereby providing indirect justification for his assumption.[14] In particular, he thinks that this conception of desires as non-cognitive (causal) dispositions would explain the grammar of desire-specifications and -attributions: 'the simple conception of desire

---

[12] Alvarez, *Kinds of Reasons*, p. 70
[13] Hyman, *Action, Knowledge, and Will*, p. 107
[14] Ibid. pp. 108-10

proposed here explains why the object of a desire, *like the object of any disposition*, is standardly identified in English with an infinitival clause, and not, as philosophers who believe desires are propositional attitudes would prefer, a sentential clause.'[15] The thought is that simple physical dispositions like the fragility of glass can be specified by use of an infinitival phrase to say what glass is disposed to do in certain circumstances (*ceteris paribus*). So we can say that, e.g., glass is disposed *to shatter* when struck. Then the fact that desires are conceived as dispositions explains the grammar of how their content is specifiable.

This suggests the following argument:

(H1) Any disposition is specifiable with an infinitival phrase

So

(H2) If desires were dispositions, then they would be specifiable with infinitival phrases.

(H3) The contents of desires are specifiable with infinitival phrases.

(H4) If (H2) and (H3), then (H3) is best explained by conceiving of desires as dispositions.

So,

(H5) (H3) is best explained by conceiving of desires as dispositions.

There are a number of objections to this argument. Nothing is said here to support the tight link between the grammatical data in (H1) and (H3) and the alleged nature of desires in (H5). In particular, nothing is said to explain (H1) – no explanation has been offered for why dispositions are specifiable with an infinitival phrase, and, therefore, we have no idea what would be good about the explanation of the grammar of desire specifications in terms of their being conceived as dispositions. That is, we have no reason to accept (H4). Furthermore, it is not clear how neat the analogies between simple physical dispositions and desires, or the grammar of their specifications, really are. Hyman's argument, then, is uncompelling.

Let us work backwards through these objections. Once we have done so, I hope to have assembled the materials to say something about why dispositions are specified

---

[15] Ibid. p. 110, emphasis added.

with infinitives, and also to explain why desires might be specified similarly, while avoiding a commitment to desires being non-cognitive.

First, we should note that Hyman himself acknowledges that there are important differences between simple physical dispositions and the dispositional attitudes which he conceives desires to be:

> The basic difference is that a simple physical disposition is manifested by causing or undergoing change, whereas a desire is manifested in two main ways: first, by purposive or goal-directed behaviour, specifically, behaviour aimed at satisfying the desire—in other words, at getting what it is a desire to have, or doing what it is a desire to do; and second, by feeling glad, pleased, or relieved if the desire is satisfied, and sorry, displeased, or disappointed if it is frustrated.[16]

In short, desires are manifested by certain sorts of rational behaviour – intentional actions of relevant kinds, and intelligible psychological reactions to the consequences of or interferences with such actions. Physical dispositions, though, are not manifested in such ways; they are manifested by the things which have them 'causing or undergoing change'. Of course, this is not surprising but is the obvious and predictable consequence of drawing an analogy between simple physical things and properties, and rational agents and their psychological attitudes. There should be no requirement that analogies be perfect.

But there are concomitant differences. As G. E. M. Anscombe argues, 'the term "intentional" has reference to a *form* of description of events [….] we can speak of the form of description "intentional actions", and of the descriptions which can occur *in* this form, and note that of these some are and some are not dependent on the existence of this form for their own sense.'[17] Her point is that there are certain descriptive terms which can only be used to describe an event as 'voluntary or intentional,'[18] e.g. 'paying' – '*A* is paying off her gambling debts', or, more saliently for our purposes, '*A* desires to pay off her gambling debts'. Such terms can only be used within descriptions of, e.g. what someone does, what they want, or intend, etc. On its own, this claim does not entail a grammatical difference between attributions of desire and physical dispositions. It is plausible, though, that at least some of these descriptive terms will have a distinctive grammar, which will not be explicable merely in terms of their occurring in specifications of desires, even though they occur within infinitival phrases.

---

[16] Ibid. p. 107
[17] G. E. M. Anscombe, *Intention*, §47, pp. 84-5
[18] Ibid. p. 45

If this is right, then the *specific* grammar of desire specifications, at least using such essentially intentional terms within infinitival phrases, would not be fully explicable by the mere fact (if it were one) that desires were dispositions, and so a further explanation would have to be given of their grammar which would explain their distinctive features as well as their occurring within infinitival phrases. But then the explanation in terms of desires being dispositions would seem redundant. The point is not really dependent on Anscombe's notion of the form of intentional description. It would be enough if there were any specific differences between the grammar of desire attributions and the specifications of physical dispositions, since these would need to be explained in terms other than of desires being dispositions.

There is another significant grammatical disanalogy between specifications of physical dispositions and attributions of desires. Desires, we have seen, are attributable by use of a subjunctive clause, but this is not true of physical dispositions. For example, we may say either '*A* desires to drink a cup of tea' or '*A* desires that she **drink** a cup of tea', but while we may say 'glass is disposed to shatter', we cannot also say 'glass is disposed that it shatter'.

What emerges is that, despite giving an argument for the view of desires as dispositions based on the grammar of desire attributions and the comparison with the grammar of specifications of physical dispositions, Hyman gives no serious consideration to the actual linguistic data or linguistic theory. Because of this, he does not even provide a suggestion about what would be good about his proposed explanation of the grammar of desire attributions. Furthermore, because he does not give even cursory consideration to what alternative explanations there might be, he fails to say anything compelling in support of his claim that the account of desires as dispositions provides the *best* explanation of the grammatical data. As I will now argue, however, an engagement with elementary linguistic theory suggests a quite different explanation. Whether precisely this alternative is in the end correct need not be decided here; it is enough if it provides a more plausible and deeper explanation than that suggested by Hyman.

To begin, let us ask: Why is it that physical dispositions are specifiable with an infinitival phrase? To answer this fully, we would need a detailed linguistic analysis of infinitives, and a worked out theory of their use and semantics. I shall make no attempt to provide these here, but will only make some preliminary comments.

Infinitival phrases consist of the infinitive particle 'to' followed by a verb or

verb phrase. In linguistics terminology, the infinitive particle 'to' is the head of the resultant phrase, which is what makes it an infinitival phrase.[19] However, as Radford *et al.* suggest, while the traditional label *infinitive particle* suggests that this 'to' is a kind of word unrelated to any other kind of word in English, this is actually quite implausible. As they say, 'In some respects, infinitival *to* seems to resemble an auxiliary like *will*, in that both are typically used in a clause with future time reference'.[20] Moreover, 'to' and 'will' can occupy the same position between two clauses, which 'makes it plausible to suggest that the two are different exponents of the same category.'[21] The core function of an auxiliary term like 'will' is to mark the tense of a phrase, and Radford *et al.* therefore suggest that we 'assume that finite auxiliaries and infinitival *to* both belong to the category T of Tense-marker and differ only in that auxiliaries are *finite* (and so are overtly inflected for the past/non-past distinction), but infinitival *to* is *non-finite* (and so is not inflected for the past/non-past distinction).'[22]

As Hyman suggests, a specification of a physical disposition using an infinitival phrase specifies the kind of change that the thing which has that disposition typically causes or undergoes. Now, with the analysis of the infinitival 'to' as belonging to the category of Tense-marker, we can see that the main difference between saying, e.g. 'glass will shatter' and 'glass is disposed to shatter' is that the first is overtly inflected for the future tense, while the second is not inflected for the past/non-past distinction. The infinitival construction for the specification of a disposition, therefore, may be explained by the fact that dispositions are not indexed either to specific times or tenses, i.e. they are *non-finite*.

This provides the beginnings of an explanation of why dispositions are specified with an infinitival phrase. Does it help to show what would be good about explaining the grammar of desire specifications in terms of a conception of desires as dispositions? It seems not, for it indicates that the infinitival grammar of desire specifications will be explained in terms of their not being indexed either to specific times or tenses, but 'typically [being] used in a clause with future time reference'.[23] We should also notice that the same explanation can be given for the fact that desires may be specified with a subjunctive clause: '*A* desires that she drink a cup of tea' contains no overt tense

---

[19] Andrew Radford and others, *Linguistics*: *An Introduction*, pp. 257-9
[20] Ibid. p. 259
[21] Ibid.
[22] Ibid.
[23] Ibid.

inflection in the specification of the content of the desire, but does have future time reference.

Recall Alvarez's comment that 'what is distinctive about the concept of desire is its conceptual relation to *agency*.'[24] Hyman's claim that desires are manifested by purposive actions seems to express a similar view,[25] although he explicitly views the conceptual relation as one deriving from the dispositional nature of desire. The conceptual relation between desire and agency, I think, provides the key to understanding the grammar of desire attributions, and why their content is specified by a construction which lacks overt tense inflection, without conceiving of desires as dispositions.

A common claim is that intentional action, at least typically, is goal-directed behaviour; in the limiting case the goal may simply be to perform the intentional action itself. A second common claim, which is associated in particular with Anscombe and Davidson, is that an action may fall under multiple descriptions, under some of which it may be intentional while under others it may not be.[26] According to Jennifer Hornsby, this idea of actions being intentional under descriptions is more perspicuously expressed by utilising a contrast between *actions* and *things done*. An action, in Hornsby's terminology, is a particular dated event, such as *my eating of breakfast*, to which a particular agent (me) is related in a special way. A thing done, though, is a universal, and as such a repeatable thing, so I can do the same thing on several occasions, and someone else may do it too, for instance I *eat breakfast* most days, and so, probably, do you. Some of the things we do are intentional; others are not. Someone can do one thing with the goal or intention of doing something else, and, if successful, their doing the first thing was (identical to) their doing the second.[27]

Specifications of things that one does, such as [to] eat breakfast, [to] pump water, [to] flip the switch, or [to] drink a cup of tea, seem to carry implicit infinitive particles, as I have indicated in square brackets. Since whenever someone acts they do something, the grammatical similarity between the specification of what they do and the desire they had serves to emphasis the conceptual relation between them. This emphasis

---

[24] Alvarez, *Kinds of Reasons*, p. 70

[25] Note that, for Hyman, intentional actions *manifest* desires; they are not merely *symptoms* of them; see *Action, Knowledge, and Will*, §§5.3-5.6.

[26] See e.g. Donald Davidson, *Essays on Actions and Events*; G. E. M. Anscombe, *Intention*, and 'Under a Description'.

[27] Jennifer Hornsby, 'Anomalousness in Action', p. 624-5, and 'Which Physical Events Are Mental Events?', pp. 64-5

is also secured by the similarity between the specification of a desire with a subjunctive clause and the specification of a corresponding thing done, since the verb they have in common appears in the base form in each.

Furthermore, it is commonly recognised that there is a conceptual relation between desire and the future, namely that the object of desire (what is specified as its content) must be something in the future, or at least not believed to be in the past, although the desire need not be directed at any definite time in the future.[28] The fact that the infinitive participle grammatically functions as a tense-marker, but without overtly inflecting the tense, and that it is typically part of a clause with reference to the future, makes it suited to express the content of desire and to reveal this conceptual relation. The specification of one's goal brings these different elements together. Having a goal is often thought to be a matter of having a desire, where one's goal is specified in the desire's content. Doing one thing with a goal or intention of doing something else makes one's goal or intention amount to what Anscombe labels an intention with which one acts, and if one is successful, then one's doing the latter is identical to one's doing the former. So the infinitival specification, in the content of one's desire, of a thing done as one's goal, exhibits a conceptual relation between the content of one's desire and one's agency. But, since having a goal obviously is a matter of being directed to something future, the infinitival specification of one's goal in the content of one's desire also exhibits the conceptual relation between desire and the future. We have, then, the beginnings of an explanation of why the grammar of content specifications of desires is, grammatically, as it is. It emphasises that desires regard the future, and that they have an essential relation to one's power to act.

Obviously, this suggestion could be developed and refined much further, but I will not pursue this here. What I want to highlight is that the explanation we have just sketched carries no commitment to desires being dispositions, and begins to explain why the grammar of the content specification of desire, which does not allow a propositional clause, is appropriate, without this being in any way due to the supposed non-cognitive character of desire. So none of the arguments we have considered from the grammatical data to the view that desire is non-cognitive are compelling. Of course, nothing I have said justifies the contrary view, that desire is cognitive, either. My argument for that view will have to wait until §3.8.

---

[28] If it does have reference to a particular future time, this will not be achieved by the infinitive itself, but by the addition of a time specification, e.g. '*A* desires to φ *next Saturday at 3.15pm.*'

2.5 The Propositional Formulation of the Content of Desire

Before moving on to a second strategy for grounding the Humean's interpretation of (EED) and acceptance of (FAD) in response to the problems (P1) and (P2), let me say why I think it is relatively unproblematic for us to routinely formulate desire attributions schematically as if they took a propositional content-clause, *so long as we remember that this misrepresents the form of desire attributions for the sake of simplicity and theoretical convenience.*

It is an obvious fact that desires may be contradictory, incompatible, or inconsistent, but how should we represent what it is for one desire to contradict, be incompatible with, or inconsistent with another? The simplest kind of case is one in which one desires to φ and desires not to φ, but even here, while it is clear that there is a contradiction, it cannot be represented in the standard logical way, for contradictions are standardly represented by the mutual assertion of a proposition *p* and its negation ~*p*. Since the contents of the desires are represented by infinitival phrases, we have no propositions or their negations in the content of the desires. A more complex case of contradiction or inconsistency is one where the contents of the desires are not '…to φ' and '…not to φ', but instead are of the form '…to φ' and '…to ψ', but where these are still obviously incompatible. Take the man who wants to drink whisky all the time, but also wants to stay sober. Setting orders of priority amongst one's desires, and attempting to make one's complete set of desires consistent, is one of the functions that practical deliberation serves, and so we need to be able to represent these sorts of incompatibilities in order to represent the reasoning processes that one may engage in in deliberation.

Another sort of case is where one considers the intelligibility of one's desires. For example, we have already seen that desires must be future directed, or at least their object cannot be something that one believes to be past. But this allows the possibility that someone might desire to do something which they have already done and can only be done once, without realising that they have already done it. For instance someone might want to write her first book, not realising that she has already written two books (perhaps she had a head injury). If she were to realise that she has written two books, then, in deliberation, she should be able to realise that her desire is defective. But that means that to represent this reasoning we need to be able to represent the relations between the content of her desire (to write her first book) and the content of her

recovered belief (that she has written two books) such that we can provide a model which shows how in considering these she can realise that her desire is defective in having a past object. But the content of the belief is propositional, while the content of the desire is infinitival, so how are we to do it?

A final sort of case is one in which someone can explain their desire in the form 'I want to φ so that *q*'. For example, I might say 'I want to deposit this cheque so that my account stays in credit'. In order to represent the reasoning that makes this kind of explanation intelligible, as it obviously is, we need to be able to represent the relation between what replaces φ – 'deposit this cheque' – and what replaces *q* – 'my account stays in credit'. To do that, we need to be able to represent a piece of subjunctive reasoning about means that would contribute to some end, roughly like this: 'If I were to deposit this cheque, then my account would stay in credit'.

We can accommodate all three sorts of case by simplifying our treatment of the content of desire so that we treat it as if it were propositional, so we begin by treating the desire to φ as if it were the desire that *p*. We can then use this simplified schema for desire to give a straightforward representation of the relation between the content of a desire and its satisfaction conditions: a desire that *p* is satisfied if and only if *p*.[29] That may look unsatisfactory before taking account of the fact that the representation of the content of desire has been simplified for theoretical convenience, since it fails to distinguish the satisfaction of desire from the truth of belief.[30] But if we remember that, via a mechanical linguistic transformation to the simplified propositional specification, the match between the propositional clauses of the desire and the propositional specification of a state of affairs perspicuously highlights the relation between what is desired and the desire's satisfaction conditions, then the failure of the simplified satisfaction schema to distinguish between satisfaction and truth need not concern us.

We can then say that, in the simple case of directly contradictory desires, the contradiction can be represented as a direct incompatibility between the satisfaction conditions of the desires. The desire to φ is transformed as the desire that *p*, which is satisfied if and only if *p*; the desire to not φ is transformed as the desire that ~*p*, which is satisfied if and only if ~p. So the desire to φ and the desire to not φ are directly contradictory because they have directly incompatible satisfaction conditions. The more complex case of the incompatible desire to φ and the desire to ψ can be represented in a

---

[29] Cf. Bernard Williams, 'Consistency and Realism', pp. 187-91
[30] Nick Zangwill, 'Directions of Fit and Normative Functionalism', p. 173, expresses this concern.

similar way. The desire to φ is transformed as the desire that *p*; the desire to ψ is transformed as the desire that *q*; then, bringing to bear our other knowledge, we can see that *p* only if ~*q*, and *q* only if ~*p*, and so see that the desire to φ and the desire to ψ are inconsistent because they have incompatible satisfaction conditions, even though the incompatibility is not immediate.

The same tactic can be used for the second and third sorts of case. For the second, if someone wants to write her first book, her desire is satisfied when and only when she writes her first book, so we can (mis)represent the desire propositionally as the desire that she writes her first book (she desires that *p*). But if she has already written two books (*q*), this implies that she has written at least one book, and *a fortiori* that she has written her first book (*p*). Therefore, we can represent her reasoning from the fact that she has written two books (*q*) to the conclusion that her desire is defective because its object is something past; *q* implies *p*, but, since it is already the case that *p*, her desire that *p* is defective. Finally, for the third sort of case, we can transform the desire to φ as the desire that *p*, and then represent the explanation as 'I desire that *p* so that *q*'. The piece of subjunctive reasoning can be represented as 'If *p*, then *q*', with 'if…then…' interpreted as the subjunctive conditional. Since in the logical notation the subjunctive conditional does not require grammatical changes in the atomic propositions, unlike the English in which the subjunctive is represented by a change in the verb form, we can see the direct affinity between the subjunctive reasoning and the explanation of the desire as a means to an end. We could make the connection between them even more explicit if we represented the subjunctive reasoning as: 'If my desire to φ were satisfied (i.e. if it were the case that *p*) then it would be the case that *q*'.

The mechanical transformation from '*A* desires to φ' to '*A* desires that *p*', provides a simple way of representing incompatible, inconsistent, or contradictory desires, via a perspicuous way of representing the relation between the content of a desire and its satisfaction. It also provides a simple way of representing the reasoning processes that one can go through in prioritising desires and making them consistent, as well as those through which one can realise the defectiveness of a desire due to its temporal aspects. Finally, it provides a convenient way of representing the relation between means-end reasoning and means-end explanations of desires. So long as we are not misled by the theoretical simplification, treating desires as if they were desires that *p* is harmless.

In the rest of this chapter I discuss a second strategy for grounding the Humean commitment to (EED) and (FAD) in terms of the supposed distinction in respect of the directions of fit of belief and desire. This strategy treats desires as desires that *p,* but, in light of the forgoing discussion, this need not overly concern us.

2.6 Teleological Explanation and Desire

A second Humean strategy for answering (P1) and (P2) is to give a 2-stage argument; first, appeal to the teleological character of the rational explanation of action, and then explain this teleological character in terms of a *direction of fit* account of belief-desire psychology. Very roughly, the idea is that if an attitude has *mind-to-world* direction of fit, its content is meant to fit how the world is, whereas if it has *world-to-mind* direction of fit, the way the world is is meant to fit the content of the attitude.

A prominent example of this strategy can be seen in Smith's teleological argument for a Humean theory of motivating reasons. Smith claims that it is an *a priori* constraint on an account of the explanation of action that it accommodates the teleological character of such explanation.[31] Moreover, he claims that teleological explanation is essentially explanation in terms of the pursuit of a goal and that 'the Humean's theory is alone able to make sense of motivation as the pursuit of a goal.'[32] The teleological character of rational explanation entails the Humean theory of motivating reasons, he claims, because it requires that a reason which explains an action be based on some psychological attitude which has a world-to-mind direction of fit, and that a reason motivates someone in virtue of their having such an attitude, as well as another which has a mind-to-world direction of fit. Since, according to Smith, the independently most plausible conceptions of desire and of belief are in terms of their having, respectively, world-to-mind and mind-to-world directions of fit, this means that if the explanation of action is teleological, then it is explanation in terms of the agent's beliefs and desires, which is just what the Humean theory says.[33] Smith's teleological argument, more formally, is that:

---

[31] Smith, 'The Humean Theory of Motivation', p. 44
[32] Ibid.
[33] Ibid. pp. 45, 50-5

(TA1) Having a motivating reason is, *inter alia,* having a goal

(TA2) Having a goal *is* being in a state with which the world must fit

(TA3) Being in a state with which the world must fit is desiring.

So:

(TA4) Anything that constitutes a motivating reason consists, at least in part, in being in a state of desiring something.[34]

This amounts to a belief-desire theory of reasons, which we found in §1.4 is unsatisfactory, but it can be easily amended along Williams' lines without deviating from the spirit of the theory. In fact, though, Smith's argument is formally invalid,[35] and it has also been claimed that by itself it fails to touch on the fundamental disagreement between Humean and anti-Humean theories.[36] For present purposes, what is important to see is that, with regard to (P1), there is a significant shift from Hume's view to Smith's Humean view. In Hume's view, desire is fundamental in motivation because it is a non-representational causal impulse, while belief is representational. As Smith points out, though, there is no reason why an anti-Humean theory of motivation, say in terms of beliefs only, could not take beliefs to be causal as well.[37] Smith's Humean view, in contrast to Hume's, takes desire to be fundamental in motivation because it takes having a desire to amount to having a goal. This is not to say that it does not *also* take desire to be causal, just that its being causal is not the property of desire that drives Smith's Humean theory. Correspondingly, it takes belief to be less fundamental to motivation just because it has the contrary direction of fit, rather than, as in Hume, because it is not an original existence, and so not a causal impulse.[38] On such an account of practical reasons, they are not merely desire-based because motivational *force* flows from desire to desire in deliberation, but because *goal-directedness* flows through it.

As for (P2), the account of desire as an attitude with world-to-mind direction of fit seems crafted precisely so as to be capable of supporting (FAD). In particular, it looks designed precisely so as to be capable of supplying an argument for this thesis in

---

[34] Ibid. p. 55. I have modified Smith's formulation of the conclusion, which in its original form includes obviously unsupported claims about beliefs as well as desires.

[35] Dancy, *Moral Reasons*, p. 30, and *Practical Reality*, p. 91; Alvarez, *Kinds of Reasons*, p. 121-2

[36] Wallace, 'How to Argue about Practical Reason', pp. 370-1; Pettit, 'Humeans, Anti-Humeans, and Motivation', p. 531. Cf. Margaret Olivia Little, 'Virtue as Knowledge: Objections from the Philosophy of Mind'; Mary Clayton Coleman, 'Directions of Fit and the Humean Theory of Motivation'.

[37] Smith, 'The Humean Theory of Motivation', p. 44.

[38] Both claims, though, are attempts to spell out the representational nature of belief.

the context of a theory which, unlike Hume's, acknowledges that desires have content. It is the fact that the distinction in respect of direction of fit promises to justify (FAD), thereby imposing a distinctively Humean interpretation of (EED) which could be used to support (RDB), which is at the core of this second strategy.

On this second strategy, then, the distinction in respect of direction of fit is put to work carrying a surprisingly heavy load. It looks as if the core features of the Humean theory of reasons are supposed to rest almost entirely on this distinction. In §§2.8-2.9, however, I will argue that the two main Humean accounts of what this distinction amounts to turn out to be incoherent. The distinction, therefore, cannot bear the load. First, though, let us consider the metaphorical origins of the supposed distinction.

2.7 Directions of Fit: Metaphorical and Normative Origins

An early explicit formulation of the distinction in respect of direction of fit is given by Mark Platts:

> The distinction is in terms of the *direction of fit* of mental states with the world. Beliefs aim at the true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit with the world, not vice versa. Desires aim at realisation, and their realisation is the world fitting with them; the fact that the indicative content of a desire is not realised in the world is not yet a failing *in the desire,* and not yet any reason to discard the desire; the world, crudely, should be changed to fit with our desires, not vice versa. [39]

As Platts acknowledges, this way of characterising the distinction is highly metaphorical, and he does not himself endorse it.[40] However, there are several aspects that are worth drawing attention to. The notion of direction of fit here is first given as what looks like a hint towards a correspondence theory of truth (what it is for a belief to be true, it is suggested, is for that belief to 'fit' how the world in fact is), and second as a hint towards an account of desire realisation (what it is for a desire to be realised is for

---

[39] Mark Platts, *Ways of Meaning: An Introduction to Philosophy of Language*, pp. 256-7. Platts attributes the idea to Anscombe, claiming that she gives it in §2 of *Intention*. John Searle, *Expression and Meaning: Studies in the Theory of Speech Acts*, p. 3, also attributes a linguistic variant of the idea to Anscombe, but claims (as is more common) that it comes in §32 of *Intention*. Although it has become standard to attribute the distinction to Anscombe, I think this is actually a mistake. For a detailed discussion of this see Richard Moran and Martin J. Stone, 'Anscombe on Expression of Intention: An Exegesis', esp. pp. 67-9.

[40] Platts, *Ways of Meaning*, p. 257

the world to 'fit' the desire).[41] Crucially, it is the *same* relation of *fitting* in both cases, and in each case that relation holds between the same two sorts of things; *fitting* is an asymmetrical relation which holds between ordered pairs of facts and psychological attitudes.[42]

Another important feature of metaphorical accounts is that the metaphors in question are essentially normative. In the quote above, Platts characterises the distinction as the view that falsity is a decisive *failing* of a belief; false beliefs '*should be changed*' to fit the world.[43] But for desire, a lack of fit is not a '*failing in the desire*'; rather the world '*should be changed*' to fit the desire. John Searle gives a similarly metaphorical, but normative characterisation:

> If my beliefs turn out to be wrong, it is my beliefs and not the world which is [sic] *at fault*, as is shown by the fact that I can correct the situation simply by changing my beliefs. It is the *responsibility* of the belief, so to speak, to match the world, and where the match fails I repair the situation by changing the belief. But if I fail to carry out my intentions or if my desires are unfulfilled I cannot in that way correct the situation by simply changing the intention or desire. In these cases it is, so to speak, *the fault of the world* if it fails to match the intention or the desire[44]

One prominent Humean account of the distinction, we shall see (§2.8), attempts to eradicate the normativity along with the metaphors. Another tries to preserve the normativity without the metaphor, but at the same time attempts to give a non-cognitive analysis of it (§2.9). Both accounts, I will argue, are entirely unsatisfactory.

As I have already noted, these formulations are highly metaphorical, for, as Ralph Wedgwood puts it, 'Beliefs are not little archers armed with little bows and arrows: they do not literally "aim" at anything.'[45] It is only agents who literally aim at things,[46] have responsibilities, and can be at fault.[47] If this is not attended to it is easy to misleadingly slide between, say, speaking about someone's *belief* aiming at truth and

---

[41] As Zangwill, 'Directions of Fit and Normative Functionalism', p. 173, points out, saying that the belief that *p* is true iff *p*, and a desire that *p* is satisfied (or realised) iff *p*, fails to distinguish truth from satisfaction, and so fails to distinguish belief from desire (but see §2.5 above).

[42] Cf. John Milliken, 'In a Fitter Direction: Moving Beyond the Direction of Fit Picture of Belief and Desire', pp. 565-7

[43] As Daniel Whiting ('Does Belief Aim (Only) at the Truth?', p. 288) notes, there are (at least) two interpretations of the idea that beliefs aim at truth, formulable either as truth being a necessary or a sufficient condition for the aim; only the interpretation of the aim as a necessary condition has this implication for false beliefs.

[44] John Searle, *Intentionality*, p. 8, emphasis added.

[45] Ralph Wedgwood, 'The Aim of Belief', p. 267.

[46] Cf. Alvarez, *Kinds of Reasons*, pp. 70-1, and Michael Morris, 'Mind, World and Value', p. 307.

[47] Perhaps e.g. institutions can do so in a derivative sense.

someone aiming to satisfy their desire. The former metaphorically attributes an aim to an attitude, while the latter non-metaphorically attributes an aim to an agent. But neither are helpful for understanding the nature of belief or desire, because of the metaphorical character of the first, and the fact that the second doesn't say anything about desire itself at all.[48] If there is a fundamental contrast, which can metaphorically be put in terms of different aims of belief and desire, then we have to cash out these metaphors for both. I think that there is something to the idea that, metaphorically speaking, belief and desire have aims, and, I think, the metaphors are most naturally understood as alluding to constitutive norms of belief and desire, or at least norm-involving teleological functions.[49]

If we were to cash out these metaphors normatively to give a characterisation of the distinction in respect of direction of fit, a better formulation might be, for example, that truth is the norm of belief, so that, roughly, there is something wrong or defective with believing a falsehood, or so that one ought to believe something only if it is true.[50] Meanwhile, one might say that satisfaction is a norm of desire, so that there is something wrong or defective with desiring something without one's desire being satisfied, or that one ought to bring about what one desires. But once we make what is being claimed clear, it is just obvious that the proposed norms of desire are false. One should only realise what one desires if one's desires are good, or at least permissible,[51] and as Platt's says, that a desire is not satisfied 'is not yet a failing *in the desire*'.[52]

What needs to be emphasised is that the metaphorical formulations of the distinction contain grounds for suspecting that there is a deep tension here. On the one hand, when we focus on belief, it may seem that the account of direction of fit which would most plausibly be applied to it would be a normative account. On the other, it looks like this normative cashing out of the metaphorical formulations of world-to-mind direction of fit is obviously false.

---

[48] It is not clear that the Humean has the resources to make these distinctions; cf. Korsgaard, 'The Normativity of Instrumental Reason', p. 233. I discuss issues related to this further in Ch. 4.

[49] Wedgwood, 'The Aim of Belief', p. 267, takes roughly the same line.

[50] Conor McHugh ('The Truth Norm of Belief') argues that if truth is a norm of belief, it is an evaluative, rather than a prescriptive, norm. The difference roughly corresponds, respectively, to the alternatives I have formulated here. See Alexander Greenberg and Christopher Cowie, 'Is the Norm on Belief Evaluative? A Response to McHugh', for a contrary view, and Peter Railton, 'On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action', for a discussion of some difficulties involved in claiming that truth is a prescriptive norm of belief.

[51] Cf. Lloyd Humberstone, 'Directions of Fit', p. 67; Alvarez, *Kinds of Reasons*, p. 69; Morris, 'Mind, World and Value', pp. 306-7

[52] Platts, *Ways of Meaning*, p. 257

2.8 Directions of Fit: The Functionalist Account

The most plausible non-normative approach to cashing out the metaphors at the root of the distinction in respect of direction of fit is a functionalist approach. This draws on a generally non-normative philosophy of mind, and applies it to this distinction in particular. Unsurprisingly, this is probably the most prominent account of the distinction in the literature.

Smith, for example, takes it that the distinction gets at a deep truth about the essential nature of beliefs and desires, and proposes that we should cash out the metaphors in terms of these psychological attitudes playing different kinds of functional roles. The most important attractions of this approach for the Humean are that, having admitted that desires have content (see §1.3), it promises to provide alternative justifications to Hume's for maintaining the Humean interpretation of (EED) and (FAD). A functionalist cashing out of the normative metaphors of beliefs and desires 'fitting' something (or *vice versa*), locations of 'failings' and what 'should be changed', promises to expunge the normativity in the metaphors, just as it does *qua* naturalistic theory of mind, leaving no essential grounds for rational evaluation, thus answering (P2). There being a fundamental contrast between belief and desire in respect of their directions of fit, meanwhile, promises to vindicate the Humean's insistence on (EED). It promises to do so, moreover, in a way that rules out beliefs or knowledge counting as desires, in the attenuated sense, thereby supporting (RDB) and promising to answer (P1).

According to Smith's influential version of this sort of account:

> [T]he difference between beliefs and desires in terms of direction of fit comes down to a difference between the counterfactual dependence of a belief and a desire that *p,* on a perception that *not p:* roughly, a belief that *p* is a state that tends to go out of existence in the presence of a perception that *not p,* whereas a desire that *p* is a state that tends to endure, **disposing the subject in that state to bring it about that *p.*** Thus […] attributions of beliefs and desires require that different *kinds* of counterfactuals are true of the subject to whom they are attributed [….] this is what a difference in their directions of fit *is.*[53]

Smith's formulation, and in particular the clause I have put in boldface, may be taken to suggest that the fundamental contrast between belief and desire is that desire disposes an agent to *act* so as to satisfy the desire, while belief does not dispose an

---

[53] Smith, 'The Humean Theory of Motivation', p. 54, boldface added.

agent to act. But this is in fact both false and misleading, and distracts from what I think is the more important feature of Smith's characterisation. Let me explain. Even the Humean acknowledges that desire alone cannot produce intentional action – it needs to be combined with a means-end belief.[54] Presumably, then, one member of the set of dispositions given in a functional specification of a means-end belief that if one were to φ then *p,* would be that an agent with this belief would be disposed to φ in condition C, where C includes the presence of a desire that *p*. So there would be, on this functionalist view, both conditions in which someone with this means-end belief, and in which someone with this desire, would be disposed to φ. This means that it could not be the fact, if it were one, that an attitude involved the mere disposition to φ which answered (P1). Since part of the point of the distinction in respect of direction of fit was to respond to this problem, a psychological attitude involving this disposition could not be all that was of central importance to having the world-to-mind direction of fit, since it is a disposition shared by desires *and beliefs*.[55]

The more important feature of Smith's formulation, which this misunderstanding obscures, is the fact that it is not an arbitrary choice of his to specify the conditions for the manifestation of the dispositions as ones in which the agent has 'a perception that *p*' or 'a perception that not-*p*'. These conditions are supposed to make sense of one term of the relation being the relevant *facts* – the relevant way the world is (see §2.7 above).[56] The only reason that 'perceptions,' despite seeming to be psychological states rather than facts themselves, might plausibly achieve this, and so provide suitable conditions so as to qualify as an account of directions of fit at all, is that they are plausibly thought to be psychological states to the extent that they are cognitive mediations *with the world itself*, and so with the facts themselves. This

---

[54] Sinhababu formulates his Humean view so as to allow that desires might produce action alone; see 'The Humean Theory of Motivation Reformulated and Defended', p. 465

[55] One possible response would be that C necessarily includes having the desire that *p* and the belief could only dispose the agent to φ, nothing else. The conditions under which an agent with the desire that *p* will be disposed to do something, on the other hand, may include any means-end belief, where the end is that *p*, and so the desire might dispose them to do any number of things (whatever might be specified as the means). It is obscure, however, what relevance the versatility and multiplicity of the dispositions to act involved in a dispositionally characterised state should have to anything illuminatingly described as its 'direction of fit'. (This is, of course, a simplification, since in complex real-life reasoning a belief might be related to many desires *and* beliefs, and in virtue of these relations might contribute to the disposition to do any number of things too.)

[56] As will become clear below, how we take Smith's phrase 'a perception that…', makes a significant difference to how we should treat his argument. The point made in the main text is that direction of fit is, roughly, what Zangwill ('Directions of Fit and Normative Functionalism', p. 194) calls a 'vertical' relation – a relation between one's psychological states and the world – rather than an 'horizontal' relation – a relation between psychological states.

difference between the counterfactual behaviour of beliefs and desires in response to (possible perceptions of) *ways the world may be*, Smith thinks, is enough to license our speaking of the difference between them in terms of their having opposite directions of fit.[57]

The functionalist account of the distinction in respect of direction of fit, however, is untenable. Lloyd Humberstone argues that Smith's functional characterisation of the distinction faces a trilemma, and this trilemma, I think, reveals general problems for the functionalist account.[58] Humberstone's trilemma arises from three ways that Smith's phrase 'a perception that *not p*' may be interpreted. The first, and least plausible interpretation, is that it refers to specifically sensory perception. For many philosophers, this will be an unattractive interpretation because they take it that negation is not a property available to sensory detection. Such a restrictive interpretation of what Smith means by 'perception', then, would be incompatible with the perception literally having negative content. Even if one does not sympathise with this view, however, there is a further problem with the interpretation. Since Smith is giving a constitutive account of the fundamental difference between beliefs and desires, *any* belief would have to be such that it would tend to 'go out of existence in the presence of a perception that *not p*',[59] and therefore '*p*' must be substitutable for *any* sentence which could give the content of a belief. But if one believed, say, that having a degree in mathematics gives one better employment prospects than one in computer science, it is clear that one could not *literally* have a sensory perception that this was not the case.

On the second interpretation, Humberstone thinks, 'a perception that *not p*' would be parsed as 'coming to believe that *not p*', or, since '…perceives that…' is a factive construction, 'coming to *know* that *not p*'.[60] According to the third interpretation, the phrase should not be interpreted doxastically, as implying belief or knowledge that not-*p*, but instead as it perceptually merely appearing that not-*p*, where this carries no implications about the beliefs which may or may not be formed regarding *p*, but which

---

[57] Smith, *The Moral Problem*, p. 115. Although Smith is explicitly denunciating the normativity in the metaphors, he continues to trade on it to make his teleological argument plausible (§2.6). Thus premise TA2 says, 'Having a goal is being in a state with which *the world must fit*', rather than the much less rhetorically powerful 'Having a goal that *p* is being in a state which tends to persist in the face of finding out that not-*p*'.

[58] Humberstone presents his objection as a dilemma, dismissing what I call the first horn without explanation of what is wrong with it; see 'Directions of Fit', p. 63

[59] Smith, 'The Humean Theory of Motivation', p. 54

[60] Humberstone, 'Directions of Fit', p. 63. A construction is factive if and only if the whole construction is true only if the contained propositional clause is true, e.g. 'A perceives that *p*' is true only if '*p*' is true, 'A knows that *p*' is true only if '*p*' is true etc.

would serve as the ground for such beliefs. So, for example, if you were to help someone disguise himself as Elvis Presley, then stood back and evaluated the disguise, it would perceptually appear to you that you were looking at Elvis (on the assumption that you've done a good job), though you would not believe that the person in front of you actually is Elvis.[61]

If, however, we interpret a perception that not-*p* as implying belief or knowledge that not-*p*, then in giving this characterisation of the mind-to-world direction of fit we are forced to appeal to the notion of belief with this very direction of fit; the mind-to-world direction of fit is *presupposed*. As Humberstone says, 'you cannot informatively characterise a fundamental disanalogy between the ways in which beliefs and desires relate to their objects by contrasting them in a respect itself specified by reference to one of those two ways'.[62]

If, on the other hand, we interpret 'perceives that *not-p*' as a mere perceptual appearance, which would ground, but does not entail, a belief about whether or not *p*, the previous circularity objection is avoided.[63] But what is left now is that the belief that *p* tends to evaporate when one is faced with the perceptual appearance that *not p*, while the desire that *p* tends to persist. But Smith means to give an account of the fundamental nature of belief and desire. A difference of a mere tendency in their counterfactual behaviour, however, makes it look like they only have a *typical* direction of fit, rather than their direction of fit being essential to them. If belief and desire are to have distinctive opposite directions of fit, and this consisted in how they would behave in the face of appearances, then either such counterfactual behaviour would have to have strict generality, or else they could only be said to have a usual direction of fit. On the latter supposition, though, a direction of fit could be neither essential to, nor constitutive of, what belief and desire are.

If Smith were to revise his claim, though, and assert strict generality to the asymmetrical counterfactual behaviour of belief and desire, his claim would be manifestly false. David Sobel and David Copp give numerous counterexamples to such behaviour for both beliefs and desires. For beliefs, they claim there are two kinds of

---

[61] Consider also the Müller-Lyer illusion.

[62] Humberstone, 'Directions of Fit', p. 64. David Sobel and David Copp,'Against Direction of Fit Accounts of Belief and Desire', p. 46, fn. 1 and accompanying text, are misled by Humberstone's confusing discussion of circularity into thinking that he takes Smith not to be offering an account of what belief and desire consist in. In fact, Sobel and Copp, as well as Humberstone, take Smith (correctly in my view) to be offering just this.

[63] Humberstone reports ('Directions of Fit', p. 64, fn.10) that Smith himself maintains that he intended this second interpretation in 'The Humean Theory of Motivation'.

counterexamples: 'stubborn beliefs' – beliefs held in the face of appearances to the contrary (e.g. confidence in the existence of God in the face of the problem of evil) – and beliefs in necessary truths, which no appearance or 'introduced state' could tend to undermine. For desire, they give the counterexample of the fair-weather sports fan, who desires the team she supports to do well, but when they do not switches her allegiance to another team. 'Must we say then' they ask rhetorically, 'that [she] does not really root for the 49ers and instead merely believes that they will win?'[64]

So, if Smith's proposal were to be salvaged, he would have to say that how a psychological attitude behaved in the face of appearances *was* constitutive of the attitude's direction of fit. But then the characterisation of belief and desire would fail to support a fundamental asymmetry in their nature.

It might be thought that this is a false trilemma, and that there are more possibilities than it allows. But even if this were true, it would not help the account. Sobel and Copp again produce an extensive range of alternative 'introduced states' which might be given as interpretations of 'a perception that *not p*', such as that it is a desire, intention, or contentment that *p*, and even including the possibility that while the content is specified, the kind of attitude itself is left open. For each possibility, they argue, there are either counterexamples or a failure of strict generality, resulting in a failure to characterise what these attitudes essentially are.[65] As Sobel and Copp note, it should not be surprising 'that we cannot find an introduced state that counts as in some way a perception with the content that not *p*, that is *not* itself a belief, but that interacts with the belief that *p exactly as if* it were an incompatible belief.'[66]

Another strategy for avoiding these objections is suggested by Mary Clayton Coleman. According to her, the root of the functionalist account's vulnerability to the objections is that it conceives of the dispositions constitutive of an attitude's direction of fit atomistically; conceiving of them holistically instead, she suggests, might enable the functionalist account to avoid Humberstone's, and Sobel and Copp's objections. On this alternative view, she explains:

---

[64] Sobel and Copp, 'Against Direction of Fit Accounts of Belief and Desire', pp. 47-8
[65] Ibid. pp. 49-51. Notice that each of these possibilities would be an example of a horizontal relation, rather than a vertical relation (see fn. 56 above). In consequence the relation characterised would not be the relation of 'fitting'. This reinforces the earlier point that 'a perception that *not p*' is not an arbitrary choice of introduced state for the functionalist account.
[66] Ibid. p. 49. Cf. Barry Stroud, 'Sense-Experience and the Grounding of Thought', pp. 85-7

> [The functionalist] could claim (very roughly) that a particular mental state has mind-to-world direction of fit just in case it has most of the mind-to-world-constituting dispositions. This is what we might call numerical or statistical holism […] because it is about what percentage of direction-of-fit-constituting dispositions each belief has [….] In short, each belief has most of the mind-to-world constituting dispositions [….] Instead of saying that most beliefs have each and every constituting disposition, it says that each and every belief has some majority of the set of constituting dispositions (and each belief need not have the same majority of that set).[67]

Coleman's thought is that there may be many dispositions which are, holistically, constitutive of an attitude's direction of fit, not just its dispositions in response to a perception that not-*p*. So long as an attitude has *most* of these constitutive dispositions, it will count as, say, a belief, even if it lacks the disposition to disappear in the face of a perception that not-*p* (e.g. in cases of stubborn beliefs); similarly, *mutatis mutandis*, for an attitude qualifying as a desire (e.g. Sobel and Copp's fair-weather fan).[68]

But there are difficulties with Coleman's strategy.[69] Even assuming, as she allows,[70] that we may finitely specify the sets of dispositions characterising psychological attitudes, or the sets of dispositions constitutive of each direction of fit, the holistic criterion for an attitude having one or the other direction of fit is ambiguous. It could either mean (i) that a statistical majority of an attitude's characterising dispositions are dispositions constitutive of a direction of fit, or (ii) that a statistical majority of the dispositions constitutive of a direction of fit in fact characterise some attitude. In either case, nothing has been said to rule out the possibility that we are wrong about which attitudes we (purportedly) take to have one or the other direction of fit. What is perhaps more problematic for (ii) is that it is hard to see how its proponent could rule out the possibility that a single attitude may be characterised by a majority of the constitutive dispositions of *each* direction of fit.[71] This might be so even for attitudes that are not even controversially claimed to be 'besires' – attitudes with bi-

---

[67] Coleman, 'Directions of Fit and the Humean Theory of Motivation', p. 131

[68] Strictly, Coleman should speak of types of dispositions, which different beliefs or desires could have in common, rather than particular dispositions characterising particular beliefs or desires. I shall ignore this for the sake of simplicity in what follows. Cf. Mark Van Roojen, 'Humean Motivation and Humean Rationality', esp. pp. 46-9, for a similar suggestion towards holism, based on the role of principles of charity in interpretation.

[69] One might suspect that Coleman's strategy will be unsuccessful even before considering the difficulties I am about to discuss. Smith's account seemed to fail because it could not deal with the possibility of someone having atypical dispositions. Simply increasing the number of dispositions we take into account might seem to multiply, rather than solve, the problem.

[70] Coleman, 'Directions of Fit and the Humean Theory of Motivation', p. 131, fn. 2

[71] Coleman herself might welcome such a result; see ibid. p. 139

directional fit.[72] For (i), on the other hand, if a slim majority of an attitude's characterising dispositions were constitutive of one direction of fit, but of minor importance relative to other constitutive dispositions of that direction of fit (e.g. they had nothing to do with how one perceives things to be), the attitude would still qualify as, say, a belief – even if the minority of its characterising dispositions included important constitutive dispositions of the contrary direction of fit.[73]

Even if we allowed that, at least in ordinary cases, a sufficient number of the dispositions characterising an attitude would be dispositions constitutive of a *single* direction of fit, Coleman's suggestion faces another, deeper, problem. As soon as we try to specify a range of the dispositions which are supposed to be holistically constitutive of a direction of fit, we realise that many, if not most, of them will be dispositional relations between different psychological attitudes of the agent. For example, Coleman suggests that one of the dispositions constitutive of someone's attitude that *p* having mind-to-world direction of fit, and so being a belief that *p*, is that if they discover that if *p* then *q*, they will be disposed to believe that *q*.[74] But dispositional relations such as these are 'horizontal' relations – relations between different attitudes of an agent – rather than 'vertical' relations – relations between an attitude and the world, or the facts.[75] The intuitive but metaphorical characterisations of directions of fit seem to require that *fitting* be a relation between an attitude and the facts. It was just this that motivated Smith's focus on having a perception that not-*p* as the salient condition for the manifestation of dispositions constitutive of each direction of fit; their being responses to a perception that *p* or that not-*p* is what made it plausible that *the world* is in the picture at all. Coleman's horizontal relations flaunt this apparent condition on an adequate cashing out of the original metaphors.

This shows that it would be necessary to give an account of why just *these* dispositions (whichever they turned out to be) should be regarded as constitutive of each direction of fit. That means, however, that it would actually be *this* account, which explained why *these* dispositions were related to mind-to-world direction of fit, and why *those* dispositions were related to world-to-mind direction of fit, that would purport to give the account of what these directions of fit actually *are*. But if this is the case,

---

[72] The unfortunate label comes from J. E. J. Altham, 'The Legacy of Emotivism'.
[73] This plausibly assumes that there would be gradations of importance amongst the constitutive dispositions, and in particular between vertical and horizontal dispositional relations.
[74] Coleman, 'Directions of Fit and the Humean Theory of Motivation', p. 131, but see also p. 135.
[75] Zangwill, 'Directions of Fit and Normative Functionalism', p. 194

then Coleman's suggested holistic dispositionalism would not really be a functionalist account of what directions of fit *are* at all, and she has said nothing about how to provide such an account.

The holistic functionalist account of directions of fit, then, is a red herring, and we are remanded to Sobel and Copp's objection: whatever the 'introduced state' specified in the conditions for the manifestation of the constitutive dispositions of directions of fit is supposed to be, it would have to interact with a belief (or desire) exactly as if it were a contradictory belief, without actually being a belief at all. It is unsurprising that we cannot find such an introduced state.

But the problem is more fundamental to functionalist accounts of directions of fit than Sobel and Copp's discussion brings out. In saying that what is required is something which interacts with a belief that *p* as if it were a contradictory belief, it is implicitly recognised that the interaction which is sought is a distinctively *rational* and *normative* interaction, which in the case of interacting beliefs will exhibit logical relationships between their propositional contents. Any divergence from such logical relations will *ipso facto* be cases of irrationality, or at least less than full rationality.[76] So the interaction that is described is the interaction that rationally *should* occur, not that which *would* occur in possible circumstances (though it might be that things would occur as they rationally ought to). Any proposed account of directions of fit which is entirely non-normative (functionalist attempts would be paradigm examples), will in the end be unsatisfactory simply because they attempt to describe what rationally should happen by appealing solely to what they claim would happen. Since what would happen and what should happen come apart, they will always lack the resources to count some cases of what would happen as *mistakes* or *errors*, and will thus be forced either to treat them as counterexamples, or not to count the strictly general patterns of what *would* occur as constitutive of the phenomena whose essential nature they meant to provide accounts of.[77]

---

[76] By the latter I mean to suggest cases in which it would be unnatural to convict someone of irrationality because e.g. they simply failed to *notice* a contradiction. See further Scanlon's distinction between 'narrow' and 'wide' construals of rationality in *What We Owe To Each Other*, ch. 1.4; and also Derek Parfit, *On What Matters*, pp. 33-6

[77] Compare Saul Kripke's argument against dispostionalist attempts to solve the paradox of rule-following in *Wittgenstein on Rules and Private Language*, pp. 22-37

2.9 Directions of Fit: Higher-Order Attitudes and Normativity

A second prominent Humean approach to providing a non-metaphorical account of the distinction in respect of direction of fit, attempts to accommodate the normativity involved in the metaphorical characterisations from which it begins (§2.7). It attempts to do this, however, just to the extent that is required to accommodate the possibility of there being mistakes or errors of the kind that plague the functionalist account. However, as a broadly Humean account, it also avoids acknowledging any sort of full-blooded, value-based, normativity. In effect, it does this in a parallel way to that in which Williams attempts to secure the normativity of reasons. As we saw in §§1.4-1.5, Williams attempts to do this by relativising it to the prior desires, or 'subjective motivational set', that an agent has. In exactly the same way, this second Humean attempt to give an account of directions of fit takes directions of fit to be normative, but claims that the normativity is relative to a higher-order psychological attitude of the agent. This strategy, though, is still supposed to be capable of answering problems (P1) and (P2).

Perhaps the most explicit example of this strategy is given by Humberstone.[78] According to him, there are two conditions of adequacy on a non-metaphorical account of directions of fit, which effectively prohibit the main difficulty we have been considering with the functionalist account. The first I shall call the *universality* condition:

> (UC)　An account of directions of fit must hold universally for all possible instances of states with one or the other direction of fit.

The second I shall call the *normative* condition:

> (NC)　An account of directions of fit must make sense of the normativity of the original metaphorical characterisations.[79]

Humberstone adopts alternative terminology, using '*thetic*' and '*telic*' in place of 'mind-to-world' and 'world-to-mind' directions of fit. This is primarily because this

---

[78] Much of what David Velleman says suggests a very similar view to Humberstone's; see especially J. David Velleman, 'The Guise of the Good', and also 'On the Aim of Belief'.
[79] Humberstone, 'Directions of Fit', pp. 71-2

terminology pertains both to the distinction as applied to linguistic phenomena and to psychological attitudes.[80] He also seeks to leave room for there to be multiple psychological attitudes with each direction of fit, which, in virtue of this fall into two broad classes. For instance, Humberstone thinks that not only beliefs, but perceivings, predictions, guesses and recollections all have the *thetic* direction of fit, while, for example, intentions, wishes, hopes, as well as desires have the *telic* direction of fit.[81] With this in mind, let us turn to Humberstone's proposal, first considering how it meets (UC), before going on to consider how it meets (NC).

Humberstone suggests that considering the contrast between belief and, say, idle supposition or imagining, will help us in giving an account of the *thetic* direction of fit. The difference here, he says, is that while a belief that *p* could be false, and in this respect is like a supposition or imagining that *p*, which could be done even in the circumstance that not-*p*, it *matters* in regard to the belief (but not to the supposition) whether or not *p*. It matters, moreover, because it has to do with the criterion of *success* for belief: 'If the criterion of success is *truth*, then the propositional attitude is *belief*'.[82] This, he thinks, means that it is a constitutive, and not merely a regulative, principle, that 'unless the attitude-holder has what we might call a controlling background intention that his or her attitudinising is successful only if its propositional content is true, then the attitude taken is not that of belief.'[83]

This is not to say, though, that it is incoherent to conceive of cases in which one would succeed in something by falsely believing that *p*, and in which it would therefore be instrumentally beneficial to have this false belief. Nor is it to suggest, implausibly, that it is unintelligible that one might be 'unconcerned at the thought that *some or other of one's present beliefs were false*'.[84] Humberstone describes a case of the first kind, in which, from a prudential perspective it would not be a mistake to get into a position of

---

[80] Ibid. p. 60. Searle uses 'word-to-world' and 'world-to-word' for this purpose (*Expression and Meaning*, pp. 3-4)

[81] Humberstone, 'Directions of Fit', pp. 63-6; cf. Smith, *The Moral Problem*, p. 117. Velleman claims that similar classes of attitudes fall into two kinds: those which regard a proposition as true, and those which regard a proposition as to be made true. This, he thinks, is what is marked by the distinction in respect of direction of fit. According to him, however, only beliefs *aim* at the truth ('The Guise of the Good', pp. 110-6). It might be thought that this requires the different attitudes in each class to be distinguished from each other by further *differentia*, as Smith, Humberstone, and Velleman seem to suppose, but this evinces an unnecessary commitment to what Anton Ford, 'Action and Generality', calls an 'accidentalist' account of the relationship between a general kind of thing and a particular instance or species of it. I discuss this further in Ch. 4.

[82] Humberstone, 'Directions of Fit', p. 73; cf. Velleman, 'The Guise of the Good', pp. 112-4

[83] Humberstone, 'Directions of Fit', p. 73

[84] Ibid. p. 67

having this false belief, but that nevertheless in this position one *mistakenly believes* that p.[85] This brings out the fact that, in specifying the higher order intention which Humberstone thinks is constitutive of the *thetic* direction of fit, we must be wary of a scope ambiguity. The intention is:

(CIB)   For any p (I intend that (I do not believe that p when not-p)).

It is not the alternative:

(CIB*) I intend that (for any p (I do not believe that p when not-p)).

Humberstone's statement of the logical form of the intention as: '*Intend*(-Bp/-p)'[86] fails to disambiguate the scope of the intention. But by making the ambiguity clear and seeing that the quantifier falls outside the scope of the intention, we can understand Humberstone's view that the incoherence of nonchalance in the face of acknowledged falsity in one's beliefs 'only arises for some *particular* belief cited as possessed in spite of its falsity'.[87]

Humberstone applies this treatment of attitudes like belief with *thetic* direction of fit to attitudes like desire with *telic* direction of fit. They too, he suggests, are subject to a constitutive higher-order intention that it be the case that p if one desires that p, or as he puts it '*Intend* (p/Wp)'.[88] Again, we can disambiguate the scope of the intention; it is:

(CID)   For any p (I intend that (it be the case that p if I desire that p).

It is not the alternative:

(CID*) I intend that (for any p (it be the case that p if I desire that p)).

As for (NC), he thinks that we should take our lead from Anscombe's use of the notion of a *mistake* in *Intention* §32. Anscombe describes a man who goes shopping

---

[85] Ibid. p. 68
[86] Ibid. p. 75
[87] Ibid. p. 67
[88] Ibid. p. 75

with a shopping list, while a detective follows him around making another list of the items he buys, suggesting that:

> if the list and the things the man actually buys do not agree, and if this and this alone constitutes a *mistake*, then the mistake is not in the list but in the man's performance […] whereas if the detective's record and what the man actually buys do not agree, then the mistake is in the record.[89]

Humberstone takes it that the idea here is that the notion of a mistake is normative, but that the normativity in play is relative to the agent's prior (higher-order) intentions. Inspired by Anscombe's description, he suggests that 'one only makes a mistake when one thwarts one's own goals.'[90] Since attitudes with the *thetic* and *telic* directions of fit, on Humberstone's view, are subject to constitutive principles according to which they imply higher-order intentions, these higher-order intentions provide goals against which we can make such normative evaluations of actual or possible conduct (including mental acts): 'the "ought" in "what ought to fit what" is given as a matter of compliance with [the] intention'.[91] The normativity in play, moreover, legitimises our speaking of what we are dealing with in terms of directions of fit between mind and world: *beliefs* ought to fit the *world* because the higher-order intention that they do so is constitutive of belief, while the *world* ought to be as we *desire* it to be because the higher-order intention that *it* do so is constitutive of desire.

With this non-cognitivist account of normativity, as relative to one's prior higher-order intentions, Humberstone can make sense of the varieties of rational deviance which plagued the functionalist account. It may well be that someone believes that *p*, finds out that not-*p*, but nevertheless continues to believe that *p* (perhaps they ignore the fact that they have found out that not-*p*, or fail to notice the contradiction, or just sustain contradictory beliefs). So long as their belief that *p* is subject to the constitutive intention for *thetic* direction of fit, however, we can make sense of these possibilities as varieties of error. Similarly for desire: it may be that one desires that *p*, finds out that not-*p*, and stops desiring that *p*. Again, this has no tendency to undermine

---

[89] Anscombe, *Intention*, §32, p. 56

[90] Humberstone, 'Directions of Fit', p. 67. Philippa Foot claims that irrationality is constituted by intentionally thwarting one's own ends: 'Irrational actions are those in which a man in some way defeats his own purposes, doing what is calculated to be disadvantageous or to frustrate his ends.' ('Morality as a System of Hypothetical Imperatives', p. 310). At least one of Humberstone and Foot must be wrong, for otherwise irrationality would be a matter of intentionally making a mistake, which is absurd. I am inclined to think they are both wrong.

[91] Humberstone, 'Directions of Fit', p. 75

our original view that the agent really did desire that *p*. So long as it was, when possessed, subject to the constitutive intention for *telic* direction of fit, we can make sense of this deviation from typically associated dispositions – for example, by interpreting the agent as having been fickle in their desire. But since the normativity involved, which grounds such evaluations, only obtains relative to the agent's own intention, it is a thin enough conception of normativity to be tolerated by the Humean; it has no full-blooded connection with value.

The proposal also accommodates a further condition of adequacy highlighted by Zangwill. It is a fundamental problem, he says, that the functionalist account takes the dispositions it describes to be constitutive of the directions of fit. Instead, Zangwill argues, the following is a condition of adequacy:

(Z)   The dispositions one has when one believes or desires that *p* ought to be *explained by* the account of what it is to have a belief or desire that *p*.[92]

Humberstone's account provides for this: giving up the belief that *p* when one finds out that not-*p* may be explained in terms of compliance with the constitutive intention for belief as an attitude with *thetic* direction of fit, while maintaining a desire in such circumstances makes sense because this complies with the constitutive intention for *telic* direction of fit.

Unfortunately, though, while it is an improvement on the functionalist account, Humberstone's account is untenable, for two reasons. First, while it succeeds in accommodating (Z), and is also able to characterise deviations from appropriate dispositions as errors of various kinds because it accommodates (NC), it is only able to do so by getting the direction of explanation wrong in another place. The important question to ask is *why* it should be plausible that beliefs and desires should be subject to necessary and constitutive higher-order intentions like (CIB) and (CID). Presumably, the natural answers are, roughly, that beliefs are the kind of thing that one ought to have only if, or which are defective unless, they are true, and desiring that *p* is to take it, in some way, that it would be good if *p* (cf. §2.7). But this is just to say that the intrinsic, full-blooded normativity involved in the nature of belief and desire is prior to, and so explains the plausibility of, the higher-order intentions – intentions which were

---

[92] Zangwill, 'Directions of Fit and Normative Functionalism', p. 179

supposed to be essential for accommodating (NC), on a non-cognitive interpretation of the relevant normativity.

Second, Humberstone's account turns out to be circular in exactly the way that he objects to when discussing the functionalist account (see §2.8 above). Humberstone appeals to higher-order intentions as what *constitutes* each direction of fit, but, on his view, intention itself has one of those directions of fit – the *telic* direction. But the objection generalises. The attempt to introduce normativity, understood non-cognitively as relative to prior higher-order attitudes, depends on the higher-order attitude setting a goal against which the lower-order attitude is to be evaluated. But the distinction in respect of direction of fit was in large part introduced so as to account for an attitude being goal-directed at all. So any proposed higher-order attitude which seemed a candidate to introduce such normativity would have to itself presuppose the distinction it is supposed to account for.[93] No higher-order attitude account of directions of fit can escape this problem. Either it will turn out to be circular, or else it will be viciously regressive, requiring some new higher-order attitude of level $n+1$ for each constitutive attitude of level n.

The higher-order attitude account of directions of fit, then, does not survive scrutiny any more than the functionalist account. Neither of the two dominant Humean accounts of the distinction is workable; it therefore seems that the distinction is unable to do the work the Humean requires of it. It is unable to vindicate either the Humean interpretation of (EED) or (FAD).[94]

2.10 Conclusion

In this chapter, we have considered two strategies for filling the justificatory gap that the Humean is left with after abandoning Hume's theory of thought, and in particular, his conception of desires as original existences; we have considered, that is, strategies for answering (P1) and (P2). In §§2.2-2.5 I considered whether there were sound arguments related to the grammar of desire attributions which could support thinking of desires, or at least some desires, as non-cognitive attitudes, and thereby at least begin to

---

[93] Other candidates might be, e.g. *acceptance* of a *norm* (e.g. Gibbard, *Wise Choices, Apt Feelings*, ch. 4), a *decision* about *what to do*, or *adoption* of a *plan* (e.g. Alan Gibbard, *Thinking How to Live*, esp. ch. 3). As Scanlon, *Being Realistic About Reasons*, p. 57, notes, any of these alternatives will be a state which 'is some kind of resolution or practical commitment', just like intention.

[94] See Scanlon, *Being Realistic About Reasons*, pp. 64-6, and G. F. Schuler, *Reasons and Purposes*, pp. 25-37, for discussions of directions of fit which are in some respects similar to mine, and with which I have some sympathy.

support the Humean interpretation of (EED). I argued that those arguments are unsound, and sketched an alternative explanation of the grammatical data that does not support the view that desires are non-cognitive.

In §§2.7-2.9, I argued that a second strategy, which attempts to vindicate (FAD) and a distinctively Humean interpretation of (EED) by appealing to a supposed distinction in respect of the directions of fit of beliefs and desires, is also unsuccessful. The functionalist account, because it attempts to eradicate the normativity in the metaphors which get the distinction going, lacks the resources to count certain dispositions, or manifestations of dispositions, as mistakes. It therefore faces a dilemma of either admitting that the relevant directions of fit of beliefs and desires are not essential to them, or else admitting that it lacks the resources to explain what directions of fit are. The higher-order attitude account initially has more success, since it can accommodate both (UC) and (NC). However, it gets the direction of explanation between the higher-order intention and the attitude of which it is supposed to be constitutive wrong, and, even worse, it turns out to be circular.

If the argument of this chapter has been correct, both (FAD) and (RDB) seem to be nothing more than unjustified dogmas of Humeanism. We have no reason to think that desires do not admit of fundamental rational evaluation, or to think that desires are non-cognitive attitudes, and therefore we have no reason to accept that practical reasons are desire-based. But, as yet, we have not seen why we should *deny* those theses either. In the next chapter, I will turn to developing an alternative, radically anti-Humean account of the nature of desire, one which rejects (FAD) in particular. I will begin by considering Kant's moral psychology, because it is there, I think, that the key to a satisfactory alternative is to be found.

Chapter 3

Normative Psychology:

Rational Evaluation, Desire, and Evaluative Belief

> *The conceptual connexion between 'wanting' [...] and 'good' can be compared to the conceptual connexion between 'judgment' and 'truth'. Truth is the object of judgment, and good the object of wanting; it does not follow from this either that everything judged must be true, or that everything wanted must be good.*
>
> ~ G. E. M Anscombe[1]

3.1 Introduction

In this chapter, I develop a proposal for a radically anti-Humean philosophical and moral psychology, and in particular for an anti-Humean conception of the nature of desire. I begin with a consideration of Kant's moral psychology.

In the *Groundwork of the Metaphysics of Morals*, Kant draws a distinction between a divine will, on the one hand, and the kind of will that we human beings have, on the other. The crucial difference between these is the nature of the determination of the will by reason. A divine will is necessarily determined by pure practical reason alone, and this means that 'the actions of such a being that are recognised as objectively necessary are also subjectively necessary; i.e. the will is a capacity to choose *only that* which reason, independently of inclination, recognises as practically necessary.'[2] The will of sensuous human beings is fundamentally different; pure reason is not sufficient to determine it, since it is 'also subject to subjective conditions (to certain incentives) that are not always in agreement with the objective ones'.[3]

Kant aims to make room for a moral psychology for human beings which can accommodate actions which are not motivated by subjective desires, and sees this as essential for the possibility of genuinely moral action. In the course of this effort, he provides crucial insights into why an Humean moral psychology makes genuinely

---

[1] Anscombe, *Intention,* §40, p. 76
[2] Immanuel Kant, *Groundwork of the Metaphysics of Morals*, p. 53 (4:412) References in parentheses are to the Royal Prussian Academy of Sciences edition of Kant's Gesammelte. All references are to the Gregor and Timmermann translation unless otherwise noted.
[3] Ibid.

moral action problematic, and into what it would take for a moral psychology to avoid these problems. But his philosophical and moral psychology for human beings, who possess a 'contingently determinable will',[4] may be revealingly interpreted as working within an Humean framework and accepting certain characteristically Humean presuppositions, even while he is trying to escape from this framework and rejecting other presuppositions.

In §3.2 I consider Kant's account of the moral psychology for non-moral, 'heteronomous' action, arguing that it is in relevant respects similar to Hume's moral psychology, and presupposes certain views about the nature of 'incentives' which are similar to Hume's conception of desires, which Hume thinks of as passions. I then look at Kant's account of the moral psychology for 'autonomous' moral action, examining the crucial differences which he thinks avoid the problematic aspects of heteronomous action and allow for the possibility of genuinely moral action. I argue that there are important lessons to learn from Kant's account of autonomous action, but that his rationalist presuppositions divert his account on an unnecessary path. Most importantly Kant shows that we can and should allow that there can be motivating attitudes which are rationally evaluable, and, when separated from his rationalist presuppositions, this motivates the view that desire as such is rationally evaluable. This also motivates the rejection of characteristic Humean assumptions that Kant shares, while it also avoids the problematic aspects of heteronomous action which worry Kant.

In the rest of the chapter, I develop an anti-Humean philosophical psychology in the light of the lessons from Kant. I claim that belief should be understood in terms of a constitutive norm of truth and §§3.3-3.6 are concerned with the proper formulation, interpretation, and implications of such a truth norm of belief. In §3.7, I explore the view that desire is likewise subject to a constitutive norm, and attempt to develop an adequate formulation of such a norm. On the basis of the formulations of the truth norm of belief and of the norm of desire, I argue in §3.8 that we should understand desire as a cognitive attitude: as a certain kind of evaluative belief. I conclude in §3.9, contrasting the resulting account with Humean non-cognitivism.

---

[4] Ibid. p. 55 (4:413)

3.2 Lessons from Kant's Moral Psychology

According to Kant, there are several elements involved in agency. Of particular relevance to us are Kant's notions of inclinations and maxims. Like Hume, Kant thinks that we have sensuous and empirically determined feelings, in particular pleasures and pains, which possess a kind of motivating power. These are produced by objects affecting us through 'sensibility', i.e. our capacity for sense-based experience.[5] An habitual sense-based feeling of this sort Kant calls an inclination, and inclinations constitute what he calls 'pathological' incentives for action.[6] As Andrews Reath characterises an incentive, it is 'a subjective determining ground of the will in the sense that it is the motivational state of the subject that is operative on a particular occasion,'[7] while Henry Allison explains that 'it is the "conative or dynamic factor in volition," the actual impetus to action'.[8]

As for maxims, Kant defines these as 'the subjective principle[s] of willing'.[9] As *principles* of willing, it seems that Kant thinks of maxims as propositionally articulated, and this is confirmed by the few rough examples he provides. For instance, Kant describes the case of a man who makes a false promise of repayment in order to secure a loan; 'his maxim of the action would go as follows: when I believe myself to be in need of money I shall borrow money, and promise to repay it, even though I know that it will never happen.'[10] A maxim therefore states a policy one has adopted, a policy to perform a certain type of action in a specified type of circumstance.[11] Inclinations do not exert a direct force on the will, but rather 'they do so only by being "incorporated into a maxim,"'[12] as Allison puts it, and so 'the spontaneity and rationality of the agent

---

[5] This is not to say that Kant has a hedonistic theory; see Christine M. Korsgaard, 'Kant's analysis of obligation: The argument of *Groundwork I*', p. 56. See also Schroeder, *Slaves of the Passions*, p. 25, who points out that it is not necessary for a Humean theory to be a crude '*self-interest* theory', and that most versions are not.

[6] Compare Hume, *Treatise* II.iii.3, pp. 418-9: 'when a *passion* has once become a *settled principle* of action, and is the predominant *inclination* of the soul, it commonly produces no longer any sensible agitation' (emphasis added).

[7] Andrews Reath, 'Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination', p. 9

[8] Henry E. Allison, *Kant's Theory of Freedom*, p. 121

[9] Kant, *Groundwork*, p. 29 (4:400)

[10] Ibid. p. 73 (4:422)

[11] Although, as principles, maxims seem to belong to reason, they will not serve precisely the same role in Kant's theory of action as beliefs do in Hume's. An ordinary belief will still be required by Kant as the state through which someone recognises that the circumstance which confronts them is of the type mentioned in the maxim.

[12] Allison, *Kant's Theory of Freedom*, p. 126

are involved in heteronomous or inclination-based agency'.[13] But this also means that one adopts a maxim as a principle of one's own *as* a principle which serves the satisfaction of some inclination, while the inclination itself provides an end or goal, as in the Humean theory.

Maxims, incorporating inclination-based incentives in the case of non-moral actions, provide the contingent and *subjective* grounds of an action for a being who's will is not necessarily determined by and in accordance with reason. Kant also describes these subjective principles of the will, i.e. maxims, as the representation of laws, which are objective principles. 'The representation of an objective principle,' he says, 'in so far as it is necessitating for a will is called a command (of reason), and the formula of the command is called IMPERATIVE.'[14] So a maxim has the form of a command of reason, and such imperatives of reason are the *objective* grounds of action.

Famously, Kant distinguishes two kinds of imperative, the hypothetical and the categorical, arguing that for an action to have moral worth, the maxim on the basis of which it is performed must be in accordance with the categorical imperative. Hypothetical imperatives, he says, 'represent the practical necessity of a possible action as a means to achieving something else that one wants (or that at least is possible for one to want)' while 'The categorical imperative would be the one that represented an action as objectively necessary by itself, without reference to another end.'[15]

It is easy to see how an hypothetical imperative, as an objective principle of reason, could get a grip on agents so that they would act on the basis of a maxim which was its subjective correlate. All that would be required is that an agent actually did want the thing for which the possible action was practically necessary as a means. If one wants this thing, i.e. has made it their end, the hypothetical imperative is rationally binding, for it is a kind of irrationality to want something without this want transferring onto what is known to be a necessary means to it. Having a desire or inclination, in other words, makes it possible to act on an hypothetical imperative subjectively represented in a maxim, or to adopt the maxim as one's own; it supplies the motivating force for the action, and makes the imperative binding. But a hypothetical imperative, and so the related maxim, can be binding for one only on the condition that one has the relevant desire, which is empirically determined in sensibility. The hypothetical

---

[13] Ibid.
[14] Kant, *Groundwork*, p. 55 (4:413)
[15] Ibid. p. 57 (4:414)

imperative, the objective principle of reason, is a law of action which applies to one in virtue of something outside oneself; action from inclination according to hypothetical imperatives, therefore, is heteronomous, according to Kant, and so not (or at least not perfectly) free.[16]

To a large extent, this picture of agency is analogous to that of the Humean, despite their differences. In the case of instrumentally performed action, in accordance with hypothetical imperatives, which Kant thinks of as lacking moral worth, reason alone is insufficient to determine the will and produce action. It can only do so when its maxims incorporate empirically determined inclinational incentives. It is these incentives which provide the motivating force for the production of action, and they are able to do so only because they are non-cognitive, but rather conative, states. As empirically determined non-cognitive states of this kind, they would seem to be not, fundamentally, open to rational evaluation, like the Humean's desires.[17]

It is also plausible to think that these are the very features of inclinations which make them unsuitable, for Kant, to provide incentives for moral action. Kant thinks that for an action to have moral worth it must be performed out of duty alone, i.e. out of recognition that one's doing it conforms to the categorical imperative, the objective principle of pure practical reason, which represents an action 'as objectively necessary by itself, without reference to another end.'[18] The demands of morality, in other words, are for Kant the demands of pure practical reason as such. That the hypothetical imperative requires the addition of inclination in order to be binding on an agent rules out the imperative providing a constraint deriving from pure reason itself, and thus rules

---

[16] See, for example, Allison, *Kant's Theory of Freedom,* esp. p. 126; Reath, 'Kant's Theory of Moral Sensibility'; and Korsgaard, 'Kant's analysis of obligation', esp. p. 57, for interpretations of Kant as maintaining that we are free even when we act heteronomously, since incorporating an inclination into a maxim is an act of spontaneity and so a radically free act; we *choose* to act on the basis of empirically determined inclinations. It should be noted that the distinctions between freedom and determination, on the one hand, and autonomy and heteronomy, on the other, while related, are not equivalent. The former is, roughly, between it being possible or impossible to have done other than what one did, while the latter is between acting according to, and being bound by, a law one gives oneself, and acting according to a law that binds one in virtue of something else. Thanks to Tim Carter for urging me to make this clear.

[17] This interpretation of Kantian inclinations as providing motivating force should not be seen as implying a 'balance of forces' conception of the operation of desires in deliberation, choice, or action. For a recent defence of the interpretation of Kant as holding a balance of forces conception, see Richard McCarty, *Kant's Theory of Action*. For what I take to be more plausible arguments against such an interpretation, see Reath, 'Kant's Theory of Moral Sensibility', esp. pp. 13, 18; Allison, *Kant's Theory of Freedom*, pp. 125-7; and Korsgaard, *The Sources of Normativity*, lecture 3. See N. J. H. Dent, *The Moral Psychology of the Virtues*, pp. 96-102, for an argument that such a conception is incompatible with the possibility of practical deliberation.

[18] Kant, *Groundwork*, p. 57 (4:414)

out moral obligations being universal and rationally necessary. But additionally, Kant says:

> Because every practical law represents a possible action as good and hence, for a subject practically determinable by reason, as necessary, all imperatives are formulae for the determination of an action necessary according to the principle of a will that is good in some way. Now, if the action would be good merely as a means to *something else*, the imperative is *hypothetical*; if the action is represented as good *in itself*, hence as necessary in a will that in itself conforms to reason, as its principle, then it is *categorical*.[19]

We have seen that, according to Kant, an action is morally good – it has moral worth – just in case it is performed in accordance with the categorical imperative. But here Kant says that an action performed according to a hypothetical imperative, even if it is *compatible* with the categorical imperative, is 'good *merely* as a means to *something else*,'[20] but the implication is that bringing about the *something else* is only good to the extent that it satisfies some inclination. This is intelligible only if inclinations are themselves arbitrary, rather than rationally evaluable, since otherwise whatever made the inclination good would, it seems, also make bringing about the *something else* good, over and above its satisfying the inclination. But since Kant thinks that inclinations are empirically determined through sensibility, it makes perfect sense that he would think they were arbitrary; it is precisely their arbitrary character which seems to make them unsuitable to be the subjective ground of an action with moral worth.

If, for Kant, the broadly Humean character and aetiology of the inclinations is what disbars them from providing subjective grounds of moral action, but his practical philosophy operates within a broadly Humean framework which requires something with motivating force to provide an incentive and be incorporated into a maxim for action, then moral action requires something which plays *just the same role* as inclination. It would have to differ from inclinations, though, in just those respects that prevent their involvement in the generation of *moral* action.

Kant's notion of *respect* (*Achtung*) for the moral law is designed to achieve this; as Jens Timmermann explains, 'At the crossroads of motivation, reverence [*Achtung*] is the moral opponent of non-moral inclination. When the moral law speaks, reverence, like inclination, is available as an effective motive; but unlike inclination it is grounded

---

[19] Ibid.
[20] Ibid. first emphasis added.

in an objective, purely formal and universal law.'[21] Kant introduces this notion in his third proposition in *Groundwork I*:

> [*D*]uty *is the necessity of an action from respect for the law*. For the object as the effect of the action I have in mind I can indeed have *inclination*, but *never respect*, precisely because it is merely an effect and not activity of a will [....] Only what is connected with my will merely as ground, never as effect, what does not serve my inclination, but outweighs it, or at least excludes it entirely from calculations when we make a choice, hence the mere law by itself, can be an object of respect and thus a command. Now, an action from duty is to separate off entirely the influence of inclination, and with it every object of the will; thus nothing remains for the will that could determine it except, objectively, the *law* and, subjectively, *pure respect* for this practical law, and hence the maxim of complying with such a law, even if it infringes on all my inclinations.[22]

Kant here distinguishes several ways in which the object of respect differs from that of inclination. The object of an inclination is an intended effect of an action, and so an effect of the will; the object of *respect* is rather an 'activity of a will' itself, or, as Kant puts it, 'what is connected with my will merely as ground'.[23] The intended effect of an action serves inclination, while the 'ground' of a will outweighs inclination or excludes it from deliberative consideration, so the (moral) law itself is the object of respect. But, finally, Kant says that respect is the only remaining subjective determinant of the will once all inclination has been excluded, while its object, the law, is its objective determinant. Like inclination, then, respect for the law is a subjective determining ground of the will, and so of action.

Kant elaborates on the notion of respect for the law in a footnote:

> I might be accused of using the word *respect* just to seek refuge in an obscure feeling, instead of giving distinct information about the matter in question by means of a concept of reason. But even though respect is a feeling, it is not one *received* by influence, but one *self-wrought* by a rational concept and therefore specifically different from all feelings of the former kind, which come down to inclination or fear. What I recognize immediately as a law for myself I recognize with respect, which signifies merely the consciousness of the *subordination* of my will to a law, without mediation or other influences on my sense. The immediate determination of the will by the law and the consciousness of this is called *respect*, so that it is viewed as the *effect* of the law on the subject and not as its *cause*.[24]

---

[21] Jens Timmermann, *Kant's Groundwork of the Metaphysics of Morals: A Commentary*, p. 40.
[22] Kant, *Groundwork*, p. 29 (4:400)
[23] As Henry E. Allison, *Kant's Groundwork for the Metaphysics of Morals: A Commentary*, p. 128, says, Kant introduces 'the concept of respect […] as the subjective motivating ground or the *principium executionis*, with the still not specified practical law serving as the objective ground of the will or its *principium diudicationis*.'
[24] Kant, *Groundwork*, p. 31 (4:401)

As has been noted by others, Kant is clearly aware that his appeal to the feeling of respect as a subjective determining ground of the will, a state possessing motivating force, is liable to appear merely *contrived*, in an attempt to vindicate morality as resting on purely rational grounds.[25] In the second sentence, though, Kant admits that respect is a feeling. Given his adoption of a broadly Humean framework of agency, it seems that this is required, since it is feelings, which Hume thought of as *passions*, which are supposed to be motivating forces.[26] But Kant insists that respect is sufficiently unlike other feelings to avoid the problems of thinking of them as motivating moral action. As Timmermann puts it, '[Respect] is not an epiphenomenon of morality, but rather that which *motivates* moral action. The *motive that can give actions moral worth* is neither fear nor inclination but "simply reverence for the law", as Kant quite explicitly puts it'.[27]

In an attempt to explain the salient differences between the feeling of respect and all other feelings, Kant draws two distinctions (although he also runs them together). On the one hand, he says that respect is *self-wrought* where other feelings are *received*. This is a difference in aetiology and corresponds to Kant's distinction between autonomy and heteronomy. On the other hand, he says that respect is self-wrought *by a rational concept*, and also describes it as the '*subordination of my will to a law*' – a law which is also a 'concept of reason' – while other feelings come from the *influences of sense*, i.e. they are empirically determined. This is not a difference merely between their origins, but between being rationally evaluable or not.[28] Kant's emphasis is clearly on the difference in aetiology, but it is actually the difference between an incentive which is essentially rationally evaluable and one which is not which is important, and which does the philosophical work for him.

---

[25] See e.g. Allison, *Kant's Groundwork for the Metaphysics of Morals*, p. 132; Timmermann, *Kant's Groundwork of the Metaphysics of Morals*, p. 41

[26] Kant mustn't think of all feelings as passions, since passions are supposed to be just that: *passive*, as the etymology suggests. Kant needs the feeling of respect to not be passive, but rather 'self-wrought' or spontaneous.

[27] Timmermann, *Kant's Groundwork of the Metaphysics of Morals*, p. 42, second emphasis added.

[28] Compare Korsgaard, 'The Normativity of Instrumental Reason', p. 219: 'According to the Kantian conception, to be rational *just is* to be autonomous. That is: to be governed by reason, and to govern yourself, are one and the same thing. The principles of practical reason are *constitutive* of autonomous action: they do not represent external *restrictions* on our actions, whose power to motivate us is therefore inexplicable, but instead *describe* the procedures involved in autonomous willing. But they also function as normative or guiding principles, because in following these procedures we are guiding ourselves.'

Kant thus makes room in his philosophical system for one *feeling* that is subject to norms, and so is rationally evaluable.[29] But at the same time he imposes two inessential restrictions that we must examine, and, I will argue, reject. The first restriction is that *only a moral action* is grounded in, or motivated by, a rationally evaluable incentive. As he says, therefore, respect is a feeling which ought to be had *only* towards the law – to what is, on his view, good in itself.[30] In effect, Kant's claim is that:

(K1)   A motive is rationally evaluable only if it is a moral motive.

If we were to reject (K1), we could avoid the schism between moral and non-moral action, and, correspondingly, also avoid the implausible view that incentives of wholly different natures operate in the two (alleged) domains. What is more, there seems no good reason to accept (K1). All Kant offers in its support is a broadly Humean view of inclinations as brute hedonistic impulses, which are not themselves rationally evaluable. As we have already seen in Chapters 1 and 2, however, we do not have good grounds for accepting anything like such an Humean view. We ought, therefore, to reject (K1).

The second restriction is on the source of the norms constraining a rationally evaluable incentive. Kant's claim is that an incentive may be rationally evaluable only in terms of the structure of reason itself. The problem with rationally arbitrary incentives is a problem with the idea that there is no normative constraint from something external to the psychological states of agents themselves, in terms of which they may be rationally evaluated. Kant attempts to provide such a normative constraint, which is independent of the *actually existing* psychological states of any agent, but yet is internal to the structure of reason as such. And he tries to show how this normative constraint is practically relevant to actual agents and applies to them, by showing that it is possible for the *motivating ground* of one's actions to be subject to norms, and so to be rationally evaluable.

This is bound up with his discussion, in *Groundwork III*, of the necessity of regarding ourselves as transcendentally free. Since Kant thinks that in order to be free

---

[29] That Kant (implicitly) makes room for the possibility of a rationally evaluable incentive is noted by Skorupski, *The Domain of Reasons*, p. 27

[30] Strictly, Kant claims in the *Groundwork* that it is *impossible* for respect to have any other object.

we must have the power of autonomy, to give the practical law to ourselves, he thinks that the source of any normative constraint must lie in the structure of our own faculty of reason. But this is an independent claim of Kant's about the particular difference in aetiology between respect and other feelings which is necessary in order for respect to be rationally evaluable. To be sure, it is important for Kant's project, since it is plausible to think that it is precisely this restriction which generates the constructivist character of his moral theory.

Actually, though, all we need to make sense of moral action is to accept the following two claims that Kant makes plausible:

(K2)   It is possible for a motive to be rationally evaluable

(K3)   If something is a moral motive, then it is rationally evaluable.

(K2) and (K3), I think, are correct, and are the real lessons from Kant's moral psychology. But there is no good reason to accept Kant's second, constructivist, restriction:

(K4)   If a motive is rationally evaluable, then it is derived from the structure of Reason.

Without (K4), though, there is no good motivation for Kant's view about moral motivation in particular, his view that:

(K5)   If something is a moral motive, then it is derived from the structure of Reason.

(K3) and (K4) together entail (K5), but, since we should reject (K4), we must, *a fortiori*, reject Kant's argument for (K5). Once we reject (K1) and (K4), though, we can see that we do not need Kant's *sui generis* feeling of respect for the moral law as the moral analogue of inclination. Learning from Kant's insights in (K2) and (K3), we should reject (FAD) and, in its place, accept that:

(DRE) Desire is rationally evaluable.[31]

We can then regard desire as suitable to motivate moral action, without subscribing to a rationalist ethical theory.[32]

3.3 Belief and the Constitutive Norm of Truth

It is uncontentious that:

(BRE) Belief is rationally evaluable.

(BRE), it seems to me, is most plausibly explained by the fact that belief is subject to a constitutive norm of truth. Just because of this, (BRE) provides the most powerful motivation for thinking that there is indeed such a norm.[33]

A common first proposal for a formulation of the truth norm of belief is the following:

(TN) For any A and any p: A ought to believe that p if and only if it is true that p.[34]

As has been widely noted, (TN) is too strong. It requires one to believe *every* true proposition, which is highly implausible, since there are innumerable trivial truths, and also impossible, since there are infinite true propositions, some of which are of

---

[31] At least to some extent, something like (DRE) is accepted by e.g. Alvarez, *Kinds of Reasons*; Hyman, *Action, Knowledge, and Will*; Korsgaard, 'The Normativity of Instrumental Reason'; Nagel, *The Possibility of Altruism*; Scanlon, *What We Owe To Each Other*; Schroeder, *Slaves of the Passions*; Skorupski, *The Domain of Reasons*; Wallace, 'How to Argue about Practical Reason'.

[32] Note Skorupski's comment: 'Feeling [e.g. an incentive] has the potential for self-determination [i.e. admits of reasons] just as belief and action do. Failure to recognise this forces Kant into an unpalatable choice—either to adopt a non-Critically intuitionistic view of value, or (as he does) to dispense with value as a *basic* category in his theory of practical reason altogether. He thus distorts the content of morality, ignores reasons other than moral requirements, enters into contortions about the rationality of respect, and—joining a strange alliance with Hume—destroys the possibility of a non-instrumentalist account of hypothetical imperatives' (*The Domain of Reasons*, p. 27).

[33] For a survey of other recent motivations for endorsing a truth norm of belief, see Conor McHugh and Daniel Whiting, 'The Normativity of Belief'. Some argue that belief constitutively *aims* at truth, interpreting this teleologically, rather than normatively. I do not consider this view here, although it is similar to the higher-order attitude account of mind-to-world direction of fit rejected in §2.9.

[34] Variants of this formulation attempt to capture it as evaluative, rather than deontic, or as neither, e.g. as 'It is good for A to believe…' or 'It is correct for A to believe…'. See e.g. McHugh, 'The Truth Norm of Belief'; Wedgwood, 'The Aim of Belief'. It is sometimes claimed that belief is subject to a constitutive *knowledge* norm, e.g. Timothy Williamson, *Knowledge and Its Limits*, p. 47; a knowledge norm would imply a truth norm.

indefinite length, and so beyond our cognitive capacities to grasp. (TN) therefore flouts the widely accepted principle that *ought* implies *can*.[35]

There are some familiar but ultimately unsuccessful strategies to overcome these problems. In this section I will comment on two of these: (i) reformulating the truth norm as a conditional norm, and (ii) interpreting 'ought to' in (TN) as taking wide scope.

(TN) can be reformulated as a conditional norm to restrict it to contexts of deliberation, thereby avoiding the earlier problems.[36] The simplest conditional formulation:

(CTN)

For any *A* and any *p*: if *A* considers whether *p*, then *A* ought to believe that *p* if and only if it is true that *p*

is unsatisfactory because it cannot accommodate so-called 'blindspot' propositions such as:

(BS)   It's raining and no-one believes that it is raining,

whose truth is dependent on not being believed, but which can be considered.[37]

A more complex conditional formulation might be:

(CTN*)

For any *A* and any *p*: if *A* considers whether *p*, then *A* ought to believe that *p* if and only if *p* would be true were *A* to believe that *p*.

(CTN*) is also unacceptable. An indefinite number of propositions would be trivially true were one to believe them, because they would be true in virtue of one believing them. For example:

(DS)   I believe that a conspiracy of mice secretly controls the world.

[35] Cf. Whiting, 'Does Belief Aim (Only) at the Truth?', pp. 283-4; Pascal Engel, 'Is truth a norm?', pp. 45-6, and 'In Defense of Normativism About the Aim of Belief', p. 45.
[36] Cf. Krister Bykvist and Anandi Hattiangadi, 'Does Thought Imply Ought?', p. 280; Wedgwood, 'The Aim of Belief', p. 273; Engel, 'In Defense of Normativism About the Aim of Belief', p. 45.
[37] Bykvist and Hattiangadi, 'Does Thought Imply Ought?', pp. 281-2

On some plausible assumptions, (CTN*) would imply that if I were to consider whether (DS), then I ought to believe it. For that matter, it makes no difference whether the *de se* proposition concerns a plausible belief or not. *Whatever* followed 'I believe that…' in (DS) it would be the case that (CTN*) prescribed that one who considered it ought to believe it.[38] A truth norm of belief is supposed to ensure that cognition is subject to normative constraint by how the world is independently, but a norm which required one to believe any proposition which would be true in virtue of believing it, whenever one considered it, would not provide a suitably external constraint.[39] It is because of this that (CTN*) would imply a requirement to believe obviously frivolous and false propositions such as (DS).

The second strategy notes that (TN) contains a scope ambiguity:

(TN$_n$) For any A and any p: A ought to (believe that p) if and only if it is true that p.

(TN$_w$) For any A and any p: A ought to (believe that p if and only if it is true that p).

The objections canvassed thus far depend on interpreting the normative operator 'ought to' as taking narrow scope. This is because, as Broome has emphasised, a normative operator taking wide scope over a conditional is not 'detachable'.[40] From (TN$_w$) and it's being true that p, therefore, it does not follow that one ought to believe that p. One can satisfy its normative requirement in two ways: by believing that p when it's true that p, or by not believing that p when it's false that p. Similarly, it prohibits two combinations: that one believe that p when it's false that p, and that one not believe that p when it's true that p.

Bykvist and Hattiangadi make two complaints against (TN$_w$), and while I think the second is a decisive objection, the first is not compelling but needs careful handling before we can see what is required of a satisfactory formulation of the truth norm. Taking them in reverse order, their second, decisive objection is that (TN$_w$) cannot deal

---

[38] Bykvist and Hattiangadi reject a variant of (CTN*) claiming that it is too weak, since it doesn't prescribe that one *ought not* to believe blindspot propositions, and, if combined with a principle which does make such a prescription, it leads to contradictions (ibid. pp. 282-3).

[39] Cf. McDowell, *Mind and World*.

[40] John Broome, 'Normative Requirements'

with all cases, for example the conjunction of all the necessary truths. According to
(TN_w) one is obligated 'either to bring it about that p [the conjunction of all necessary
truths] is false or bring it about that you believe that p'.[41] Since this instance of 'p' is
necessarily true, the first disjunct is impossible, but, because 'p' is so complex, it is
ungraspable, so the second disjunct is impossible. (TN_w), therefore, falls foul of the
ought implies *can* principle.

Bykvist and Hattiangadi's first objection is that (TN_w) 'does not capture the
thought that the truth is what one ought to believe, or that false belief is faulty or
defective'.[42] Let us first separate out their objection's two components:

     (T)   The truth is what one ought to believe.

and

     (F)   False belief is faulty or defective.

That (TN_w) fails to capture (F) is quite clearly a legitimate and compelling
complaint against it. That it fails to capture (T), though, is not obviously a bad thing,
since (T) does not discriminate between, e.g., important, relevant or salient truths with
which one is confronted, and trivial, blindspot or indefinitely complex truths. On the
most charitable interpretation, however, the motivation behind (T) is the thought that
(TN_w) fails to prescribe that one believe enough truths, because it does not prescribe
that one believe *even one* truth. What they are looking for, on this interpretation, is a
way of capturing the thought that one must believe some truths, so that one cannot
withhold belief altogether while satisfying the normative demands of belief. In other
words, pyrrhonism must not be an option.

This objection, however, is not decisive, for there are at least three possible
ways to rule out pyrrhonism: (i) some better formulation of the norm of belief might be
able to accommodate (T) directly, (ii) it might be accommodated separately from giving
an adequate formulation of the norm of belief, or (iii) it might be impossible for an
agent to lack any beliefs at all, and so be impossible to withhold belief altogether while
satisfying the normative demands of belief. If (iii) were to follow from an adequate
formulation of the norm of belief, that would mean that the norm would accommodate
(T), but indirectly.

---

[41] Bykvist and Hattiangadi, 'Does Thought Imply Ought?', p. 284
[42] Ibid.

I have no knock-down argument that (i) is impossible, but it certainly faces tremendous difficulties. Intuitively, which propositions one ought to believe depends, not only on their truth, but also on a large quantity of highly complex variables (e.g. which propositions one considers, what experiences and knowledge one has, what one's conceptual resources are, what situations one finds oneself in, etc.). It is implausible that a norm is formulable which could take account of all the relevant variables, issuing exactly the right prescriptions for every proposition, agent, and circumstance. In fact, (iii) seems to me the most promising line of response. In the next section, I will defend what I think is a more satisfactory formulation of the norm of belief, before explaining how I think the anti-pyrrhonist intuition behind (T) may be accommodated.

3.4 The Truth Norm of Belief: The Proposed Formulation

In all its variations, the truth norm we have been considering has been formulated as a bi-conditional. We can break it down into its component conditionals:

(TN_{nec}) For any A and any p: A ought to (believe that p) only if it is true that p.

(TN_{suf}) For any A and any p: if it is true that p, then A ought to (believe that p).

In each of the variations we have considered, it is (TN_{suf}) which has been problematic. (TN_{nec}), though, only imposes a *constraint* on what one ought to believe, imposing *no* prescriptions. Because of this, Bykvist and Hattiangadi also reject (TN_{nec}) as the norm of belief. In particular, they object to the fact that 'It does not follow, from the falsity of p, that S ought not to believe that p', since lacking an obligation to believe that p is different to having an obligation not to believe that p.[43] But while they are correct that this rules out (TN_{nec}) as the norm of belief, it does not show that it has no place in such a norm.[44] We can simply combine it with a further norm which prohibits believing falsehoods, thereby capturing (F):

(FN)   For any A and any p: if it is false that p, then A ought to (not believe that p).

---

[43] Ibid. p. 280
[44] Whiting, 'Does Belief Aim (Only) at the Truth?', argues for a teleological variant of (TN_{nec}) as the aim of belief.

One might object to (FN) on the grounds that it is not a norm of truth but of falsity. So we might instead try:

(~TN) For any A and any p: if it is not true that p, then A ought to (not believe that p).[45]

(~TN) is inadequate alone as the norm of belief, since, for all it says, it is possible for one to refuse to believe anything whatsoever, thereby satisfying it. We saw that we must rule out that sort of pyrrhonist view, but if we take the combination of (TN_{nec}) and (~TN):

(NB) For any A and any p: (i) A ought to (believe that p) only if it is true that p;

and,

(ii) if it is not true that p, A ought to (not believe that p)

as the norm of belief, we can accommodate (F), as well as the anti-pyrrhonist intuition behind (T), or so I shall argue. We should note, however, that (NB) does not itself impose any prescription as to which truths one ought to believe, leaving it open which truths particular individuals have obligations to believe. That is just what we should want, since such obligations are relative to the complex set of conditions in which one finds oneself. It is a virtue of (NB) that it does not state that everyone is under exactly the same obligations to believe exactly the same truths.

3.5 The Truth Norm and Anti-Pyrrhonism

It is all but a platitude to say that:

(B) What it is to believe that p is to have a view as to how things are, namely that the way things are is that p.

---

[45] There are some complications facing (~TN) regarding possible truth-value gaps. There may well be ways of dealing with such complications, but I cannot consider the matter here.

Despite (B)'s truistic appearance, however, one might think that (1) it is a mere notational variant of the claim that to believe that *p* is to regard *p* as true, and one might also be impressed by Shah and Velleman's claim that (2) regarding a proposition as true is insufficient to distinguish belief from other attitudes. Shah and Velleman argue that the characterisation of an attitude as one of regarding a proposition as true, fits not just the attitude of believing, but also those of assuming, supposing, and imagining.[46] (For example, assuming that *p* is regarding *p* as true, for, say, the purposes of *reductio*, and imagining that *p* is regarding *p* as true for, say, the sake of a game.) If *both* (1) and (2) were true, though, (B) might appear on closer inspection actually to be false!

[X] Suppose that (1) is true. Then, for belief, 'regarding *p* as true' must be interpreted as taking a view as to how things are, namely that *p*. But for a plausible interpretation of, e.g. assumption, as an attitude of regarding *p* as true, 'regarding' needs to be interpreted as having the force of 'accepting', or perhaps 'presuming' (for instance, for the sake of argument). Interpreted that way, it does seem correct to say that assuming that *p* is regarding *p* as true, but it is *not* the same interpretation of 'regarding *p* as true' that is needed for (1), so (2) is false. In effect, cases of assumption would be, roughly, taking it *as if* how things are is that *p*, without that necessarily being the view as to how things are that one takes.

[Y] If one were able to find a suitable interpretation of 'regarding *p* as true' that fitted belief as well as assumption etc., it might appear that (2) would be true. But even then, my account could distinguish belief from these other attitudes in normative terms. For we could say that, if one has an attitude of regarding *p* as true which is subject to (NB), then the attitude is a belief. This would adequately distinguish belief, since, e.g. there is nothing wrong with assuming that *p* when *p* is false, for instance for the purposes of *reductio*. So belief would be the attitude of regarding *p* as true which was subject to the norm (NB), while assuming and the rest would be attitudes of regarding *p* as true that were not subject to (NB).[47]

If the conclusion of paragraph [X] were right, and if it is right that a truth norm is constitutive of belief, then this truth norm would be constitutive of having a view as to how things are, or of regarding a proposition as true. That belief is characterisable

---

[46] Nishi Shah and J. David Velleman, 'Doxastic Deliberation', p. 497; cf. J. David Velleman, *The Possibility of Practical Reason,* pp. 110, 248-50.

[47] This should not be interpreted so as to imply that there would be a generic attitude of regarding *p* as true, with belief and assumption etc. being distinguished by having further (normative) *differentiae*. Belief is not what Ford calls an 'accidental species' ('Action and Generality', p. 83).

according to (B), then, would be due to (NB), if (as I have arguead) this is in fact the norm of belief.[48] If, on the other hand, the view in [Y] were right, then having a view as to how things are, namely that *p* (or regarding *p* as true) which is subject to (NB), is believing that *p*. (B), however, would not itself be due to (NB), but together they would give a natural, but not quite platitudinous, characterisation of belief.

Since, as yet at least, no interpretation of 'regarding *p* as true' which fits belief as well as assumption etc. is forthcoming, we may provisionally accept the conclusion of [X], and take (B) to give a platitudinous characterisation of belief, in virtue of (NB) being its constitutive norm.

If, as I have claimed, it is a platitude to characterise belief according to (B), however, then an argument for its being impossible to withhold belief altogether seems to be available, and so we seem to be in a position to accommodate the anti-pyrrhonist intuition behind (T). Consider our capacity for perceptual experience. When one has a perceptual experience, one thereby takes a view as to how things are. A contentious interpretation of this (which I think is correct) claims that in having a perceptual experience one takes a view as to how things are in the external world one perceives, and when matters are aright, one thereby takes in how things are, and so gains perceptual knowledge of how things are in the world one inhabits.[49] A less contentious interpretation (which I think is also correct) claims that in having a perceptual experience one takes a view at least as to how things seem, which is still a (subjective) way things are. According to (B), though, in taking a view as to how things are, even if the way things are in question is that things seem a particular way, one has formed a belief. Since, inevitably, if one is an agent one has perceptual experiences of some kind or other, it is impossible for an agent to not have any beliefs at all, not even beliefs about how things seem to them.[50] A pyrrhonist accommodation of the normative

---

[48] The claim that the legitimacy of characterising belief according to (B) is due to (NB) deserves much fuller argument than I provide here. The departure point for such a fuller case, it seems to me, would be the idea that it is in virtue of the fundamental involvement of *truth* in belief, through its constitutive truth-norm, that belief bears on the world at all, together with the idea that in specifying what is believed, one specifies a condition under which the belief is true – a condition of *how thing would be* in order for it to be true. Cf. John McDowell, 'In Defence of Modesty', pp. 88-9

[49] This is part of what is involved in a disjunctive conception of perceptual experience. See e.g. John McDowell, 'Criteria, Defeasibility, and Knowledge', p. 9. The claim in the text should be qualified, as holding at least absent special circumstances, such as that one takes there to be special reasons to think one might be hallucinating, or something similar.

[50] In fact, being an agent requires one's having beliefs, since otherwise there would be no possibility of psychological explanation of what one *did*, and without this possibility, there would be no possibility of one's doing something at all, i.e. one could not be an agent. Cf. M. F. Burnyeat, 'Idealism and Greek

demands of belief is therefore impossible; one cannot satisfy (NB) by withholding belief altogether, just because one cannot withhold belief altogether at all. Since the impossibility of withholding belief altogether was justified by way of characterising belief by the platitudinous (B), and since, I argued, the legitimacy of (B) is due to (NB), the conclusion depends on the norm of belief (NB); so the norm of belief accommodates the anti-pyrrhonist intuition behind (T), but indirectly. Inevitably, agents have beliefs, and of the beliefs they end up with, they ought to have them only if they are true, and if they are not true, they ought not to have them.

3.6 Is the Truth Norm Normative?

Let us briefly consider one last sort of objection. It might be claimed that there is no good sense in which (NB) is genuinely normative. Such an objection could, perhaps, be based on Kathrine Glüer and Åsa Wikforss's 'no guidance argument', which purports to show that a truth norm of belief cannot provide guidance to an agent's belief-forming behaviour.[51] Glüer and Wikforss themselves are cautious in what they take their argument to establish, saying that it 'is not directed against the validity of the truth norm in general, but targets specifically the idea that such a norm, if valid, would provide guidance for belief formation'.[52] But one might think that unless a norm is capable of providing guidance, in some intuitive sense, it is no genuine norm at all.

Such a claim, though, would be wrong; we can give a perfectly good account of what is genuinely normative about the truth norm (NB), even if it could provide no guidance for belief-formation. The sense in which (NB) is normative is the sense in which it sets *standards of evaluation* for belief. Belief is evaluable in terms of truth in various ways, and (NB) sets out just what those ways are. It is not just that we *can* evaluate beliefs in terms of truth although we might not, in the way that we can evaluate someone's eating habits in terms of rules of etiquette, but need not impose these external standards. It is intrinsic or essential to belief, part of what belief itself *is*, that it is evaluable in terms of truth. That is at least partly what it means for (NB) to be a *constitutive* norm. It is precisely because truth, as it figures in (NB), is the constitutive

---

Philosophy: What Descartes Saw and Berkeley Missed', esp. p. 30; and Jennifer Hornsby, 'Agency and Actions', pp. 6-8

[51] See Kathrin Glüer and Åsa Wikforss, 'The Truth Norm and Guidance: a Reply to Steglich-Petersen', and 'Still No Guidance: Reply to Steglich-Petersen'.

[52] Glüer and Wikforss, 'The Truth Norm and Guidance', p. 758

norm of belief, which sets standards of evaluation for it, that (BRE) is true. Belief is rationally evaluable in terms of truth, as set out by (NB).

3.7 The Constitutive Norm of Desire

Having now clarified in some detail my normative conception of the nature of belief, let us turn to desire. Given what I have just said about belief, we should expect (DRE) to be explained by the fact that desire is subject to some constitutive norm. Though this is an idea which has received relatively little attention, it is not new.[53] In particular, though not put explicitly in terms of constitutive norms, it is an idea that is central to some of the attempts we examined in Chapter 2 to formulate the distinction between belief and desire in terms of a supposed difference between their 'directions of fit'. For example, Platts, introducing the supposed distinction, says:

> Desires aim at realisation, and their realisation is the world fitting with them; the fact that the indicative content of a desire is not realised in the world is not yet a failing *in the desire,* and not yet any reason to discard the desire[54]

And Searle writes:

> [I]f I fail to carry out my intentions or if my desires are unfulfilled I cannot […] correct the situation by simply changing the intention or desire. In these cases it is, so to speak, the fault of the world if it fails to match the intention or the desire[55]

As I suggested in §2.7, these comments are most naturally interpreted as characterising desire in terms of constitutive norms. We might characterise what both are getting at by saying that they take there to be a constitutive norm of satisfaction for desire, so that there is something wrong or defective with desiring something without one's desire being satisfied, or that one ought to bring about what one desires. But Platts and Searle give a fundamentally distorted view of the norm of desire; it is just obvious

---

[53] Engel passingly speculates about there being normative correctness conditions for other epistemic attitudes than belief, for example conjecture and certitude, but does not consider the possibility of norms of conative or orectic attitudes ('In Defense of Normativism About the Aim of Belief', pp. 33-4, fn. 2). McHugh and Whiting suggest the possibility of giving a normative account of desire and other conative attitudes, but do not go any farther than this ('The Normativity of Belief', p. 710). Michael Morris develops such an account, but his formulations are not ultimately satisfactory (*The Good and the True,* ch. 11.1).

[54] Platts, *Ways of Meaning,* pp. 256-7

[55] Searle, *Intentionality*, p. 8

that the proposed norms of desire are false. Roughly, one should only satisfy one's desire if what one desires is good, or at least permissible.[56]

This indicates that a better suggestion would be that desire is subject to a constitutive norm of goodness, rather than of satisfaction. Indeed, Sergio Tenenbaum goes so far as to suggest that the distinction in respect of direction of fit should be understood in terms of belief and desire being subject to the contrasting constitutive norms of truth and goodness.[57] But he does not provide any detailed proposal as to exactly how those norms should be formulated.

Let us suppose that this sets us along the right lines.[58] We may follow the path of our discussion of the truth norm of belief, and try an initial formulation of the norm like this:

> (GN) For any A and any p: A ought to desire that p if and only if it would be good if p.

Just like (TN), (GN) has a scope ambiguity:

> (GN$_n$) For any A and any p: A ought to (desire that p) if and only if it would be good if p.

> (GN$_w$) For any A and any p: A ought to (desire that p if and only if it would be good if p).

Let us begin by considering (GN$_n$). Analogously to (TN$_n$), it implies that, for any p, if it would be good if p then one ought to desire that p, but this is absurd, and quite likely impossible.[59] It is plausible that an indefinitely large range of things would be

---

[56] Cf. Humberstone 'Directions of Fit', p. 67; Alvarez, *Kinds of Reasons*, p. 69; Morris, 'Mind, World and Value', pp. 306-7

[57] Sergio Tenenbaum, 'Directions of Fit and Motivational Cognitivism', pp. 247-8. The accounts of belief and desire directly in terms of the constitutive norms of truth and goodness are adequate, and such a direction of fit construal of them is at best superfluous.

[58] Plausibly, analogous considerations to those which some take to favour a knowledge norm of belief would favour the view that, roughly, one should only satisfy one's desire if one knows that what one desires would be good, or at least permissible. This would presumably support taking the norm of desire not to be simply a norm of goodness, but of knowledge of goodness. The formulations which follow, and the argument they provide, would have to be modified accordingly. Cf. fn. 34 above.

[59] Henceforth, I will use the phrase 'x would be good', rather than the more cumbersome 'for some (any) p, it would be good if p'. Strictly, though, I should be understood according to the latter formula.

good, and, while that is not quite the same as the range being infinite, it still is implausible that it is within our limited capacities to have the same indefinitely large number of desires. Furthermore, there are many things that would be good, but only very minimally good, for instance it would be good if the third person born in Dresden this month was not born in a room with a squeaky hinge, and it would be good if market stall price signs stopped incorrectly using apostrophes. It would be ludicrous, though, to insist that *everyone* ought to desire these things![60] Finally, there are many things that would be good but which are incompatible, so one could only desire them together by having incompatible desires. Surely there can be no such obligation![61]

One might try the move to a conditional norm:

(CGN) For any *A* and any *p*: if *A* considers whether to desire that *p*, then *A* ought to (desire that *p*) if and only if it would be good if *p*.

This helps to avoid the first problem with (GN$_n$), but it fails to avoid the others. If one considers whether to desire that *p*, then it may well be true that it would be *permissible* to desire that *p* if and only if it would be good if *p*. It even seems to be a consequence of the Guise of the Good thesis that desiring that *p* would be *intelligible* if and only if one *took* it that it would be good if *p*.[62] But it is not the case that, if one considers whether to desire that *p*, one ought to desire that *p* if it would be good if *p*. For example, that would not be credible when it would be only very minimally good if *p*, and, since there is nothing to stop one considering whether to desire that *p*, and whether to desire that *q*, when *p* and *q* are incompatible, accepting (CGN) would unacceptably entail that in such circumstances one ought to have incompatible desires.

Nor would the following amended version help:

(CGN*)

For any *A* and any *p*: if *A* considers whether to desire that *p*, then *A* ought to (desire that *p*) if and only if it would be good if *p*, and were *A* to desire that *p*, *A* would not have any incompatible desires.

---

[60] It is not of course ludicrous to think that *someone* ought to desire one or another of such things.

[61] This is not to say that one *ought to not* have incompatible desires.

[62] According to the Guise of the Good thesis, one always desires something *sub specie boni*; for a good overview, see Sergio Tenenbaum, 'Guise of the Good'. A version of the Guise of the Good thesis is a consequence of the argument in §3.9 below.

(CGN*) still fails to deal with the problem of desiring minimally good things, and it also faces a problem with *de se* desires. For example, suppose that someone is distressingly ill, and considers whether to have the higher-order desire that they desire to be distressingly ill. Of course, it would *not* be good overall if they desired that they be distressingly ill, but may nevertheless be a *pro tanto* good, since it would, let us suppose, reduce their dissatisfaction with their lot. But it would be appalling to claim that they *ought* to desire that they be distressingly ill; very likely they ought to see a doctor.

(GN$_w$) fares no better than (TN$_w$). The latter implies that one ought to believe the conjunction of all necessary truths. As for (GN$_w$), because it is impossible to make it the case that only a finite number of things are good, all of which are compatible, it will still imply that one ought to desire incompatible things; very likely it will imply that one ought to desire an indefinitely large range of good things which it is beyond one's capacities to desire all together. (GN$_w$) also falls foul of an analogous intuition to (T), namely:

(G)   The good is what one ought to desire.

Just as (TN$_w$) could not accommodate (T), because what it prescribes are combinations, (GN$_w$) cannot accommodate (G), because it too prescribes combinations.

The problems facing the various formulations of (GN), it should be clear, very closely mirror those which arise for (TN). The solution, I propose, should also mirror that which I gave for the truth norm of belief, namely (NB). We should formulate the constitutive norm of desire as follows:

(ND)   For any *A* and any *p*: (i) *A* ought to (desire that *p*) only if it would be good if *p*;
and,
(ii) if it would not be good if *p*, *A* ought to (not desire that *p*).

3.8 Desire as Evaluative Belief

We are now in a position to give an argument for identifying desires with evaluative beliefs, i.e. for the view that to desire that p just is to believe that it would be good if p. We may begin by re-emphasising the fact that both (NB) and (ND) are supposed to be constitutive of their respective psychological attitudes, which means that they spell out what it is for an attitude to be a belief or to be a desire. We now have our constitutive norm of desire, but what we still require is a statement not just of the norm of belief, but of the norm of *evaluative* belief. So let us consider specifically an evaluative belief – a belief with a content clause of the general form 'it would be good if p'.[63] The following is an example of the norm which would apply to a specific evaluative belief:

(NB$_{sch}$) For any A: (i) A ought to (believe that it would be good if schools were properly funded) only if it is true that it would be good if schools were properly funded;

and,

(ii) if it is not true that it would be good if schools were properly funded, A ought to (not believe that it would be good if schools were properly funded).

We can generalise from this to give a statement of the norm of evaluative belief as such:

(NEB) For any A and any p: (i) A ought to (believe that it would be good if p) only if it is true that it would be good if p;

and,

(ii) if it is not true that it would be good if p, A ought to (not believe that it would be good if p).

We can then take the equivalence schema for the truth operator:

(E$_o$)   It is true that p if and only if p

and apply it to both clauses of (NEB) yielding the following:

---

[63] When I speak of evaluative belief, I mean specifically a belief of this form. No doubt there are other sorts of evaluative belief, but they do not concern us here.

(NEB*)

> For any *A* and any *p*: (i) *A* ought to (believe that it would be good if *p*)
> only if it would be good if *p*;
>
> *and,*
>
> (ii) if it would not be good if *p*, *A* ought to (not believe that it would be
> good if *p*).

An argument for identifying desires with evaluative beliefs must identify an attitude which is specifiable schematically as *the desire that p,* with an attitude which is specifiable schematically as *the belief that it would be good if p.* Notice that the grammar of the content clauses of the desire and of the evaluative belief have not been represented in the same way. We have represented both the desire and the evaluative belief as taking propositional content-clauses. But we have represented the desire's content-clause as taking a basic sentence, while we have represented the evaluative belief's content-clause as taking a more complex sentence formed by a basic sentence prefixed by the evaluative operator 'it would be good if…'. Because of this difference, we should not expect to be able to substitute the same sentences in the propositional variable positions in both (ND) and (NEB*).[64]

Because of this, what we need to do is to show that the conditions for the applicability of the norm (ND) are exactly the same as the conditions for the applicability of (NEB).[65] And, indeed, that is exactly what we have just done. This can easily be seen if we remove the relevant components of (ND) and (NEB*) – the components which concern the attitude and content form in each. What we get is:

(ND!) For any *A* and any *p*: (i) *A* ought to (…) only if it would be good if *p*;
> *and*,
>
> (ii) if it would not be good if *p*, *A* ought to (not…)

and

---

[64] As we saw in chapter 2, strictly, the content-giving clause in a desire specification will be in the subjunctive, rather than the indicative, mood, or else be an infinitival phrase. The following argument, however, does not depend upon my simplifying move of treating desire specifications as containing a propositional clause to be filled by a complete sentence.
[65] Cf. Morris, *The Good and the True,* p. 219

(NEB*!)

> For any *A* and any *p*: (i) *A* ought to (…) only if it would be good if *p*;
> *and,*
>
> (ii) if it would not be good if *p*, *A* ought to (not…).

(ND) and (NEB*), in other words, are precisely the same norms. They are constitutive norms, spelling out what it is for an attitude to be a desire or an evaluative belief, so, on the basis of the principle:

(SN-SA)

> The constitutive norms of attitude A and attitude B are the same if and only if A just is B,[66]

what it is for an attitude to be a desire is just the same as what it is for an attitude to be an evaluative belief.[67]

In summary, it is possible to give a plausible normative account of the nature of belief and of the nature of desire in terms of their constitutive norms. I have argued that we can give satisfactory formulations of these norms, which both avoid the standard objections to proposed truth norms, and which have important explanatory power, in particular in explaining why belief and desire are rationally evaluable. But once we have our formulations of the constitutive norms of belief and of desire in place, it is clear that the norm of desire is exactly the same norm as that of evaluative belief. Since the norms are constitutive, desire just is evaluative belief. If this is correct, it opens a

---

[66] One who thinks that to understand belief in terms of a truth norm is to give a fundamental account of what belief *is* ought to find the general idea that rational attitudes are to be understood in terms of fundamentally constitutive norms compelling. If one does find that thought compelling, as I do, it is hard to see what could be involved in the individuation of attitudes beyond their fundamentally constitutive norms. Cf. Wedgwood: 'Suppose that there were two distinct types of mental state that did not differ in any way with respect to the conditions under which they satisfy normative concepts [….] in that case, it is very hard to see how these really could be distinct types of mental state at all' ('The Aim of Belief', p. 270).

[67] This implies a version of the Guise of the Good thesis, but phenomena such as *akrasia* and *accidie* are purported to provide counterexamples to this thesis. I think that they rely on tendentious interpretations of the phenomena and an unjustified rejection of the possibility of having contradictory beliefs. See further Michael Stocker, 'Desiring the Bad: An Essay in Moral Psychology'; Velleman, *The Possibility of Practical Reason*; Smith *The Moral Problem*; and Sergio Tenenbaum, *Appearances of the Good: An Essay on the Nature of Practical Reason*, and 'Guise of the Good'. I defer discussion of phenomena such as *akrasia* until Ch. 5.

path to an anti-Humean, cognitivist theory of practical reasons, which can nevertheless be underwritten by a (cognitivist) belief-desire theory of motivation.

3.9 Conclusion

To conclude this chapter and Part I, let me briefly draw attention to the contrast with Humean non-cognitivism. We saw in §1.2 that, according to Hume, what appear to be moral or evaluative beliefs should not be understood as genuine beliefs at all, because they have motivating power. Because of this, he claims, evaluative beliefs must be understood as, in effect, disguised desires or passions. According to Hume, then, evaluative beliefs just are desires or passions, but describing them as passions properly reveals their true nature.[68]

The crucial point is that a prior non-cognitive account of the nature of desires – as 'original existences' – takes priority in Hume, and it is this which is supposed to justify the claim that the description of evaluative beliefs as passions properly reveals their nature. The Humeans' commitment to (FAD) requires that the identification of an evaluative belief with a passion result in an understanding of the evaluative belief as not open to rational evaluation. But (FAD) is incompatible with (DRE), and so with (ND). Because the argument for identifying desire with evaluative belief that I have given is premised on (DRE) and (ND), therefore, it cannot be interpreted as implying a non-cognitive analysis of evaluative belief.[69] The identification of desire with evaluative belief, in other words, is non-cognitivist only if it takes (FAD) as fundamental, but because my argument proceeds from (DRE), it takes exactly the opposite stance. We have, so to speak, run Hume's argument in reverse.[70]

---

[68] Hume, *Treatise*, II.iii.3, III.i.1

[69] The fact that (ND) is the same norm as (NEB), together with the fact that the point of the truth norm of belief, *including evaluative belief*, is to provide an external normative constraint on cognition, is another indication that my argument is essentially cognitivist. See §3.3 above.

[70] David Lewis argues that standard decision theory rules out the view of desire as belief, and this might be thought to stand in the way of a cognitivist identification of desire and evaluative belief ('Desire as Belief', and 'Desire as Belief II'). However, Lewis' arguments have been persuasively rebutted by Huw Price ('Defending Desire-as-Belief'**),** and H. Orri Stefánsson ('Desires, Beliefs, and Conditional Desirability'). Consideration of the relation between the view developed here and decision theory, however, is beyond the scope of this thesis.

Part II

Reasons, Actions, and Explanation

Chapter 4

Values, Reasons, and Actions

But no act exists except in the doing of it, and in the doing of it there is a motive; and you cannot separate the doing of it from the motive without substituting for action in the moral sense action in the physical, mere movements of bodies.

~ H. W. B. Joseph[1]

4.1 Introduction

In Part I, we focused on issues in philosophical and moral psychology which underpin theories of practical reasons and of the rational explanation of action. In Part II, we will be concerned to construct a realist account of reasons and an accompanying account of the explanation of action, both of which will be underpinned by the philosophical and moral psychology developed in Chapter 3. This will require an account of what reasons are, of what actions are, and of the relationship between actions and the reasons for which they may on occasion be performed. The overall account will also need to harmonise the roles of reasons and of psychology in the explanation of action.

In this chapter, we will be mainly concerned, first, to develop an account of the nature of practical reasons, and second, to give the framework in which to understand the nature of actions. In §4.2 we will take stock of where Part I leaves us, and highlight some issues about the acceptability of a realist view of value which are germane to an account of reasons. §§4.3-4.4 take up these issues, arguing that non-cognitivsm and expressivism, as well as error theories of value, do not provide arguments which add anything substantive to what has been considered in Part I, and are equally undermined by the criticism of Humean philosophical and moral psychology we have given. This leaves us free to accept the view that there are objective values, where this is understood according to a realist construal.

§4.5 suggests, at quite a high level of abstraction, the structure for how an account of reasons should be brought into line with the cognitivist account of desire from Chapter 3, and §§4.6-4.7 offer a proposal for how to understand the nature of

---

[1] H. W. B. Joseph, *Some Problems in Ethics*, p. 38

practical reasons within that structure, suggesting that reasons are, fundamentally, what I shall call 'thick value facts'. §4.8 argues against the standard approach to agency, which incorporates causal theories of action and of the rational explanation of action. §4.9 suggests that we should adopt the alternative conception of actions, and of how they are related to events in general, proposed by Anton Ford. I conclude in §4.10.

4.2 Taking Stock

In Part I we focused on issues in philosophical and moral psychology, in particular on the nature of desire. I argued that what makes the Humean theory of practical reasons distinctive is its two fundamental features: that it takes reasons to be desire-based (RDB), and that it takes desire not to be fundamentally open to rational evaluation (FAD). Since (RDB) is justified by the view that desires, and only desires, are motivationally efficacious (EED), together with (FAD), but (EED) is not itself distinctively Humean, (FAD) is the fundamental feature of the Humean theory.

In Chapter 2, though, I argued that, having abandoned Hume's theory of thought, (RDB), and more importantly (FAD), lack a sound justification. The Humean theory is thus poorly motivated, and these two dogmas of Humeanism should be rejected. In Chapter 3, I argued that Kant shows how we can accommodate there being genuinely moral action, by countenancing the idea that there can be rationally evaluable 'incentives', or motivating psychological attitudes. However, I argued that Kant is wrong to restrict this characteristic to a special moral incentive, respect for the moral law, and also wrong to think that any fundamental rational evaluation of an incentive would have to be in terms of the structure of reason. Instead, I suggested, we should generalise the idea, accepting that desire in general is rationally evaluable, while rejecting Kant's rationalism.

Furthermore, I argued that a psychological attitude's being rationally evaluable is best explained by its being subject to a constitutive norm. For beliefs, this norm is a truth norm, which I argued should be formulated as (NB). For desires, though, the constitutive norm is most plausibly thought to be a norm of goodness, and I argued that it should be formulated as (ND). (ND) and (NB), however, make desires a certain kind of evaluative belief. Desires, then, just are evaluative beliefs. This yields a version of a belief-desire theory of motivation, which is nevertheless a cognitivist theory of motivation.

Recall now the discussion of Williams' distinction between internal and external reasons in §1.5. There I suggested that if the external reasons theorist were to think of reasons as value-based, then a natural response to Williams' argument against external reasons would be to claim that when someone gained a belief that there was an external reason to φ, they would, roughly, have gained a belief that it would be good if they φ-ed. I further suggested that their belief that it would be good if they φ-ed would be capable of motivating them to φ. In light of our conclusion that desires just are evaluative beliefs, this suggestion is vindicated; in gaining the belief that it would be good if they φ-ed, they gain the desire to φ. But *this* desire, it is important to remember, is *not* an Humean desire, since it *is* rationally evaluable.

There is a gap, however. As Williams emphasises, the external reasons theorist needs to be able to make sense of the idea 'that the agent should acquire the motivation *because* he comes to believe the reason statement, and that he should do the latter, moreover, because, in some way, he is considering the matter aright.'[2] As his final clause makes clear, it is not enough to make sense of the idea that a belief that one has an external reason to φ provides a new motivation to φ, it must make sense of this when the belief in question is *true*. This means that, as well as believing that there is an external reason to φ, and being motivated by this belief, there must really be such an external reason, and on the present suggestion that means, roughly, that it would *really be good* if they were to φ.

In Part I, however, while I focused on philosophical and moral *psychology*, I did not directly consider arguments against the possibility of the world's containing objective values, realistically construed; nor did I argue directly that it does contain such values. Nevertheless, as I will explain in the next two sections, I think that the best arguments against the possibility of realistically construed values depend on views about philosophical and moral psychology, and these arguments are undermined by the considerations adduced in Part I. These anti-realist arguments, therefore, do not add anything substantial to the debate that has not already been considered in dealing with the Humean philosophical and moral psychology itself, or which cannot be dealt with by appealing to the conclusions concerning them which have already been reached.

---

[2] Williams, 'Internal and external reasons', pp. 108-9.

4.3 Non-Cognitivism and Expressivism

In §1.2 we considered Hume's argument for a non-cognitivist moral psychology. According to that argument (HM), what look like moral judgements should not be taken at face value, because genuine judgements are exercises of reason, but reason could not provide the grounds of actions. Since moral 'judgements' are the grounds of some actions, they should not be understood as exercises of reason, and *a fortiori* not as genuine judgements, at all. Strictly, this is an argument for a conclusion about the nature of a certain kind of *psychological* attitude – moral judgement – but it is part of a broader argument for an anti-realist view of moral value. Indeed, shortly after giving that argument, Hume goes on:

> Shou'd it be pretended, that tho' a mistake of *fact* be not criminal, yet a mistake of *right* often is; and that this may be the source of immorality: I would answer, that 'tis impossible such a mistake can ever be the original source of immorality, since it supposes a **real right and wrong**; that is, a **real distinction in morals, independent of these judgments**.[3]

While Hume claims that what look like moral judgements are really disguised passions, he also gives an account of how we come by these apparent moral judgements, by giving an account of how we come to have passions, or impressions of reflexion, generally. Hume's general empiricist methodology is to investigate a subject matter by enquiring into the impressions which are its source, and this methodology is in evidence in his enquiry into morals. By purportedly showing that the impressions at the source of morality are, fundamentally, impressions of pleasure and pain, which, via our ideas of them, give rise to passions that are directed at their usual causes, he claims to have shown that there are no morals independent of our moral 'judgements'. In short, there are no, nor could there be any, impressions of moral qualities.[4] What we have, then, is an argument of the following kind:

---

[3] Hume, *Treatise* III.i.1, p. 460; boldface added.

[4] Hume gives (at least) three further arguments against there being moral qualities independent of our moral judgments: (i) that moral qualities cannot be relations of ideas, (ii) that they cannot be 'matters of fact', and (iii) an argument that anticipates J. L. Mackie's argument from queerness (*Ethics: Inventing Right and Wrong*, ch. 1). The first two depend on the peculiarities of Hume's theory of thought, and I will not consider them further. I will consider Mackie's argument in the following section.

(HA1) The nature of a subject is revealed by the nature of the impressions from which our thoughts about it are derived.

(HA2) The impressions from which our thoughts about morality are derived are impressions of pleasure, pain, and passions, not impressions of moral qualities.

So,

(HA3) Morality concerns pleasure, pain, and passions, not moral qualities.

In a passing comment, Hume anticipates a view which comes to prominence in a 20th century extension of Hume's non-cognitivism in moral psychology to the semantics of moral language: 'when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it.'[5] Here we have a proto-expressivist account of moral language, according to which moral assertions do not purport to represent or describe objective moral facts, or express beliefs which do so, but rather express non-cognitive attitudes.[6]

What we have are two arguments which depend on Hume's particular philosophical psychology. The anti-realist argument depends on the analysis of 'moral judgments', and the account of the origin of passions. The expressivist argument, that moral assertions do not purport to describe or represent objective moral facts, or express beliefs that do so, again depends on the analysis of what seemed to be moral judgements or beliefs, since if they are just disguised passions, there are no beliefs which purport to represent moral facts for moral assertions to express.

Contemporary expressivist views are not significantly different in this respect. For example, in Allan Gibbard's version (which generalises Hume's non-cognitivism and expressivism to encompass not just moral judgements, but all normative

---

[5] Hume, *Treatise*, III.i.1, p. 469

[6] On a literal interpretation, Hume's comment is closer to what A. J. Ayer, *Language, Truth and Logic*, p. 104, calls a 'subjectivist' theory, according to which moral assertions assert that one has some non-cognitive attitude. Since subjectivism, in this sense, is wildly implausible, I think it is more charitable to interpret Hume as a proto-expressivist, and other comments he makes support this: e.g. 'these variations [of blame and praise, according to our situation of nearness or remoteness] we regard not in our general decisions, but still apply the terms *expressive of our liking or dislike*, in the same manner, as if we remain'd in one point of view' (Hume, *Treatise* III.iii.1, p. 582, emphasis added).
For early expressivist accounts of moral language see e.g., Ayer, *Language, Truth and Logic*, esp. pp. 107-8; R. M. Hare, *The Language of Morals*, pp. 20, 171-2; and more recently, Gibbard, *Wise Choices, Apt Feelings* and *Thinking How to Live*; and Simon Blackburn, *Ruling Passions*, esp. Ch. 3.

judgements)[7] what looks like a normative judgement is really a non-cognitive attitude: it is an acceptance of a norm.[8] Acceptance of a norm, according to Gibbard, is an exercise of a special 'linguistically infused' motivational system, which he calls the 'normative control system'.[9] But, crucially, this motivational system is not a cognitive system, and an acceptance of a norm is not an attitude which represents an objective normative reality. As Gibbard puts it, 'The analysis is non-cognitivistic in the narrow sense that, according to it, to call a thing rational is not to state a matter of fact, either truly or falsely.'[10]

One aspect of Gibbard's argument that deserves comment here is that his analysis of normative language is expressivistic and 'non-cognitivistic' *because* it takes it that the states of mind that normative assertions express are non-cognitive. As in Hume, the analysis of the language depends on the underlying philosophical psychology. So we need to ask why one would think that the states of mind expressed by normative assertions are non-cognitive in the first place.

A superficial answer is that Gibbard sets out to provide a naturalist account of normative life, and so rules out in advance an account which includes real normative features. Yet he wants to avoid treating normative language as 'defective or second-rate', so his account mustn't imply that it is systematically false (as it would if normative assertions expressed cognitive attitudes).[11] Perhaps a deeper answer is that Gibbard, like Hume, admits the common-sense view that 'acceptance of a norm is motivating, at least to a degree: believing I ought to stop [eating nuts] tends to make me stop.'[12] This, it seems, is thought to be related to his fundamental criticism of realist or 'descriptivist' accounts of normative judgement, namely that 'They yield meanings that are inadequate to the basic purposes to which the term 'rational' can be put [....] descriptivistic analyses miss the general element of endorsement – an element an expressivistic analysis can capture.'[13] The thought seems to be that normative judgements could be motivating only if they were endorsements, and endorsement is a

---

[7] This is a trend among philosophers generally, not just Humeans; see e.g. Skorupski, *The Domain of Reasons*; and Scanlon, *Being Realistic About Reasons*.
[8] Or, in his more recent work, an adoption of a plan. For the earlier view, see *Wise Choices, Apt Feelings*, esp. pp. 71-5; for the later view see *Thinking How to Live*, ch. 3.
[9] Gibbard, *Wise Choices, Apt Feelings*, p. 56
[10] Ibid. p. 8
[11] Ibid. pp. 7-8, 23. For an example of the contrary view, see Mackie, *Ethics: Inventing Right and Wrong*, which I discuss in §4.4 below.
[12] Gibbard, *Wise Choices, Apt Feelings*, p. 56
[13] Ibid. p. 10

kind of non-cognitive attitude which could be understood in terms of Gibbard's notion of accepting a norm.

Without that Humean claim that motivation requires a non-cognitive attitude, however, the claim that normative judgments were exercises of a special normative control system would have no tendency to support the view that those exercises were non-cognitive. Exercises of the normative control system could just as well be cognitive. Given that the sense of 'desire' that we have been concerned with in Part I is the attenuated sense in which anything that could be motivating counts as a desire, Gibbard's view is, in the end, simply a sophisticated version of the core Humean outlook, which endorses (FAD).[14] But in Chapter 2 we saw that the Humean is not entitled either to (FAD), or to an interpretation of (EED) which vindicates (RDB), and cannot defend the view that non-cognitive attitudes are fundamental to motivation in a way that cognitive attitudes are not. Gibbard does nothing further to entitle himself to that view.

4.4 Mackie's Argument from Queerness

Another influential kind of argument against the world's containing real values is given by J. L. Mackie. In contrast to non-cognitivism and expressivism, this argument accepts the *prima facie* view that normative judgements really are cognitive attitudes, but claims that they are systematically false, since there are no such values as they purport to concern. It therefore constitutes an *error theory* of normative judgement.

Mackie's argument, in essence, is this:

(MQ1) If values were objective, they would be entities, properties, or relations
     of a queer sort
(MQ2) There is nothing of such a queer sort.
So,
(MQ3) Values are not objective.

In order to evaluate the argument, we need to understand what the salient queerness mentioned in (MQ1) is supposed to be, and what justifies its rejection in

---

[14] Gibbard admits that accepting a norm could be rationally evaluated, but only in terms of a higher-order norm that one accepts; see ibid. pp. 165-6. For further criticism of this idea, see Scanlon, *Being Realistic About Reasons*, pp. 58-61.

(MQ2). According to Mackie, for there to be objective (moral) values would be for there to be something 'in the fabric of the world' that would '[back] up and validate some of the subjective concern which people have for things'.[15] Objective values, in other words, would be values which existed in the world, independently of our thought or representation of them, and independently of our motivational attitudes or 'concerns'. In the terms that I have been using, they would be, precisely, independent objective features of reality in terms of which, at a fundamental level, our subjective desires would be rationally evaluable.

The queerness that would be involved in there being objective values, according to Mackie, is that they would have to be 'objectively prescriptive':

> An objective good would be sought by anyone who was acquainted with it, not because of any *contingent fact that this person, or every person, is so constituted that he desires this end*, but just because the end has to-be-pursuedness somehow built into it.[16]

It is clear that the queerness of objective values is supposed to be that recognising them would be motivating for one who did not already have a desire for, or to do, the relevant thing (or one suitably related to such a desire), i.e. for one who is not already 'so constituted that he desires this end'.[17] It is equally clear that Mackie here thinks of a motivational state of desire in exactly the way we saw was characteristic and fundamental to the Humean view – he accepts, that is, the fundamental arbitrariness of desire (FAD) (see the italicised portion of the foregoing quote).

Why should values being objectively prescriptive or, as he elsewhere puts it, being 'intrinsically action-guiding',[18] seem queer? The answer seems to be that a broadly Humean theory of motivation is presupposed. In particular, it is the presupposition of the view that no motivating attitude could be rationally evaluable which would make the existence of objective values seem queer. With the

---

[15] Mackie, *Ethics: Inventing Right and Wrong*, p. 22
[16] Ibid. p. 40, emphasis added.
[17] As has often been noted, the claim that an objective good would actually be sought by anyone who was acquainted with it is obviously too strong, and fails to account for the possibility of *akrasia, accidie* etc. All Mackie requires for 'to-be-pursuedness' to seem queer, though, is that objective values be motivating, even if this does not translate into pursuit.
[18] J. L. Mackie, *The Miracle of Theism: Arguments for and against the existence of God*, p. 115. The full quote runs: 'Objective wrongness, if there is such a thing, is *intrinsically* prescriptive or action-guiding, it in itself gives or constitutes a reason for not doing the wrong action, and this holds also for some, if not all, other moral features. To say that they are intrinsically action-guiding is to say that the reasons that they give for doing or for not doing something are independent of that agent's desires or purposes. But the natural features on which the moral ones supervene cannot be intrinsically action-guiding or reason-giving in this way.'

presupposition in place, no ordinary motivating attitude could be cognitive, and so could not be an attitude which someone could acquire through *recognition* of something, or 'acquaintance' with it. So if objective values were objectively prescriptive or essentially action-guiding, so that recognising them was essentially motivating, they would indeed have to be things of a queer sort, and give rise to a queer sort of cognitive motivation. Once again, it is the presupposition of (FAD) that enforces the distinctively Humean conception of motivation and desire, and which drives the view that objective values would be queer.

Furthermore, with a broadly Humean theory of motivation in place, it is clear why someone would deny the existence of such queer-seeming entities. Since the sense of 'desire' in (FAD) is the attenuated philosophical sense in which *any* motivating attitude counts as a desire, even the motivation that is produced through the recognition of an objective value would be a desire. But, *ex hypothesi*, such a desire would be cognitive and rationally evaluable, and therefore the possibility of such a desire is incompatible with (FAD). The existence of objective values would therefore appear impossible, and to be something that must be rejected, as in (MQ2).[19]

Just as we saw in the case of non-cognitivism and expressivism, this argument from queerness against the reality of values depends on an Humean conception of desire, in particular on (FAD). Since we have rejected it, we are free to acknowledge the existence of real values, without being concerned that these would entail any metaphysical queerness. Of course, it may well be true that this is incompatible with a naturalistic metaphysics, or a 'scientific conception of reality',[20] but if that were an argument against the reality of value, it would at best be a question-begging one.

The upshot is that, by undermining the Humean conception of desire, and developing an alternative normative conception of desire as well as of belief in its place, the argument of Part I also undermines the best arguments against there being real values. So we may proceed without undue concern about potential objections from these sources, since they have been undermined by the criticism of their underlying philosophical and moral psychology.

---

[19] Cf. Hume, *Treatise* III.i.1, p. 465
[20] John McDowell, 'Virtue and Reason', p. 72

4.5 Reasons and Desires

Recall Williams' explanatory constraint that we encountered in Chapter 1:

(EC) R is a reason for A to φ only if (i) A could be motivated to φ for R, and
(ii) were A to φ for R, R would figure in a true explanation of *A's* φ-ing.

An account of practical reasons must be able to show how it can accommodate (EC), and to do this, it needs to show how reasons are practically relevant. Here, I will indicate at a fairly high level of abstraction what I think it would take for an account of reasons to accommodate (EC)(i). In §§4.5-4.6 I will offer a more concrete proposal for an account of reasons, which would be capable of accommodating (EC)(i) in this way.

The following is uncontroversial. In order to be motivated to φ, or to actually φ, for a reason R, one would need to be aware of R; one would need to believe that R was a reason to φ, and act in the light of how one thus takes things to be. So the first step in showing how reasons accommodate (EC) is to link up an account of reasons with the account of moral and philosophical psychology, in particular of desire, developed in Chapter 3.

According to that account, desires are evaluative beliefs about what would be good. Specifically, desires are psychological attitudes which are subject to the constitutive norm:

(ND)   For any A and any p: (i) A ought to (desire that p) only if it would be good if p;
and,
(ii) if it would not be good if p, A ought to (not desire that p)

In Chapter 1, I argued that practical reasons should be conceived of as facts, and this conception of reasons is neutral between Humean, Kantian, and realist views. The disagreement between these views can be construed as a disagreement about whether or not the facts that are reasons must stand in some relation, for instance a deliberative relation, to e.g. one's desires or subjective motivational set, the structure of reason, or whatever. I have already argued at length against the conceptions of motivating attitudes which underpin such Humean and Kantian views, but it is clear that (EC) requires there

to be *some* relation between motivating attitudes and reasons. So far, though, I have not said anything about how we are to conceive the sorts of facts that are reasons, or what the relation between reasons and desires is supposed to be.

What I suggest is this. The intrinsic practical relevance of practical reasons, and the possibility of their being *practically effective*, can be understood in terms of reasons being facts about the sorts of features that would determine what would be good. In determining what would be good, reasons determine what one ought to desire, and desires are practically motivating attitudes.

Furthermore, we need to realise that a fact such as that it would be good if *p* would be a reason that, or be determined by reasons that, someone might have for desiring that *p*, as well as for acting so as to bring it about that *p*.[21] If one desired that *p* when it would not be good if *p*, then one's desire would contravene its constitutive norm and would be rationally defective; if one desired that *p* when it would be good if *p*, though, it seems one's desire would at least be rationally permissible, even if not justified or required. Moreover, as I argued in §3.8, a desire that *p* just is a belief that it would be good if *p*. Usually, when one believes that it would be good if *p*, one will take it that it would be good if *p* in virtue of having some further good-determining features.[22] So, typically, when one desires that *p*, and *ipso facto* believes that it would be good if *p*, one has an evaluative belief concerning a reason, and one will desire that *p* because one has that belief concerning a reason. Since being motivated for a reason *R* would primarily be a matter of desiring that *p*, reasons would meet (EC)(i): having a belief that there was a reason *R* would be something that enabled one to be motivated to act for that reason.

This is not yet enough to meet (EC)(ii), however. To do so, we need to provide an account of the relation between what an agent believes and desires and their actions, and of the relation between the reasons that an agent has and the actions they perform for those reasons. Only then can we say what is involved in the rational explanation of action, and explain how (EC)(ii) can be met. These are topics which I will take up in §§4.8-4.9 and in Chapter 5. In §§4.6-4.7 I will propose an account of reasons which can accommodate (EC)(i) in the way I have just suggested.

---

[21] Since there being reasons to so act would equally be reasons to desire that *p*, an account along the lines I am suggesting would accommodate Nagel's insistence that 'if the desire is a motivated one, the explanation of it will be the same as the explanation of [one's] pursuit' (*The Possibility of Altruism,* p. 29).

[22] An exception would be if one believed that it would be good if *p* on the basis of testimony, where the testimony did not include what would be good about it.

4.6 Thick and Thin Evaluative Concepts

The proposal, then, is that reasons are to be understood as facts about features which determine what would be good (or bad). There are two natural candidates for a more concrete account of the facts which constitute practical reasons. The first is an account which takes a realist view of evaluative facts, but takes them to be reducible to naturalistic, non-evaluative facts. Such an account would, presumably, take facts about the natural, non-evaluative features of things, which determine the evaluative facts, to be reasons. The second is an account which takes facts about what would be good to be determined by such facts as, for instance, those about what would be courageous, honest, kind, provocative, or rude. We may refer to these as *thick evaluative facts*, because, on realist views of values, the evaluative features they concern are what are taken to be picked out by *thick* evaluative concepts (in contrast to *thin* evaluative concepts such as those of the good or the right).[23] Either of these two approaches seems to be compatible with the position I have developed thus far. Since I reject the possibility of a naturalistic reduction of the evaluative to the non-evaluative, however, I will assume that the second approach to reasons sets us on the right path, and give a programmatic sketch of what reasons are, taking the things which are reasons to be facts about thick values.

How, then, should the distinction between thick and thin evaluative concepts, and the evaluative features of things they pick out, be understood? There are various well-known ways of characterising the distinction, beyond mere appeal to examples, but most of these, I will argue, are not satisfactory. This is not altogether surprising. They are formulated in a context in which the philosophical mainstream is hostile to evaluative realism, and, given the fact that one primary source of interest in the distinction is that thick concepts are supposed to be more recalcitrant to anti-realist (or, at least, non-cognitivist) treatments of value,[24] there is a tension between the *prima facie* character of these concepts and the philosophical orthodoxy from which the

---

[23] Williams, *Ethics and the Limits of Philosophy*, gives the first explicit distinction between thin ethical concepts and 'substantive or thick ethical concepts' (p. 140), although the general kind of distinction was made prior to this; see Simon Kirchin, 'Introduction: Thick and Thin Concepts', pp. 7-8; Jonathan Dancy, 'In Defense of Thick Concepts', p. 279, fn. 1. The contrast between thin and thick is usually made between specifically *ethical* concepts, but this is an inessential restriction, for it can equally be made for evaluative concepts in other domains, such as in aesthetics, epistemology, and even metaphysics (consider the concept of *parsimony*).

[24] See Williams, *Ethics and the Limits of Philosophy*, pp. 141-2

characterisations are given. Those characterisations, unsurprisingly, tend to be unsatisfactory in that they are prejudicial to a realist view.

As Williams draws the distinction, thin concepts are purely 'action-guiding', whereas thick concepts are 'at the same time world-guided and action-guiding', where being 'world-guided' means that 'the use of the concept is […] controlled by the facts or by the users' perception of the world.'[25] This characterisation is unsatisfactory in several respects, primarily because it rules out in advance the possibility that thin concepts such as *good* might also be world guided.[26] In addition, it over-emphasises the connection between thick concepts and actions, at the expense of other interesting and important connections they have in other domains such as epistemology and aesthetics, while remaining silent about their evaluative character.[27]

Another way of drawing the distinction is to say that ordinary non-evaluative concepts have descriptive content, thin evaluative concepts have evaluative content, and thick concepts contain elements of both descriptive and evaluative content.[28] Again, this is unsatisfactory; it rules out the possibility that evaluation could itself be descriptive, thus presupposing a prejudicial conception of what there is to be described.[29] A variant of this construal might be to substitute 'non-evaluative' for 'descriptive'.[30] This would be an improvement, since it would allow the evaluative elements of the content of thick concepts to be descriptive, but it also insists that it is distinctive of these concepts that they combine evaluative and non-evaluative elements in their content.

This, however, presupposes that thick concepts really have dual-component contents, which is not uncontentious. Among those who accept that the content of thick concepts is evaluative, it is often assumed that they are evaluative because their content contains an element of thin evaluation. Views then diverge on the question of whether

---

[25] Ibid. p. 141

[26] See Jonathan Dancy, 'Practical Concepts', p. 56. On p. 55 Dancy also raises questions about whether the notion of the 'world' figuring in these characterisations is the same in each case, and whether the 'guidance' relation is the same in each. Cf. Edward Harcourt and Alan Thomas, 'Thick Concepts, Analysis, and Reductionism', p. 22.

[27] See fn. 23 above. For a discussion of this, see Simon Kirchin, 'Thick Concepts and Thick Descriptions', esp. §4.1

[28] Debates about how to understand thick evaluative concepts often focus on whether they should be given a cognitivist or non-cognitivist analysis. The suggestion in the main text is a cognitivist position; the parallel non-cognitivist position would take the thin evaluative concepts to express non-cognitive pro- or con-attitudes and thick evaluative concepts to combine non-evaluative content with an expression of a non-cognitive evaluative attitude. Since I have rejected non-cognitivism, I discount that analysis, focusing only on positions which take the evaluativeness involved in thick evaluative concepts to be a matter of their content.

[29] Cf. John McDowell, 'Non-Cognitivism and Rule-Following', p. 198

[30] This terminological shift is considered and rejected by Debbie Roberts, who also rejects the alternative shift to 'natural' ('Shapelessness and the Thick', pp. 492-3).

the evaluative and non-evaluative elements are 'disentangleable', as some claim, or whether, as 'anti-disentanglers' claim, they are inextricably fused together, so that they cannot be pulled apart to reveal what the non-evaluative component is independently of the evaluative. But, as Debbie Roberts argues,[31] it is not obvious that we should accept that the evaluation involved in the use of thick evaluative concepts is thin evaluation at all, and if it were not, then (i) this would undermine taking the content of thick concepts to have two components at all, and (ii) it would support the anti-disentangling view, since there would not be two components to disentangle in the first place.[32] If, however, there really were two components of their content, it is not clear what would prevent the possibility, in principle, of their being disentangled. If they were disentangleable, though, this would invite, if not entail, the non-cognitivist claim that the evaluative component was not really part of the content, but expressed some non-cognitive pro- or con-attitude,[33] while only the non-evaluative component described the world. This would bring us full circle to the problem of posing a contrast between the evaluative and the descriptive, while inviting a non-cognitivism which I have already argued should be rejected.

A third way of drawing the distinction is to say that thick concepts are more specific than thin concepts, or that they have narrower truth- or satisfaction-conditions. Such a characterisation is encouraged by the fact that thick concepts are more informative than thin concepts. Saying, for instance, that someone's action was good carries much less information about what sort of action it was than saying that their action was generous. This is not necessarily problematic, so long as we do not equate it with a similar view of which it is suggestive, namely the view that thick concepts are *determinates* of *determinable* thin concepts, e.g. *generous* would be taken as a determinate of *good*.[34]

We should not accept the latter view, because taking thick and thin concepts to be related in this way would be incompatible with the fact that at least many thick

---

[31] Debbie Roberts, 'It's Evaluation, Only Thicker'

[32] For views of thick concepts which reject their supposed dual-content, see esp. Roberts, 'It's Evaluation, Only Thicker', and 'Shapelessness and the Thick'; but also Dancy, 'In Defense of Thick Concepts', p. 268; and Kirchin, 'Thick Concepts and Thick Descriptions', p. 75. Talk of 'disentangling' is introduced by McDowell, 'Non-Cognitivism and Rule-Following', pp. 200-1.

[33] See Harcourt and Thomas, 'Thick Concepts, Analysis, and Reductionism', pp. 21-2.

[34] Harcourt and Thomas endorse the specific/non-specific distinction, and go on to argue for the determinate/determinable distinction partly on its basis. They also reject a reductionist account of the content of thick concepts (ibid.). Cf. Christine Tappolet 'Through Thick and Thin: *good* and its determinates', who defends a similar, but reductionist, view.

concepts seem to have what we might call variable evaluative valence: in some contexts they can be good, while in others they may be bad, or even neutral or mixed.[35] For example, take the thick evaluative concept *provocative*; in some cases it may be a good thing to be provocative, while in others it would be bad. We should, I suggest, be content with the admittedly somewhat vague characterisation of the thick/thin distinction in terms of a difference in the informativeness of evaluative concepts, and resist construing this in terms of the determinate/determinable distinction.[36]

This way of understanding the distinction rejects the view that the evaluation involved in thick concepts is necessarily thin evaluation and simultaneously the standard, typically non-cognitivist, view that the content of thick concepts has a dual-character (e.g. evaluative and descriptive). It also avoids commitment to thinking of thick concepts as determinates of thin concepts, while being compatible with the variability of evaluative-valence. As Roberts argues, to apply a thick evaluative concept is to 'directly [ascribe] an evaluative property'[37] to something, but not necessarily to ascribe a thin evaluative property to it. If the thick evaluative concept *courage* is correctly applied to something, this will be because that thing has the evaluative property of courage.

The variability of the evaluative-valence of thick evaluative concepts should make us reject the view that they are determinates of determinable thin evaluative concepts such as the concept *good*. But, as I will now argue, thin values such as goodness are dependent on thick values; there cannot be thin values without thick values. In fact, I think that it is barely intelligible that there should be thin values independently of thick values. This is why reasons should be understood as thick evaluative facts, rather than what might seem the simpler view that they are thin evaluative facts about what would be good.[38]

A simple way to argue that there could not be thin values without thick values would be in terms of a conception of, say, the thin value of goodness as a summative value.[39] Then, one could argue, a fact about what would be good in a situation would be a function of an enumeration of the thick values in that situation, with their evaluative-

[35] Dancy, 'In Defense of Thick Concepts'

[36] This assumes that thick concepts are indeed evaluative, an assumption which has been challenged, for instance by Pekka Väyrynen, 'Thick Concepts: Where's Evaluation?'.

[37] Roberts, 'It's Evaluation, Only Thicker', p. 87

[38] The apparently simpler view would make for a more direct linking up with the account of philosophical psychology than that suggested in §4.5.

[39] I shall henceforth speak simply of goodness, rather than thin values in general.

valence fixed, adding up to a net favouring of something. However, I do not think that a summative conception of goodness is plausible.[40]

A second, more interesting and plausible way to argue is to claim that something could not be good without there being something about it which, or in virtue of which, it was good. Since thin evaluative concepts are thin precisely because they lack the informativeness that is characteristic of thick evaluative concepts, it is the applicability of thick evaluative concepts which show what it is about something that is of value when it is good.[41]

We can develop this basic line of thought and put it in terms of the widely accepted thesis that the evaluative supervenes on the non-evaluative. The supervenience of the evaluative on the non-evaluative means that it is impossible to make sense of a difference in the evaluative without there also being a difference in the non-evaluative, but not conversely – there could be a difference in the non-evaluative without this entailing a difference in the evaluative. That means that, in a given situation, one would need a way to distinguish the relevant non-evaluative supervenience base, i.e. the non-evaluative facts which would make the difference to the obtaining of the evaluative facts, from those non-evaluative facts which were not relevant to determining the evaluative facts.

The characteristic informativeness of thick evaluative concepts provides just such a way of determining the relevant non-evaluative facts in a given situation, because their applying in that situation provides information about which sorts of non-evaluative facts in the situation the thick value fact supervenes on. This might be interpreted in two quite different ways, which involve two correspondingly different interpretations of thick evaluative concepts. On the first of these, the content of thick concepts should be understood as having a dual character, one side being non-evaluative and the other being thinly evaluative. Then the non-evaluative part of the content of a thick concept would provide the relevant non-evaluative way things would have to be for the evaluative concept to apply, while the other part would be the thin evaluation of goodness which applied as a result. In principle, the non-evaluative part could be 'disentangled' from the evaluative part, so that in grasping a thick concept, one would

---

[40] This is in part because different values need not be commensurable; cf. Wiggins, 'Weakness of Will, Commensurability, and the Objects of Deliberation and Desire'.

[41] Cf. Tappolet, 'Through Thick and Thin', p. 211. As Tappolet makes clear, the claim should strictly be that something could not be good *pro tanto* without there being something that was good about it.

grasp the relevant non-value and the supervening value together.[42] But we should not accept this view of the role of thick evaluation in understanding the supervenience of the evaluative on the non-evaluative, at least not if we reject a view of thick concepts as having such a dual-character, along with the view that the evaluative is reducible to the non-evaluative, as I earlier suggested that we should.

On the second interpretation of the role of thick concepts in determining the relevant non-evaluative supervenience base, which I favour, the content of thick concepts should be understood as unitary, meaning that, unlike on the first interpretation, there would be no non-evaluative component of their content which could directly provide the relevant non-evaluative way things would have to be for a thick concept to apply. Instead, the application of a thick evaluative concept to a thing would ascribe an evaluative property to it, and the unitary evaluative content of a thick concept itself, on this interpretation, would provide the characteristic informativeness about what something is like in being valuable in that way.

Thick values, though, are still evaluative, so, according to the supervenience thesis, they would also supervene on the non-evaluative. Because the evaluative is not reducible to the non-evaluative on this view, though, the way that something would have to be in order for a thick evaluative concept to apply could not be codified in purely non-evaluative terms, in such a way that the non-evaluative codification would yield a correct verdict about whether the evaluative concept applied in every possible case.[43] The uncodifiable, non-evaluative way something was in order for a thick concept to apply to it would not be part of the *content* of a thick concept, for there is no non-evaluative component of the content of thick concepts, and there is therefore no possibility, on this view, that a non-evaluative component could be disentangled from a thinly evaluative component.

It then seems plausible to claim that it would only be in virtue of understanding the application of a certain thick evaluative concept in a particular case that one could be in a position to see which non-evaluative facts were '*normatively* crucial' to the thick evaluative concept's applying in that case (in the sense that they were responsible for

---

[42] Note that this could be a two-component view of the content of thick concepts which was nevertheless cognitivist; cf. Tappolet, 'Through Thick and Thin', pp. 213-4.

[43] That the evaluative is in this way non-evaluatively 'shapeless' is a view McDowell argues for in 'Non-Cognitivism and Rule-Following'. For recent discussions of the shapelessness hypothesis, see Roberts, 'It's Evaluation, Only Thicker', esp. §5.2; and Simon Kirchin, 'The Shapelessness Hypothesis'.

the evaluative property's instantiation).[44] Thick evaluative facts would be essential for there being thin evaluative facts just because it would be thick evaluation which would provide the (uncodifiable) route from an understanding of things being a certain way evaluatively, to an understanding of the normatively relevant sort of way they were non-evaluatively, on a case-by-case basis. Thick evaluative concepts, picking out thick values, would be required to provide the route from any supervening value, including thin values, to the non-values on which they supervened.

4.7 Value-Based Reasons

In §4.6 I claimed that it is inadequate to draw the distinction between thick and thin concepts in terms of the thick being both action-guiding, or 'practical' as Dancy puts it, and at the same time world-guided, while thin concepts are purely action-guiding. But that needn't prevent us from recognising the importance of having such a capacity to thick concepts, or the properties they pick out. We should agree with Dancy when he argues that:

> Somehow we have to make sense of the idea that these concepts have an intrinsic practical relevance. Other, non-evaluative concepts may be such that their applicability makes a practical difference on occasion, but that relevance is extrinsic to them. It may make a difference that this car goes faster or uses less petrol than that one, but that difference is not intrinsic to those features.[45]

According to the view he expresses, thick concepts are (or, rather, their applying in particular cases is) intrinsically relevant to how one should act, in contrast to non-evaluative concepts which *may* be relevant, but only in virtue of their relation to something else. This is not to be confused with the claim that thick concepts are *necessarily* practically relevant; they are neither necessarily relevant in a particular way – e.g. in a pro way – since they may have variable evaluative valence, nor are they guaranteed to be practically relevant in every case in which they apply, since the possibility of variability makes it plausible that they might sometimes be neutral. It is just that when they *are* practically relevant, in one way or another, this is not because of something else, but is intrinsic to them. Non-evaluative features are, by default, practically irrelevant, and become relevant in virtue of something else, which explains their relevance; thick evaluative properties are by default practically relevant, and when

---

[44] Cf. Roberts, 'It's Evaluation, Only Thicker', pp. 92-3
[45] Dancy, 'Practical Concepts', p. 56

they are not, this is to be explained by something else. This is so even though *what* practical relevance the applicability of a thick concept has *is* dependent on what other features are present in a particular context.[46]

It is distinctive of thick properties, then, that they are practically relevant, but this does not directly connect them with actions. They are connected with actions by their intrinsic practical relevance being determined in a given context in such a way that they favour or disfavor certain sorts of actions. In other words, thick evaluative facts favour or speak against certain sorts of actions (in particular contexts), and these sorts of actions are determined as actually being good or bad by the total confluence of thick evaluative facts which characterise a given situation. This conception of the distinction between thick and thin evaluative concepts and properties suggests an attractive view of value-based practical reasons. It is ordinarily of practical relevance that φ-ing would be, for instance, provocative, and the fact that it would be provocative would count as a reason to φ because, in a certain context, this fact would favour φ-ing. Similarly, it would count as a reason against φ-ing because, in that context, being provocative would disfavor φ-ing. In general, thick evaluative facts constitute practical reasons when they would favour or disfavour performing a certain sort of action in a given context.

Non-evaluative facts may also be reasons for action, but not intrinsically, and not fundamentally, since they depend on value-based reasons. For example, the fact that the cliff is unstable might be a reason not to walk too close to the edge, but this non-evaluative fact would only be a reason because, in the context, there was a thick evaluative fact which provided a reason to not risk falling down the cliff, the fact that falling down the cliff would be tragic, say. This is the evaluative realist analogue of the Humean claim that a non-evaluative fact is a reason only in virtue of standing in the appropriate relation to one's desires.[47]

---

[46] Ibid. pp. 56-7. Compare Dancy's account of the motivational significance of beliefs in *Moral Reasons*, pp. 22-6

[47] Several philosophers take reasons to be non-evaluative facts, but do not concede that this is only so in virtue of evaluative facts; see e.g. Dancy, *Practical Reality*, p. 104; Scanlon, *Being Realistic About Reasons*, pp. 30-3; Parfit, *On What Matters*, pp. 31-2. Dancy also admits that evaluative facts are reasons, but does not acknowledge the relation between them and non-evaluative reasons that I have asserted; see Jonathan Dancy, 'Why There Is Really No Such Thing as the Theory of Motivation', p. 6. Scanlon admits that beliefs about reasons have a special rational relation to intention and action, which 'factual' beliefs have a weaker form of 'insofar as they are beliefs about things that *are* reasons' (*Being Realistic About Reasons*, p. 64).

The Humean theory of practical reasons gives special importance to instrumental reasons and instrumental reasoning or deliberation.[48] Indeed, as we saw in §1.5, Williams argues that internal reasons can be discovered in deliberation, but that it is hard to see how, through deliberation, one could come to truly believe that one has an external reason, and this is supposed to be part of his case against external reasons. The excessive focus on instrumental reasons and instrumental reasoning has been thought by some to be a problem with the Humean theory, and because of this, it has been suggested that we need to enrich an Humean account without abandoning its spirit, or that we need to abandon the Humean account altogether.[49] I have already rejected the Humean view of practical reasons on other grounds, but I am also sympathetic to the view that deliberation involves more than mere instrumental reasoning. My account is able to accommodate both instrumental and non-instrumental reasons and reasoning, but it is also a consequence of it that instrumental reasons in particular should be understood in a way that is interestingly richer than usual.

First, non-instrumental reasons and reasoning. One kind of non-instrumental reason is obviously accommodated; there may well be some actions, or kinds of action, which are non-instrumentally good – their goodness is not derived from their instrumental relation to something else which is good. For example, φ-ing might simply be generous, and the generosity of φ-ing, in the context, might itself make φ-ing non-instrumentally good. Whatever features would make actions good in themselves would constitute non-instrumental reasons to perform them. Correspondingly, a non-instrumental form of practical reasoning might be deliberating about which kinds of actions would be non-instrumentally good to perform, for example considering whether doing the generous thing would be (non-instrumentally) a good thing to do.

Another sort of non-instrumental reason to φ might be that φ-ing would constitute, or non-instrumentally contribute to, ψ-ing, when it would be good to ψ.[50] For example, one has a reason to care for one's child so that one is a good parent, and it would be good if one were a good parent. Caring for one's child, however, is not

---

[48] See, for example, Hubin, 'What's Special about Humeanism?'; Neil Sinhababu, *Humean Nature: How desire explains action, thought and feeling*, pp. 3-4.

[49] Examples of the former might be thought to include Williams, 'Internal and external reasons'; and David Schmidtz 'Choosing Ends', esp. pp. 247-55; examples of the latter might be thought to include Scanlon, *What We Owe To Each Other*, ch. 1; and Wiggins, 'Deliberation and Practical Reason'.

[50] This sort of non-instrumental reason is important for non-consequentialist interpretations of Aristotle's ethics. In particular, on this sort of interpretation one would not have an instrumental reason to be virtuous (for the sake of *eudaimonia*) – being virtuous would be partly constitutive of *eudaimonia*; see e.g. J. L. Ackrill, 'Aristotle on *Eudaimonia*'.

instrumentally related to being a good parent, but is itself partly constitutive of it, and (at least ordinarily) it is not something that one would do in order to instrumentally bring it about that one was a good parent. Again, there is a corresponding form of deliberation: one might consider what sort of actions would constitute (perhaps partly) ψ-ing.[51] Cataloguing the various sorts of non-instrumental reasons there could be is beyond the scope of this work; suffice it to say that my account has no problem with admitting that there are such reasons.

Second, instrumental reasons and reasoning. One obvious kind of instrumental reason, which I shall take as a model, would be the kind of reason that one can have to act prudentially. For example, enrolling in a pension scheme so that one will have an income in retirement might be a prudent thing to do, and the reason that one had for doing so would seem to be an instrumental reason. On an Humean model, the instrumental reason would be desire-based; one would have a reason to enrol in a pension scheme because it was instrumentally related to having an income in retirement, and this was something that one desired. The reason would simply be the non-evaluative fact that the former was instrumentally related to the latter. This cannot be how instrumental reasons are understood on my account, for it takes reasons to be thick evaluative facts. There certainly can be these sorts of Humean instrumental relations, but it will be a further question whether, when there is, that means there is an instrumental reason.

On my view, we can say the following about whether there would be an instrumental reason, when there is an Humean instrumental relation. If it would be good if one were to have an income in retirement, there would be thick values which would make it good. Enrolling in a pension scheme would be instrumentally good in virtue of being instrumentally related to having an income in retirement. The values which would make it good to have an income in retirement would make it instrumentally valuable to enrol in a pension scheme, and so provide instrumental reasons to do so.

It is plausible, however, to think that *prudence* is a thick evaluative concept,[52] which picks out the evaluative property of prudence. On this view prudence is a non-instrumental thick value that something typically has in virtue of, roughly, standing in an Humean instrumental relation to something in one's future. So, in our example, enrolling in a pension scheme would be instrumentally valuable in virtue of being

---

[51] Williams, 'Internal and external reasons', pp. 104-5
[52] See e.g. Tappolet, 'Through Thick and Thin', p. 210, who considers this example in passing.

instrumentally related to something – having an income in retirement – which is valuable, and it would be prudent in virtue of standing in that Humean instrumental relation. But the fact that enrolling in a pension scheme would be prudent would not itself be a fact of instrumental value, but of non-instrumental value. If, as would be possible, it being prudent would itself favour enrolling in a pension scheme, then the fact that it would be prudent would be a non-instrumental reason to enrol in a pension scheme in addition to the instrumental reasons.

Moreover, in another sort of example in which the end desired is not valuable, φ-ing might still be prudent, perhaps because it would nevertheless stand in an Humean instrumental relation to the end. There would then be a further question about whether prudence favoured φ-ing. Plausibly, this would depend on the context; sometimes an action's being prudent favours performing it, but other times it speaks against performing it. Usually, there will be many alternatives, some of which will be characterised by other thick values. For instance, one might have a choice between acting prudently and acting boldly, or loyally. One would need to consider whether, in this context, acting prudently really did favour φ-ing, or whether it spoke against it and acting loyally favoured ψ-ing.

Whatever the case, though, my account denies that there could be merely instrumental reasons to do something, in anything like the way that the Humean claims. Either there will be a genuine evaluative instrumental reason that favours the means, in which case there will be a non-instrumental reason that favours the end, or there will be a non-instrumental reason of prudence (or the like) that favours the means, or both. But there could be no reason to φ when there was no value in φ-ing, nor in what φ-ing would be instrumental in bringing about.

4.8 The Standard Causal Approach to Agency

Let us now return to the question of how an account of reasons could accommodate (EC)(ii). In §4.5, I claimed that in order to see how this condition could be accommodated, we would first need an account of the relation between an agent's having certain beliefs and desires and their actions. In this section I will consider the 'standard approach'[53] to agency, which takes the relation between an agent's attitudes and their actions to be a causal relation, and takes this to provide an account of the

---

[53] The phrase is G. E. M. Anscombe's, 'Practical Inference', §1.

nature of action itself. I will argue that the resulting causal theory of action should be rejected.

The standard approach to agency encompasses a causal theory of action and of the rational explanation of action. A causal theory of action takes actions to be particular events (usually bodily movements), which are describable in terms of their effects.[54] However, not all bodily movements are actions, for example sometimes one has a muscle spasm and one's leg moves, but these are not cases in which there is an action that occurs of one moving one's leg.[55] Bodily movements, on a standard causal theory of action, get to be actions in virtue of being caused (in the right way) by an agent's beliefs and desires. A causal theory of the rational explanation of action, correspondingly, takes rational explanation to be in terms of an agent's beliefs and desires, which rationalise, and in addition cause, their actions.[56]

This standard approach to agency is, as Hornsby emphasises, an 'events-based account'.[57] The conception of causality which it makes central to its account of actions and to the account of rational explanations in terms of an agent's beliefs and desires is one of event-causation. It also takes both bodily movements and psychological attitudes to be events, where the former are actions in virtue of being caused (on the specified conception of causality) by the latter.[58] The phenomenon of agency, according to the standard approach, 'is delimited when it is said which events are actions',[59] and this is done by saying which bodily movements are appropriately caused by agents' beliefs and desires.

The normative character of the philosophical psychology I defended in Chapter 3 does not in itself make it incompatible with a version of the standard approach to agency, since, as Davidson has demonstrated, one can accept that the psychological is essentially normative without denying that it is also causal.[60] The normative character of the psychological may well present an obstacle to the reduction of psychological

---

[54] Cf. Donald Davidson, 'Agency', and 'Problems in the Explanation of Action', p. 103

[55] Hornsby ('Agency and Actions', pp. 3-9) argues that sometimes one does something intentionally but there is no event of one's doing it at all, which causes trouble for what she calls an 'events-based account of agency'.

[56] Cf. Davidson, 'Actions, Reasons, and Causes'

[57] Hornsby, 'Agency and Actions', p. 4

[58] Davidson ('Actions, Reasons, and Causes', p. 12) makes this explicit, but it is more common for philosophers to speak of mental events, states, and processes interchangeably. For sustained and compelling criticism of treating these interchangeably, see Helen Steward, *The Ontology of Mind: Events, Processes, and States*.

[59] Hornsby, 'Agency and Actions', p. 4; cf. Michael Smith, 'The Standard Story of Action: An Exchange (1)'

[60] See especially Donald Davidson, 'Actions, Reasons, and Causes', and 'Mental Events'.

properties or kinds to physical properties or kinds, however, but again, Davidson shows that one can accept the irreducibility of the psychological without rejecting a naturalistic monism or a standard approach to agency.[61]

If there is an incompatibility with the standard approach, it is due to the platitudinous characterisation of what it is for someone to believe something:

> (B) What it is to believe that *p* is to have a view as to how things are, namely that the way things are is that *p*.

At least *prima facie*, it is very hard to accept that *having a view as to how things are*, namely that *p*, could be a matter of there being a particular (event) inside one's mind or brain, or even that *A's having a view as to how things are* could be a matter of there being a particular (event) in *A*'s mind or brain, since having a view as to how things are is a property of a whole person. As will become clear, the reply that something can be a property of a whole person which one has *in virtue* of the person having a particular (event) inside their mind or brain is at least not *obviously* available, since this is of a piece with claiming that a whole person is reducible to psychological particulars, which would seem to make the position circular.

Those who follow the standard approach take it that there is an event-causal relation between certain psychological attitudes an agent has (conceived as mental particulars) and their actions, but they are nevertheless happy to admit talk of an *agent* causing their actions, so long as this is taken as elliptical for talking of the agent's *attitudes* causing their actions. For example, Davidson says that 'we understand [talk of an agent causing an event which is their action] only when we can reduce it to the case of an event being a cause'.[62] Velleman, meanwhile, disagrees with the details included in a simple version of the standard approach, denying that an agent's causal role can be captured in terms of the causal role of their beliefs and desires:

> In this story, reasons [i.e. belief-desire pairs] cause an intention, and an intention causes bodily movements, but nobody – that is, no *person* – *does* anything. Psychological and physiological events take place inside a person, but the person serves merely as the arena for these events: he takes no active part.[63]

---

[61] See Davidson, 'Mental Events'
[62] Donald Davidson, 'The Logical Form of Action Sentences', p. 128
[63] J. David Velleman, 'What Happens When Someone Acts?', p. 123

Velleman thinks that what is needed is for the standard story to be supplemented with an additional ingredient of the same kind: a mental particular. In quite explicit terms, he explains his acceptance of the approach, and his objection to its details:

> One is surely entitled to hypothesize […] that there are mental states and events within an agent whose causal interactions constitute his being influenced by a reason, or his forming and conforming to an intention.
> [….] But the states and events described in a psychological reduction of a fully human action must be such that their interactions amount to the participation of the agent. My objection to the standard story is not that it mentions mental occurrences in the agent instead of the agent himself; my objection is that the occurrences it mentions in the agent are no more than occurrences in him, because their involvement in an action does not add up to the agent's being involved.[64]

To involve an agent in the story, according to Velleman, it must contain, in addition to the usual ingredients, further special psychological attitudes, which 'are functionally identical to the agent, in the sense that they play the causal role that ordinary parlance attributes to him'.[65] It turns out, in his view, that the special psychological attitude which plays the right functional role, is a special 'desire to act in accordance with reasons' which 'is functionally identical to [the agent.]'[66] By adding this attitude to the standard story, the standard approach is supposed to be capable of vindicating talk of *agents* causing their actions.

But the standard approach is untenable, whether in its simple form or in a more complex form such as Velleman suggests. This is because it gets the relationship between an agent and their actions, the relationship between an agent and what they cause, and the nature of actions wrong. In the process, it leaves agents themselves out of the frame of agency altogether. To see this, recall Hornsby's distinction between a *thing done*, which is a universal, and an *action*, which is a particular (§2.4). Exercises of agency occur when someone does something intentionally, and when one does something intentionally, they cause something to happen or, perhaps, to be the case. But what an agent causes to happen are not their actions at all, but the consequences, results, or effects of their action. For example, someone might lock the door; what they do would be to lock the door, whereas what they cause would be the door's locking, or the door's being locked. Their action would be their locking of the door, and the door's

---

[64] Ibid. p. 124-5
[65] Ibid. p. 137
[66] Ibid. pp. 141-2

being locked would be the consequence of their action.[67] If this is right, then in taking agency always to be a matter of an agent causing an action (where this is reducible to, or constituted by, attitudes inside them causing an action) the standard approach will have given a mistaken view of the relation between an agent and their actions, and of the relation between an agent and what they cause.

But this may be too quick. According to Davidson's influential account, basic actions, which he characterises as 'the ones we do not do by doing something else',[68] are always movements of the body, and actions are usually described in terms of their effects, to which they stand in event-causal relations. When it comes to basic actions, these are all, according to him, 'mere movements of the body';[69] every action one performs directly is a mere movement of one's body, although they may be described, in terms of other things to which they stand as cause, as one's doing something else. It is now a familiar point that such a formulation is ambiguous as to whether 'movement' is supposed to be interpreted as the transitive or intransitive form of the verb.[70] It is uncontroversial that if it is interpreted as the transitive form (so that the claim is that someone's moving their body is a basic action) it is at least true that it is an action, whether or not it is true that basic actions are always or ever instances of one moving one's body. If, however, 'movement' in 'movement of one's body' is interpreted as the intransitive form of the verb, (so that the claim is that every basic action is an instance of one's body moving) it is much more controversial that the phrase concerns actions at all. But Davidson claims that when, and only when, we are dealing with someone's basic movement of their body, their movement of their body is identical to their body's movement.[71] So the event an agent would have caused in, say, moving their hand, namely their hand's moving, would be identical to their moving their hand, i.e. to their action. This may make one think that it is not in general true that an agent causes the effects of their action *rather than* their action itself.

But we should not accept Davidson's view. The argument he gives for it is that it is required to avoid a regress. He starts by saying that 'if we ask what makes a particular case of an arm going up a case of an arm being raised, a natural answer is that

---

[67] Cf. Hornsby, 'Agency and Actions', esp. pp. 16, 18; Maria Alvarez and John Hyman, 'Agents and their Actions', p. 233
[68] Davidson, 'Agency', p. 59
[69] Ibid.
[70] See esp. Jennifer Hornsby, *Actions*, ch. 1.
[71] Davidson, 'Problems in the Explanation of Action', pp. 101-5; for discussion see Jennifer Hornsby, 'Actions in Their Circumstances', esp. pp. 112-3.

the agent made his arm go up [….] Perhaps this "making" is the thing that must be added to the arm going up to make it the raising of an arm.'[72] He goes on, suggesting that this line should be resisted, because 'if what marks the difference between something an agent does and what happens to that agent is a prior act of the agent – a making happen – then for the prior act to be something the agent does, another antecedent is required, etc.'[73] The thought here is that, if we try to explain what turns something that happens to an agent into an action of theirs – their arm going up into their raising their arm, as it might be – in terms of their performing some other action – their making their arm go up – we shall have to explain what makes *this* into an action of theirs (rather than something that merely happens to them) in terms of their performing some *third* action, and then explain what makes *this* into an action… and so on regressively. This is clear from Davidson's follow-up comment: 'It will not help to suggest that to do something an agent must do something else, such as try to perform the desired act, for trying is itself an act, and so would require a prior trying, etc.'[74]

The mistake here is to think that we should be trying to explain what turns something that happens to an agent into an action of theirs in the first place, never mind explaining this in terms of their performing some other action. We should reject this as misguided from the start, because an event that is an action is *not* the sort of thing that *might not have been an action*.[75] One's arm going up, then, should never have seemed a candidate to be an action in the first place, and the question, 'what makes a particular case of an arm's going up a case of an arm being raised' should be rejected rather than answered.

We are left, then, saying that the agent's raising their arm was an action of theirs, although their arm's rising was not, so why does this not start us along Davidson's regress? The answer is that the regress is only generated when at every step the event which is supposed to be the agent's action is also supposed to be an event which is not *essentially* an action, and in which the agent is not already essentially involved. That is what makes it seem necessary to ask what makes *it* an action, and how the *agent's* involvement can be secured, thus setting us on the path of the regress. If, in contrast, we say that the agent's action was *their raising* their arm, which was *their*

---

[72] Davidson, 'Problems in the Explanation of Action', p. 102
[73] Ibid.
[74] Ibid.
[75] Jennifer Hornsby, 'Agency and Actions', p. 19; 'The Standard Story of Action: An Exchange (2)', pp. 59-63; and see further §4.9.

causing their arm's rising, while rejecting any attempt to construe this as a matter of some event occurring in their mind or brain to which an agent's role might be reduced (e.g. an intention, a desire to act in accordance with reasons, or a trying of the sort Davidson considers), the regress is avoided at the first step.

We then do not need to identify one's basic actions with certain movements that one's body undergoes. Instead, we can say simply that someone's basic actions are events of their doing something, where they do not do that thing by doing something else. Many cases of someone moving their body will thus count as basic, in this sense, but so, equally, will many other sorts of action, such as Hornsby's example of an agent's carrying a suitcase, by which action they cause it to be carried (by them), but where they do not carry it by doing something else.[76] But in no case, including cases of basic action, is one's action the same event that one causes.

If, as I have just argued, an agent is ineliminable from a proper picture of her basic actions (where these need not be restricted to her bodily movements) then she will also be ineliminable from her non-basic actions – her doing those things which she does indirectly, by doing something else – since any of her non-basic actions will depend on her doing something directly. Without agents in the frame, no event that could remain in their absence should even seem to be a candidate for being an action. The conception of the nature of actions adopted by the standard approach, therefore, which sees actions as events which might not have been actions at all, and which only get to be so in virtue of being appropriately caused by an agent's attitudes, is unacceptable. As Hornsby puts the point:

> Those who speak as if an action were an event one candidate for whose cause is an agent make it seem as if an action might be identified independently of any agent. But an event that merits the title 'action' is *a person's* intentionally doing something. And

---

[76] In 'Actions in Their Circumstances', p. 107, Hornsby suggests this and a range of other examples. However, she claims that, in a sense of 'cause' in which causality is not supposed to be a relation between events, these are examples in which an agent causes the event which is their action: 'The causality here is internal to an event: Ann's carrying the suitcase *is* the event of its being carried', and goes on to make the same claim about Davidson's example: 'a person's raising her arm (when she raises it) is indeed the same as her arm's rising (then), as Davidson said. But this is not a special case, as Davidson thought' (p. 114). I agree with Hornsby that moving one's body is not a special case, and that other sorts of actions can be basic, but disagree that in any case of basic action, the event which is one's action is identical to its consequence. In Hornsby's example, it may be true that Ann causes it to be the case that the suitcase is carried (which is not a particular), but it is not true that she causes her carrying of it, or its being carried *by her* (which is a particular) – she just carries it.

such events do not belong in a causal order from which people themselves might be missing.[77]

The causal theory of action is a proposal about how it is that, among all the events there are, some of them qualify as actions. But if the forgoing is correct, the causal theory of action is untenable. We therefore need a different account of what actions are, and of how the distinction between events which are actions and those which are not is to be understood. In the following section, I will suggest that the account of the relationship between actions and events proposed by Anton Ford should be embraced, and that it avoids the problems which arise for the causal theory of action that we have just considered.

4.9 Action as a Categorial Species of Event

Actions, we saw in §4.8, must be understood as events of an agent's doing something, which occupy the position of causing what the agent herself causes in doing something. An agent's role in their doing something, furthermore, is not reducible to that of certain attitudes, conceived as particulars inside them, causing their action. Agents must be irreducibly involved in the picture of agency and actions.

This insists that actions are a fundamentally different kind of occurrence to other so-called 'natural' events. But they are events nonetheless. Making sense of this requires the relationship between the phenomena falling under the concept *action* and the phenomena falling under the more general concept *event* to be thoroughly reconceived. Anton Ford provides just such a thoroughgoing reconception, and I suggest that his account simply be adopted in place of the causal theory of action.

According to Ford, the standard approach in the philosophy of action incorporates a causal theory of action because it pursues the project of searching for what he calls an '"accidentalist" account of intentional action.'[78] On the accidentalist account, any occurrence is, in the first instance and most fundamentally, an event; in the domain of agency, the (fundamental) category of *event* is subject to a first division into *action* and *mere event*, with a second division in the category of *action* into, as Ford puts it, *unqualified action* and *qualified action* (the latter including, e.g., unintentional,

---

[77] Hornsby, 'Agency and Actions', p. 19. This does not imply that one cannot know that there was an action without knowing who the agent was, or that there are not cases in which one cannot tell whether an event was an action or not; cf. Hornsby, 'The Standard Story of Action: An Exchange (2)', pp. 60-1.
[78] Ford, 'Action and Generality', p. 79

weak-willed, half-hearted, or other sorts of action which fall short of what Velleman calls 'action *par excellence*').[79] It then supposes, Ford says, that 'the primary task of the philosopher of action is to hound down the differentiae – to say what an event must be, over and above being an event, in order to be an action, or to say what an action must be, in addition to being an action, in order to be an action in the full and proper sense.'[80] This presupposition is clearly in place in the accounts we considered in our discussion of the causal theory of action.

An accidentalist approach requires criteria for the individuation of events *just as such*, for instance like Davidson's suggestion that for an event A to be the same event as B is for A and B to have exactly the same causes and effects, or his later suggestion that it is for A and B to have the same spatiotemporal boundaries.[81] But we could not begin to answer, for instance, whether an event A occurred in the same place as event B, or whether the causes and effects of A were the same as those of B, unless we already knew what sorts of event A and B are supposed to be, since such knowledge is essential for knowing what sorts of location, or what sorts of causes and effects, could be truly predicated of them.[82] So, as Anscombe replied to Davidson, 'the demand for a criterion of identity of particular occurrences *just as such* is not a reasonable one';[83] it is only reasonable to demand a criterion of identity of occurrences of a specific sort, e.g. of weddings, makings of vows, ceremonies, or enterings into contracts.

The trouble with Anscombe's rejection of the standard approach, Ford suggests, is that while she argues that nothing could be the extra ingredient added to an event to make it into an action, or to make an action into an intentional action, she does nothing to lift the presuppositions underlying the feeling that there *must* be some, and so fails to bring others to so much as see the possibility of an alternative. The presuppositions in question, Ford argues, concern the variety of generality at play in the relation between

---

[79] Velleman, 'What Happens When Someone Acts?', p. 124

[80] Ford, 'Action and Generality, pp. 76-7

[81] See Donald Davidson, 'The Individuation of Events', and 'Reply to Quine on Events'.

[82] Cf. Jennifer Hornsby, 'Physicalism, Events, and Part-Whole Relations', and 'Which Physical Events are Mental Events?'; and Anscombe, 'Under a Description'.

[83] Anscombe, 'Under a Description', p. 217. Davidson defends his proposed criterion of individuation, despite its deficiencies, on the grounds that it is no worse off than criteria of individuation for material objects ('The Individuation of Events', p. 180). But the demand for a criterion of identity of particular material objects *just as such* is not a reasonable one either, but only of tables, horses, statues etc.; see David Wiggins, *Sameness and Substance Renewed*; and Hornsby, 'Actions in Their Circumstances', p. 127.

event and *action*, or *action* and *intentional action*, where the first category of each pair is more general than the second.[84]

Ford distinguishes three varieties of generality: accidental generality, categorial generality, and essential generality. In a case of accidental generality, an 'accidental species' of an 'accidental genus' is a member of the species in virtue of being a member of the genus, and furthermore possessing additional differentiae. The genus is therefore metaphysically prior to the species, which is a 'derivative kind of thing'.[85] Ford gives the example of a snub nose, which is an accidental species of the genus *nose*. A snub nose is fundamentally a nose, which is snub by having the additional (less fundamental) feature of being concave. In a case of categorial generality, the relation between species and genus is different, in that 'no quality can account for the contrast between a categorial species and its genus – or, at any rate, no quality that is logically independent of the species.'[86] Something is a member of a categorial genus in virtue of being an instance of one of its categorial species, which is more fundamental than the genus. For example, a horse is a species of the genus *animal* not in virtue of being fundamentally an animal, together with some further, less fundamental, features which makes it into a horse; it is an animal in virtue of being a horse.[87] Finally, an essential species is 'neither posterior to its genus, like an accidental species, nor prior to its genus, like a categorial species [….] An essential species and its genus are coeval.'[88] For example, he says, pure gold is an essential species of the genus *gold.* The criteria for something being pure gold is not independent of the genus itself, for all that something has to be in order to be pure, rather than impure gold, is *gold and gold alone*. Similarly, for a circle to be a *perfect* circle, it must simply have the property of being circular *without defect*. However, something is not a member of an essential species independently of what it is for it to be a member of the genus. It is not, as it is in the case of a categorial species, that something is a perfect circle in the first instance, and a circle only secondarily, as if it were only one way amongst several that something could be circular. A thing's membership of the genus *circle* and its membership of its essential species *perfect circle*

---

[84] Ford, 'Action and Generality', pp. 80-1
[85] Ibid. p. 83
[86] Ibid. p. 85
[87] Ibid. pp. 85-8
[88] Ibid. p. 91

are equiprimordial; imperfect circles, meanwhile, are just circles deficient in their circularity.[89]

The standard approach presupposes that the relation between each of the traditional divisions – the divisions of events into mere events and actions, and of actions into qualified and unqualified actions – is an example of accidental generality, so that the genus *event* is most primordial, and with the addition of some differentiae an event becomes an action, and with further differentiae becomes either an unqualified or qualified action. But this presupposition, Ford argues, is in fact quite implausible in both cases.

Unqualified actions, he argues, are actually *essential* species of the genus *action*. This explains, first, why philosophers of action are primarily concerned with intentional, free, wholehearted, or rational actions, but only interested in *un*intention, *un*free, *half*-hearted, or *ir*rational actions to the extent that they help us to understand the former sorts of actions.[90] The former sorts being essential species makes it appropriate to take them as constituting the proper subject of philosophical attention, as is generally, if implicitly, assumed. But if the standard approach were correct, there would be no more reason for intentional action to be the central focus of philosophical attention than there would be for (Ford's example) nocturnal action to have that privilege, which would be absurd.[91] Furthermore, Ford claims that it is distinctive of species of essential genera, that there can be no conflict between the various species of an essential genus; necessarily, being a member of any essential species is compatible with being a member of every other essential species of that genus, precisely because an essential species and genus are equiprimordial. As he says, 'The anti-privative forms of a thing are always compatible in just this way: the infinite respects in which a body may fail to be deformed, or a shape irregular, or a sample impure, or a law unjust – these necessarily harmonise. But the same is not true of accidental species.'[92] It is quite clear that there could be no obstacle, in principle, to an action being intentional, *and* free, *and* wholehearted, *and* rational, and so on.

What is more important for our purposes, though, is the relation between actions and events. Where the standard approach takes this also to be a relation between an

---

[89] Ibid. pp. 90-1
[90] For examples of such interests, see e.g. Anscombe, *Intention*, Gary Watson, 'Free Agency'; Harry G. Frankfurt, 'Identification and wholeheartedness'; Broome, 'Normative Requirements'; Velleman, *The Possibility of Practical Reason*.
[91] Ford, 'Action and Generality', pp. 95-7
[92] Ibid. p. 98

accidental species and its genus, most saliently on the proposal that actions are events (bodily movements) which, *in addition*, have the right causal aetiology, Ford argues that the relation of *action* to *event* is actually one of a species to categorial genus. If this is right, then actions are events, but not in virtue of being fundamentally events, which, by virtue of further less fundamental differentiae are rendered actions. Actions are events in virtue of being, fundamentally, actions, since *action* is a species of the less fundamental, categorial genus *event*, just as *horse* is a species of the categorial genus *animal*.

If we accept Ford's account of the kinds of generality governing the relations between events, actions, and intentional actions, the repudiation of reductive or analytical ambitions of the account of actions which I suggested earlier should not be regarded as objectionable. Actions are instances of agents' doing things, and we needn't be troubled by the irreducible involvement of an agent in their actions. On the contrary, this secures the presence of agents in the account. Nor should we be troubled by the fact that this account refuses to supply differentiae which, added to something which is already an event, makes it into an action. This would only be needed if the relation between *action* and *event* were one of species to accidental genus, and we have seen that we can instead regard it as one of species to categorial genus. As Anscombe urges, we can reject the need for criteria of individuation for events as such, being satisfied with identifications of specific sorts of occurrences. For example, knowing that there was a poisoning of *A* and a killing of *A*, and that these were actions, we can ask whether someone's poisoning *A* was the same action as someone's killing *A*.

The upshot of this discussion is that there is no difficulty in providing an account of the nature of action, and of the relation between actions and events, facing those who reject the causal theory of action. Indeed, the account of the nature of action that we can give is actually better than the causal theory, yielding a more plausible view of the relations between event, action, and intentional action, on the one hand, while ensuring that agents are included in the picture of agency from the very beginning.[93]

---

[93] Some philosophers (e.g. Morris, *The Good and the True*; Alvarez and Hyman, 'Agents and their Actions') argue that for agents to be included in a picture of agency, actions must not be taken to be events, and so reject the claim that action is any sort of species of event. For example, Morris agrees that actions are not caused by anything, in particular not by an agent's attitudes, and on the basis of this claims that actions are not events (*The Good and the True*, p. 136). However, it is the combination of all three of the views (i) that actions are events, (ii) that the individuation of events just as such is in terms of their causal relations, and (iii) that actions are not caused, which is inconsistent. Morris, having accepted (iii), mistakenly assumes (ii) and rejects (i). But we should follow Anscombe in rejecting (ii), and with it the

4.10 Conclusion

I have argued that we should accept a realist view of value, and take practical reasons to be thick value facts. Such an account meshes well with the philosophical psychology developed in Chapter 3, and provides the first step in showing how to accommodate (EC). Furthermore, I have argued that the standard causal approach to agency should be rejected, and suggested that we adopt Ford's alternative view of action as a categorial species of event. Chapter 5 will focus on the rational explanation of action, arguing that, given this alternative conception of actions and how they stand to events in general, we can offer a satisfactory non-causalist account of rational explanation, which can harmonise the explanatory roles of reasons and of psychological attitudes.

---

accidentalist conception of actions, which leaves us free to accept both (i) and (iii). (This fully allows that other sorts of event are caused.)

Chapter 5

The Explanation of Action

Almost anyone not under the influence of theory will say that, when a person is weak-willed, he intentionally chooses that which he knows or believes to be the worse course of action when he could choose the better course; and that, in acting in this way, the weak-willed man acts not for no *reason at all – that would be strange and atypical – but irrationally.*

~ David Wiggins[1]

5.1 Introduction

Since §4.5, we have been working towards showing how a realist account of value-based reasons can accommodate the following explanatory condition:

(EC)   *R* is a reason for *A* to φ only if (i) *A* could be motivated to φ for *R*, and
(ii) were *A* to φ for *R*, *R* would figure in a true explanation of *A's* φ-ing.

In Chapter 4, I argued that a conception of reasons as thick evaluative facts is able to accommodate (EC)(i), in the manner proposed in §4.5. But I also noted that in order to accommodate (EC)(ii), we need to provide an account of the relation between what an agent believes and desires and their actions, and of the relation between the reasons that an agent has and the actions they perform for those reasons. In this chapter I take up these tasks, with a view to providing an account of the explanation of action, which encompasses what I take to be two distinct, but closely related, kinds of rational explanation: explanation in terms of an agent's psychology ('psychological explanation'), and explanation in terms of their reasons ('reasons-explanation'). The resulting account brings together the account of belief and desire that I defended in Chapter 3, and the accounts of reasons and of actions that I recommended in Chapter 4.

In §4.8, I argued against causal theories of action, which take the relation between an agent's believing or desiring something and their actions to be a causal

---

[1] Wiggins, 'Weakness of Will, Commensurability and the Objects of Deliberation and Desire', p. 241

relation. The resulting rejection of causal theories of action strongly suggests, even if it does not entail, that a causal conception of psychological explanation – the kind of rational explanation which is taken as central by the standard approach – is also mistaken. This is not to suggest that the idea of psychological explanation itself is mistaken, but just that it would be a mistake to think that psychological explanation was a species of causal explanation.

There is, however, a well-known and influential argument of Davidson's which might seem to force a causal conception of psychological explanation upon us. §5.2 considers that argument, and suggests that it should not be seen as forcing such a causal conception on us, but rather as imposing a condition of adequacy, which an alternative to a causal conception of psychological explanation must meet. §5.3 introduces the idea that one can act because of a fact, and so be responsive to it, if and only if one has knowledge of that fact. This promises to allow us to accommodate (EC)(ii). But, in contrast to other versions of this idea, I argue that it is only tenable if explanation in terms of one's knowledge ('knowledge-explanation') and reasons-explanation are not equivalent, and, moreover, if we reject a causal conception of the character of psychological explanation.

§5.4 draws on Hornsby's account of the relation between psychological explanation and reasons-explanation in order to fill this idea out, claiming that knowledge-explanation is a special kind of explanation in terms of belief ('belief-explanation'), while at the same time an understanding of belief-explanation is parasitic on an understanding of knowledge-explanation. §5.5 suggests how a non-causal conception of psychological explanation can meet Davidson's condition of adequacy, and §5.6 offers an account of rational explanation, which harmonises psychological explanation and reasons-explanation. It proposes an account of the relation between an agent's psychological attitudes and their actions, and between their reasons and their actions, and explains the character of the explanations that are thereby provided.

This account of the rational explanation of action is augmented in §§5.7-5.8, by explaining the roles of an agent's character in the explanation of action. These allow us to understand why an agent acts in the light of certain sorts of considerations rather than others, and how this affects both the ethical status of their actions, and the manner in which they are performed. Furthermore, it allows us to explain such phenomena as *akrasia* and *accidie*, which might otherwise be thought to be problematic for a cognitivist theory. Seeing how these phenomena can be understood further clarifies

what is involved in the non-causal account of rational explanation proposed in this chapter.

5.2 Rational Explanation and Davidson's Challenge

There is a famous argument of Davidson's, which might seem to force a causal conception of psychological explanation upon us in spite of our rejection of a causal theory of action. Following Alfred Mele, I will refer to this argument as 'Davidson's Challenge'.[2] Davidson begins his argument thus:

> [A] person can have a reason [i.e. a belief-desire pair] for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action *because* he had the reason. Of course, we can include this idea too in justification; *but then the notion of justification becomes as dark as the notion of reason until we can account for the force of that 'because'*.[3]

The middle sentence here states the fundamental challenge – an account of the character of psychological explanation must substantiate the psychological 'because'. The first sentence points to the kinds of examples in which the challenge appears in a particularly stark form, cases in which an agent has more than one set of attitudes which would rationalise their action, and in which they do not act because of one set of rationalising attitudes, but instead because of another set. Call these 'stark cases'. In stark cases, it is particularly striking that we need to be able to say in virtue of what it is true that they acted because of one rather than another set of attitudes, but the problem is general, applying even in cases in which an agent acted when they only had one set of rationalising attitudes as a candidate to figure in a psychological explanation.[4] The final clause which I have italicised, together with two other statements Davidson goes on to make, form the core of his argument. He continues:

> One way we can explain an event is by placing it in the context of its cause; cause and effect form the sort of pattern that explains the effect, in a sense of 'explain' that we understand as well as any. If reason [i.e. a belief-desire pair] and action illustrate a different pattern of explanation, that pattern must be identified.[5]

---

[2] Alfred Mele, *Motivation and Agency*, pp. 38-9
[3] Davidson, 'Actions, Reasons, and Causes', p. 9, second emphasis added.
[4] For further discussion of numerous examples, in a Davidsonian spirit, see Mele, *Motivation and Agency*, ch. 2.
[5] Davidson, 'Actions, Reasons, and Causes', p. 10

The summation of the argument comes in the following comment:

> I would urge that, failing a satisfactory alternative, the best argument for a scheme like Aristotle's [causal account] is that it alone promises to give an account of the "mysterious connection" between reasons [i.e. rationalising attitudes] and actions.[6]

The resulting argument seems to be something like the following:

(DC1)   An account of the explanatory character of psychological explanation must be able to explain the force of the psychological 'because'.

(DC2)   Causal accounts of the explanatory character of psychological explanation are able to explain the force of the psychological 'because'.

(DC3)   The explanation of the force of the psychological 'because' given by causal accounts fits attitudes and actions into an explanatory pattern that we understand as well as any.

So,

(DC4)   Psychological explanation is a species of causal explanation.

(Strictly speaking, Davidson's challenge is given by (DC1) alone; the whole (DC) argument is supposed to support a causal theory of psychological explanation on the basis of Davidson's challenge.)

Clearly, the form of argument here is an appeal to the best explanation. It appears that (DC3) is supposed to tell us what is good about the proposed explanation of the force of the psychological 'because' as having causal-explanatory force. But it is actually not obvious that (DC3) should be relevant at all. It would be relevant, it seems, only if it were assumed that there was only one kind of explanation, or assumed that having any account of the character of the explanation is better than having none. But the former assumption is obviously false, while the latter, as Morris points out, 'is legitimate only when the adoption of the [explanatory] model plays an instrumental role in the pursuit of some further project, as in the technological application of science; not when the goal is just the understanding provided by that model itself'.[7] Since a philosophical account of the character of psychological explanation is not in this way instrumental, the assumption is not legitimate. This is because, even if it were true that

---

[6] Ibid. p. 11
[7] Morris, 'Mind, World and Value', p. 310

we understood the pattern of causal explanation as well as any explanatory pattern, that supposed fact would not show that the causal account gave the best explanation of the psychological 'because'. Rather, it would show that if psychological explanation were a species of causal explanation, we would understand the force of the psychological 'because' as well as we understand any explanatory 'because', which is quite another matter.

Of course, the most prominent alternatives to a causal account, and those that Davidson was attacking, are teleological accounts of psychological explanation. Both Davidson and Mele compellingly argue that existing versions of these fail Davidson's challenge.[8] Even so, given the problems with the standard causal approach to agency which I discussed in Chapter 4, we ought to be sceptical about whether the mere fact that a causal view manages *a* response to Davidson's Challenge really should support the view that it gives a *good* response.

Furthermore, there is no reason to think that teleological or context-placing accounts are the only possible rivals to causal accounts of the character of psychological explanation. Indeed, I will propose a different sort of rival account. Because of these considerations, I think that we should regard Davidson's challenge as emphasising a condition of adequacy on an account of psychological explanation, but as doing no more than this. The upshot is that accommodating (EC) requires an account of psychological explanation, which requires an account of the relation between an agent's attitudes and their actions, and the accounts of these must meet Davidson's challenge.

5.3 Knowledge-Transparency

According to (EC), an account of reasons must show how there could be *true* reasons-explanations. What is required in order to make sense of the possibility of true reasons-explanations is that we make sense of the idea that someone could be *responsive* to the facts themselves. We need, that is, to be able to see that someone could be *aware* of the facts, take them into account, and respond to them in such a way that they count as acting in the light of the facts.[9]

---

[8] The account given in A. I. Melden, *Free Action*, is Davidson's primary target. For a more recent version, see e.g. Scott R. Sehon, 'An Argument Against the Causal Theory of Action Explanation'; for criticism of this and other more recent versions of teleological accounts of action explanation, see Mele, *Motivation and Agency*, ch. 2.

[9] Cf. Alvarez, *Kinds of Reasons*, p. 25

In order to make the two forms of rational explanation – psychological explanation and reasons-explanation – compatible, we should start by attending to the fact that it is essential to making sense of reasons-explanations that we acknowledge that responsiveness to the facts involves an agent's being *aware* of them. The thought that an agent's responsiveness to the facts involves an awareness of them appears already to bring acting for reasons, and so reasons-explanation, within the scope of an agent's psychology.[10] I propose that we take this appearance seriously in explaining how reasons accommodate (EC), and in giving our accounts of reasons-explanation and psychological explanation.

When an agent is *aware* of the facts, the facts themselves fall within the scope of the agent's cognition, and this is tantamount to saying that the agent has *knowledge* of the facts. This suggests that we can gloss the thought that an agent's responsiveness to the facts brings reasons-explanation within the scope of an agent's psychology as follows: to act on one's *knowledge* of the facts is to be responsive to them, so to act on one's *knowledge of reasons* is to act *for those reasons*.

Both Hyman and Hornsby defend very similar views to this. Hyman, for example, claims that *knowledge*-explanations are 'transparent', where *belief*-explanations are not: 'In every case, *if we can explain the act (belief, desire, feeling, etc.) in terms of the agent's knowledge of a fact, then we can also explain it directly in terms of the fact known*. Knowledge, unlike belief, is transparent: we can look straight through it to the fact.'[11] Hornsby, meanwhile, defends the idea that when we can give a reasons-explanation of an agent's action, this is in virtue of the agent's knowing the relevant facts, and there being a corresponding knowledge-explanation; conversely, she claims that 'where there is a [psychological explanation] from *X* 's knowledge [of *p*], the fact that *p* was a reason *X* had for φ-ing.'[12] Using Hyman's terminology, let us say that both he and Hornsby offer versions of *knowledge-transparency*.

Psychological explanation can be represented in the following schema:

(PE)   *A* φ-ed because she believed that *p* and desired that *q*,

while reasons-explanations can be represented in the schema:

---

[10] Cf. Jennifer Hornsby, 'A Disjunctive Conception of Acting for Reasons', p. 251
[11] Hyman, *Action, Knowledge, and Will*, p. 157
[12] Hornsby, 'A Disjunctive Conception of Acting for Reasons', pp. 251-2

(RE)   *A* φ-ed because *p.*

According to knowledge-transparency, what allows the possibility of true instances of (RE) is that there are true instances of the knowledge-explanation schema:

(KE)   *A* φ-ed because she knew that *p.*

When, and only when, there is a true instance of (RE), there is a true instance of (KE). It is not implausible to think that what knowledge-transparency amounts to is just the claim that (RE) and (KE) are equivalent. For instance, this might be suggested by one of Hornsby's formulations: '*A* φ-ed because *A* knew that *p,* so that *A* φ-d because *p.*'[13] Furthermore, both Hyman and Hornsby embed their versions of knowledge-transparency within accounts of rational explanation which take psychological explanation to be a species of causal explanation. However, as I will now argue, if either knowledge-transparency amounted to the claim that (RE) and (KE) are equivalent, or if psychological explanation were a species of causal explanation, then knowledge-transparency would be untenable. Neither Hyman's nor Hornsby's version of knowledge-transparency, therefore, is acceptable.

Consider, to begin with, the following argument, which endorses both that equivalence thesis and a particular causal account of psychological explanation:

(Tr1)   Knowledge-explanations are psychological explanations.

(Tr2)   Psychological explanations identify psychological attitudes in the causal aetiology of actions.

(Tr3)   If knowledge-explanations are equivalent to reasons-explanations, then either they do not identify psychological attitudes in the causal aetiology of actions, or reasons are (standardly) facts about psychological attitudes in the causal aetiology of actions.

(Tr4)   Knowledge-explanations are equivalent to reasons-explanations.

So

---

[13] Ibid. p. 252

(Tr5)   Either knowledge-explanations do not identify psychological attitudes in the causal aetiology of actions, or reasons are (standardly) facts about psychological attitudes in the causal aetiology of actions.

(Tr6)   Reasons are not (standardly) facts about psychological attitudes in the causal aetiology of actions.

So

(Tr7)   Knowledge-explanations do not identify psychological attitudes in the causal aetiology of actions.

But

(Tr8)   If (Tr7), then knowledge-explanations are not psychological explanations

So

(Tr9)   Knowledge explanations are not psychological explanations.

(Tr9) is the negation of (Tr1), so this purports to be a *reductio ad absurdum* of knowledge-transparency, when interpreted according to (Tr4), and when psychological explanation is taken to be causal explanation, as this claim is construed in (Tr2). In fact, while it could be claimed, against (Tr4), that knowledge-explanation is necessary and sufficient for reasons-explanation, but that they are not equivalent, it is hard to see what could even be meant by claiming that psychological explanation is a species of causal explanation if this claim is not to be construed according to (Tr2). To be clear, I am *not* claiming that the (Tr) argument directly threatens either Hyman's or Hornsby's position, since in fact I take it that each of them reject one, or another, of its premises; Hyman, I think (although he is not altogether clear on this matter), would reject (Tr4), while Hornsby would reject (Tr2). What I will argue against each of them is that by rejecting one, or the other, of these premises, respectively, but not both of them, their versions of knowledge-transparency are rendered unsatisfactory. Before examining their rejection of, respectively, (Tr4) and (Tr2), however, let me consider what may seem the more obvious premise to reject: (Tr3).

It may be initially tempting to reject (Tr3); after all, (Tr3) is not *obviously* true. But on closer inspection, its rejection turns out not to be plausible. The *explanantia* in (RE) and (KE) are different; in the former it is that *p*, but in the latter it is that *A* knew that *p*. Since (i) an *explanans* gives a *reason why* the *explanandum* is true, (ii)

explanation is factive, and (iii) the fact that p is different to the fact that A knew that p,[14] (RE) and (KE) do not seem to explain the same thing – A's φ-ing – equivalently (contra (Tr4)); they explain A's φ-ing by giving *different reasons why*. For their explanations to be equivalent (as per (Tr4), which is the antecedent of (Tr3)), one would have to be assimilated to the other, i.e. either (KE) would have to not identify psychological attitudes in the causal aetiology of actions in their *explanantia*, or the reasons identified in the *explanantia* of (RE) would (standardly) have to really be facts about what an agent knew, so that 'p' in (RE) would be substitutable for clauses of the form 'A knew that p' (in which a psychological attitude in the causal aetiology of actions would be identified). This, however, is precisely what is claimed by (Tr3).

Now let us consider Hyman's apparent rejection of (Tr4). It seems that Hyman rejects (Tr4) on the basis of the sorts of considerations adduced in the first half of the previous paragraph. So, on Hyman's view, there is a knowledge-explanation when, and only when, there is a reasons-explanation, but this needn't imply that they are equivalent.[15] If they were not equivalent, however, and psychological explanation were a species of causal explanation, what sort of explanation would be given by an instance of (RE)? If it were suggested that reasons-explanation were also causal explanation, the natural way to understand how these could be compatible would be by taking a reason to causally explain one's knowledge, and one's knowledge to causally explain one's action. But there would be nothing to prevent a reason from causing one to have a merely true belief, which would in turn cause one's action. If this is right, then the relation that a reason would stand in to one's action when it causally explained it would be no different from when it did so via true belief. But the point of insisting on knowledge-transparency is to insist that knowledge allows explanation by the facts, where mere true belief does not, so Hyman seems to be committed to denying that reasons causally explain one's actions, via one's knowledge causing them. This, however, leaves Hyman with nothing to say about what sort of explanation reasons-explanation is supposed to be. In effect, by denying (Tr4), Hyman fails to explain what the explanatory character of reasons-explanation is. This is what makes his version of knowledge-transparency unsatisfactory.

---

[14] See Hyman, *Action, Knowledge, and Will*, pp. 137-9, 147-58

[15] That this is his view is suggested by comments such as that 'true belief plus causation […] are insufficient to make a standard because-explanation true. Whereas knowledge is *sufficient*' (ibid. p. 157, emphasis added), and many other similar formulations. But see his comment that 'doing something because of a fact and doing it "really in consequence of knowing" the fact are one and the same thing' (also on p. 157) for an indication that he may accept (Tr4) after all.

Hornsby, meanwhile, would reject (Tr2). Since she insists that psychological explanation is a species of causal explanation, she would have to reject the following conditional from which it would be validly inferred:

(BEC)　If A φ-ed because she believed that p, her believing that p was in the causal aetiology of her φ-ing.[16]

(BEC), though, would be the obvious way of giving substance to the claim that psychological explanation is a species of causal explanation. Yet Hornsby provides nothing in the way of an alternative conception of what it would *be* for psychological explanation to be a species of causal explanation. In fact, she rejects the thesis of the nomological character of causation, and thereby rejects the very explanatory pattern that Davidson claims would allow us to understand the explanatory force of the psychological 'because' as well as any, if psychological explanation were a species of causal explanation.[17] It seems, then, as if Hornsby not only fails to give an account of the explanatory character of psychological explanation, but that she actually undermines what motivation comes from Davidson's challenge for thinking that it is a species of causal explanation in the first place, even while she appeals to Davidson's argument to support her insistence that it *is* a species of causal explanation.[18]

What we should take from this discussion, I suggest, is that to give a viable interpretation of knowledge-transparency, in the service of accommodating (EC), we need to reject *both* the interpretation of it on which (KE) and (RE) are equivalent, and the view that psychological explanation is a species of causal explanation. At the same time, though, we need to give an alternative, non-causal, account of the character of psychological explanation and of reasons-explanation.

---

[16] See e.g. Jennifer Hornsby, 'Causality and "the mental"', pp. 134-5; 'Agency and Causal Explanation', pp. 134-5; 'Anomalousness in Action', pp. 627-9
[17] See e.g. Jennifer Hornsby, *Simple Mindedness*, pp. 78-80; 'Agency and Causal Explanation', pp. 138-9
[18] See e.g. Hornsby, 'Agency and Causal Explanation', p. 133; 'Anomalousness in Action', p. 627; 'Actions in Their Circumstances', pp. 124-5

5.4 Rational Explanation and Knowledge-Translucency

What we require is a way to understand knowledge-explanation as a special kind of non-causal psychological explanation, which allows us, in Hyman's phrase, to 'look straight through'[19] to a reasons-explanation, but without this implying that the two are equivalent. This means that we require a unified non-causal account of psychological explanation and of reasons-explanation, which (i) provides an understanding of the explanatoriness of psychological explanation capable of answering Davidson's challenge, (ii) treats knowledge-explanation as a kind of psychological explanation, and (iii) explains what a reasons-explanation, which can be 'seen through' a knowledge-explanation, adds to the psychological explanation already afforded by the latter. In this section I shall set the groundwork for the general shape of what such a unified account of psychological explanation (in general), knowledge-explanation (in particular), and reasons-explanation, might look like. In §§5.5-5.6, I will tentatively offer a proposal for an account of this general shape.

What we need to make sense of, I suggest, is that knowledge-explanations allow us to see straight through them to reasons-explanations, but at the same time do not become strictly transparent, so that all the explanation that can be seen is the reasons-explanation. Since knowledge-explanations are psychological explanations, the psychological character of their explanatoriness must not become invisible when they are given; we need to be able to see through them to reasons-explanations while simultaneously seeing that what we are looking through are, precisely, psychological explanations. A version of knowledge-transparency which makes sense of this, would be aptly characterised as 'knowledge-*translucency*'.

Hornsby's disjunctive approach to rational explanation supplies a way to provide for knowledge-translucency. On Hornsby's approach, we are to understand psychological explanation in terms of an agent's beliefs disjunctively, as follows:

(DisA) If A φ-d because A believed that p, then
    EITHER: A φ-ed because A knew that p, so that A φ-d because p.
    OR: A φ-ed because A merely believed that p.[20]

What this disjunctive approach achieves is an understanding of knowledge-explanation as itself a special case of belief-explanation. This is problematic in

---

[19] Hyman, *Action, Knowledge, and Will*, p. 157
[20] Hornsby, 'A Disjunctive Conception of Acting for Reasons', p. 252

Hornsby's account, since she thinks that belief-explanation is a species of causal explanation, and so is committed to thinking of knowledge-explanation in the same way. But, having rejected the view that psychological explanation is a species of causal explanation, we can accept her (DisA) without being forced to accept that knowledge-explanation is a species of causal explanation. Hornsby also argues that, while knowledge-explanation is treated as a form of belief-explanation on the disjunctive approach, an understanding of belief-explanation is parasitic upon an understanding of knowledge-explanation, or explanation by the facts known:

> [W]hen one explains a person's acting in some way by saying what they believe, one relies upon a grasp of what is actually a reason for so acting [....] Thus a distinctive sort of explanatory interest is in play when [belief-explanations] are seen to be at work; but the understanding achieved by knowing [a belief-explanation] is rested in an understanding of the operation of [knowledge- or reasons-explanations].[21]

What Hornsby seems to have in mind is that in giving a psychological explanation of someone's action, we explain their doing what they did in terms of their *taking* something to justify their doing it (i.e. *taking* there to be a reason to do it), but to understand this sort of explanation depends on already understanding what it would be for there to *actually* be a reason for them to do it, which might have explained their action.[22] This may be so even though knowledge-explanation is a form of belief-explanation. This means, she claims, that psychological explanations and reason-explanations 'must stand or fall together'.[23]

We can incorporate these two features – that knowledge-explanation should be treated as a form of belief-explanation, and that an understanding of belief-explanation is parasitic on an understanding of knowledge-explanation, or explanation by the fact known – in one stroke, by treating cases in which one believes that p as cases in which it is as if one knows that p.[24] Knowledge will count as belief, on this approach, since it is trivially true that if one knows that p it is as if one knows that p. When a belief-explanation can be given, this will be such that it is as if a knowledge-explanation can be given; knowledge-explanation, then, will be a special case of belief-explanation

---

[21] Ibid. p. 258-9

[22] This would mean that what Davidson, 'Actions, Reasons, and Causes', p. 9, describes as a 'somewhat anaemic sense in which every rationalisation justifies' an agent's actions would depend on an understanding of a full-blooded sense of justification.

[23] Hornsby, 'A Disjunctive Conception of Acting for Reasons', p. 259

[24] Note that this is not supposed to provide an analysis of belief. Cf. Williamson, *Knowledge and Its Limits*, pp. 45-8.

because it is trivially true that when a knowledge-explanation can be given, it is as if a knowledge-explanation can be given. Furthermore, an understanding of belief-explanations is essentially parasitic upon an understanding of knowledge-explanations, since one must understand the character of knowledge-explanation to be able to understand an explanation which is as if it were knowledge-explanation. Providing an account of the explanation of action, then, will begin by giving an account of the explanatory character of knowledge-explanation, and the reason-explanation it affords, and then explain a parasitic or derivative explanatory character of mere psychological explanation.

5.5 Davidson's Challenge: The Shape of an Answer

If psychological explanation in general, and knowledge-explanation in particular, is to be genuinely explanatory, it has to answer Davidson's challenge. It must explain the force of the psychological 'because', so that it can explain the difference between A's φ-ing *because* she believed that p and desired that q, and A's φ-ing *while* (but not because) she believed that p and desired that q. There is a simple way to answer Davidson's challenge – we just need to insist that the *essential nature* of an action is actually dependent on those attitudes because of which the agent performed them. If we insist on this, then, from the fact that some *particular* action is the one we are considering, it will follow that the agent performed it *because* she had certain specific beliefs and desires, and there is no room for a contrast with a counterfactual case in which the agent did *that very action*, but not because she had those very same beliefs and desires.[25] The relation between an action and the attitudes an agent had, *because* of which she performed it, on this view, is, not one of causation, but of partial constitution: attitudes are partially constitutive of the actions they explain.

This is not an unprecedented view, although those who have espoused it do not always seem to have been fully cognizant of the significance of what they claimed, and it has often been incoherently *combined* with a causal view of the character of psychological explanation. Take, as a first example, Kant's view of action. For Kant, 'an action from duty has its moral worth *not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon, and therefore does not depend upon

---

[25] This is not to say that any identification of an action is sufficient to provide knowledge of its essential nature, and so knowledge of which of the agent's attitudes explain it. This is something we can learn, by learning why the agent did it.

the realisation of the object of the action but merely upon the *principle of volition* in accordance with which the action is done'.[26] It is plausible to interpret Kant as holding the view that it is an *essential property* of an action that it has the moral worth it does. So if, say, Kant's shopkeeper treated his customers honestly from duty, he would have performed a different, morally worthy, action, than if he had done so from self-interest, to achieve some advantage for himself, which would have resulted in an action lacking moral worth.[27] What the nature of one's action is, therefore, would depend, on this interpretation of Kant, on the maxim in accordance with which one performed it; that is why, in order to consider the moral status of *actions themselves*, we have to attend to their *psychological grounds*, on Kant's view. As we saw in §3.2, however, Kant combines this with a quasi-Humean causal view of the explanation of action.

As a second example, take Aristotle's account of action. On J. L. Ackrill's interpretation, 'It would appear […] that *what* action precisely has been performed – what action is genuinely explained by the *arche* [originating principle] in the agent – depends on what the object of thought and desire was.'[28] Furthermore, the point of focusing on the role of the *arche* is to restrict our attention to actions which are done intentionally (*hekousiōs*), and, as Ackrill explains, for Aristotle 'a man's actions, properly speaking, for which he can be praised or blamed, are confined to what he does if not from choice at least *hekousiōs*.'[29] Once again, it seems clear that the *essential nature* of an action, in Aristotle's view, depends on the attitudes – the thoughts and desires, or the *arche* – which explain it, and it is natural to understand this dependence as a constitutive dependence. It is not just its occurrence that is explained in giving a psychological explanation, but its essential nature – '*what* action […] has been performed' – which makes someone's action praiseworthy or blameworthy. However, like Kant, Aristotle combines this with a causal account of the explanation of action.

Third, Morris argues that answering Davidson's challenge, and so saying what is explanatory about psychological explanations if they are not causal, requires meeting a general constraint on explanations, which is 'that it meet something like the following counterfactual condition: if what we offer as explanation had not been true, there would not have been the thing we are trying to explain'.[30] But this, he realises and exploits, is

---

[26] Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. by Mary Gregor, p. 13 (4:399-400)
[27] For the example see ibid. p. 11 (4:397)
[28] J. L. Ackrill, 'Aristotle on Action', p. 99
[29] Ibid. p. 94
[30] Morris, 'Mind, World, and Value', p. 315

yielded by a constitutive relation just as well as by a causal relation; on his account, beliefs and desires are literally 'intrinsic to behaviour'.[31] Morris's focus is on explaining what it is for someone to believe or desire something at all, which for belief, he thinks, is for it to be 'legitimate, in virtue of how she is at that time, to count what she does as foolish or otherwise according to whether or not it is false that p', while 'someone has a desire at a certain time just in case it is legitimate, in virtue of how she is at that time, to count what she does as bad or otherwise according to whether or not that is a bad thing to want'.[32] Because of this focus, he takes it that the account of what it is to believe or desire something directly provides a constitutivist account of psychological explanation. What it is, according to him, for someone to have φ-ed because they believed that p, is for it to be legitimate to count their φ-ing as foolish or otherwise according to whether or not it was false that p. This is supposed to be satisfactory because it secures a counterfactual dependence between one's action, and one's belief that p, when the latter explains the former, and the character of the explanation is supposed to be exhausted by such counterfactual dependence, which is secured by the constitutive relation between them. But in fact it is not satisfactory.

First, as I will argue in §5.6, we can give a constitutivist account of psychological explanation which is capable of giving a more interesting and informative explanation of the explanatory character of psychological explanation, which goes beyond mere counterfactual dependence. Second, though, it is not clear that Morris's account is sustainable, even if we were to ignore the thinness of what it takes to be involved in the explanatory character of psychological explanation. The reason for this is that it does not acknowledge a sufficiently large gap between what it is to believe or desire something, and what it is for one's actions to be explained by one's believing or desiring something. As we saw a moment ago, Morris takes the criteria for one's believing something and for one's actions being explained by one's believing something to be identical. That makes it seem as if, on the one hand, it is constitutive of one's believing that p that it is legitimate to count some of one's actions as foolish or not according to whether or not it is false that p, but, on the other hand, for some particular action of one's to be explained by one's believing that p is for it to be one of those actions which is constitutive of one's believing that p. That would mean, however, that according to Morris, someone's believing that p would be partially constituted by

---

[31] Ibid.
[32] Ibid. pp. 308-9

their doing something which was partially constituted by their believing that *p*. But this has a look of circularity to it; it's just not clear that constitution running in both directions like this is intelligible.

Morris's account, therefore, lacks the resources to distinguish between what it is for one to believe that *p* and what it is for one's action to be *explained by* one's believing that *p*. But we need to be able to distinguish these if the former is to be constitutive of the latter, and so secure the counterfactual dependence involved in psychological explanation. In order to make the required distinction, the criteria for the former and for the latter cannot be identical, as they are on Morris's account.

Finally, take Dancy's discussion of Davidson's challenge:

> Suppose that there are two distinct representations [i.e. psychological attitudes] present to me, each one sufficient for me to act in these circumstances, and each standing as a reason for the same act as the other. What is the matter of fact which makes it the case that the first of these is in fact the one that motivates me and the second not?[33]

The answer he gives is this:

> As it stands, the question makes an unjustified assumption, namely that the identity of an act is independent of its motivation, so that one and the same act can be done for different reasons. If this assumption is straightforwardly false, so that a difference in motivation necessarily makes a difference in act, [the] question evaporates. And my present view is that the assumption is false. We do, of course, naturally speak of repaying money to escape further trouble rather than out of a sense of shame or simply because we see it is the right thing to do. But I hold here that the act in question differs as a *moral* object according to which of these did in fact motivate it, and that we should not say that a person who pays his debts out of duty does the same thing as one who pays them out of fear.[34]

Firstly, from Davidson's perspective, what Dancy here calls an unjustified assumption would not appear unjustified at all; on standard (Humean) conceptions of causality, it is axiomatic that the nature of the cause cannot be inferred from the nature of the effect. So, really, what Dancy calls an unjustified assumption *follows from* the assumption that psychological explanation is causal explanation. This suggests both that thinking that the *nature* of actions is dependent on the attitudes because of which the agent performed them is incompatible with psychological explanation being a species of

---

[33] Dancy, *Moral Reasons*, p. 42
[34] Ibid. pp. 42-3. In *Moral Reasons*, Dancy also accepts a causal account of the character of psychological explanation, although he has since rejected such a conception of it; see e.g. Jonathan Dancy, *Practical Reality*, and 'On Knowing One's Reason'.

causal explanation, and also that there may be something question-begging about Davidson's argument for the causal view.

Secondly, and more importantly, however, it again seems clear that Dancy is taking the essential nature of an action to be partly constituted by the agent's attitudes because of which they performed it. What is particularly striking about his way of putting that view – that an action 'differs as a *moral* object' according to which attitudes explain it – is that it crystalises a feature of each of the examples I have given; Kant, Aristotle, Morris, and Dancy all take the essential nature of actions to be crucial to their correct moral, ethical, or more generally evaluative appraisal.

Thirdly, however, there is an important mistake in how Dancy elaborates this view, which we must take care first to illuminate and then avoid. The issue turns on his use of the phrases 'one and the same act' and 'does the same thing'. When Dancy speaks of acts in the first of these, he would most naturally be interpreted as speaking of *actions*, which are particular events of an agent's doing something intentionally; when he speaks of doing the same thing in the second, though, he is most naturally understood as speaking of a universal, a *thing done*, of which *the agent's doing it* (their action) is an instance (see the discussion of Hornsby's distinction between *actions* and *things done* in §2.4). Now, the contrast which is relevant for the stark case of Davidson's challenge is one between two possible worlds, where it is assumed that it is possible *in each* that someone performed the *same action*, but where their performance is explained in each by *different* psychological attitudes. The claim that the essential nature of an action is constitutively dependent on the attitudes because of which the agent performed it rules out this being a contrast between genuine possibilities. But Dancy proceeds from correctly invoking and rejecting *this* contrast, to speaking of two people who repay debts out of different motives as not doing the same thing. It is possible to interpret him as misleadingly using the second phrase here to speak of (particular) *actions*, but then his claim would be utterly trivial, for no two people could ever (individually) perform numerically identical actions, and this has nothing to do with their possibly differing motivations.[35] But if, as is the more natural interpretation, he is here talking about (universal) *things done,* then what he says is clearly false. The person who repays their debts out of duty, and the person who does so out of fear, manifestly have something in common that they do, namely repay their debts. (This is

---

[35] The qualification in parentheses is needed to allow the possibility that there can be genuinely *collective* action.

true even with respect to the previous contrast between possible worlds: in the two worlds the same agent does the same thing, even though they do not perform the same action.)

In fact, it is quite clear that what Dancy really does mean is captured by the second interpretation, on which it is not trivial but false. For in a note to this passage, he says: 'I have borrowed this view from Joseph', [36] and Joseph, even more explicitly than Dancy, expresses the false view:

> The same act, an objector might say, may surely be done from different motives; and therefore the act must be something, irrespective of the motive. But are they really the same act? Different acts, having different motives, may work themselves out into the same movements of bodies; but these are not the acts. A man who was fond of oysters might eat a plateful put before him for the sake of their flavour; a man who loathed them might do so to avoid hurting his host's feelings; a man who loathed or was indifferent to them might do so to prevent his neighbour, whom he knew to be fond of them and he disliked, from having two portions [….] *They are not three instances of one act, viz. eating a plateful of oysters.*[37]

Joseph's conclusion in the final sentence of this passage is an extremely clear formulation of the mistake I just accused Dancy of making. Joseph is wrong to say that these three actions are not all eatings of oysters, since something that all three men do (intentionally) is eat oysters; but even though all three actions are eatings of oysters, their essential natures are different (and the differences are not just those which are the trivial consequence of their being performed by three different men). This, however, has no tendency to undermine the view that *particular actions*, in their essential nature or as 'moral objects,' to use Dancy's phrase, are constitutively dependent on the agent's attitudes which truly explain them.

5.6 The Rational Explanation of Action

Having clarified these matters, let us return to our main topic: the nature of rational explanation, encompassing both psychological explanation and reasons-explanation. Davidson's challenge can be answered by taking the relation between an agent's action and her attitudes because of which she performed it to be a relation of partial constitution: an agent's attitudes are partially constitutive of the actions they explain.

---

[36] Dancy, *Moral Reasons*, p. 58, note 2
[37] Joseph, *Some Problems in Ethics*, p. 45, emphasis added. Much of chapter II and the beginning of chapter III of Joseph's book is concerned to argue that, as he puts it, an act 'includes' its motive, which I take to be another, somewhat confused, version of the view that the motivating psychological attitudes of an action are partly constitutive of it.

The foregoing discussion also brought out the fact that the essential nature of actions is crucial to their ethical and generally evaluative appraisal. So what we need now is to say how this relation of partial constitution is to be understood, to explain the way in which this is crucial to the relevant appraisal, and to see what this reveals about the character of psychological explanation.

The normative account of belief and desire in Chapter 3 affords a straightforward and appealing way to do all of these things. According to that account, belief is subject to the constitutive norm:

(NB)　For any *A* and any *p*: (i) *A* ought to (believe that *p*) only if it is true that *p*;

　　　and,

　　　(ii) if it is not true that *p*, *A* ought to (not believe that *p*)

Desire, meanwhile, is subject to the constitutive norm:

(ND)　For any *A* and any *p*: (i) *A* ought to (desire that *p*) only if it would be good if *p*;

　　　and,

　　　(ii) if it would not be good if *p*, *A* ought to (not desire that *p*)

Roughly, the basic beliefs and desires involved in psychological explanation would, schematically, be the desire that *p,* and the means-end belief that if one φ-ed one would bring it about that *p*.[38] We can characterise the way in which these attitudes are partially constitutive of one's action as follows:

---

[38] If one believed that if one φ-ed one would bring it about that *p*, one would have a belief suitable, according to the rough formulation, to figure in a rationalising explanation. But φ-ing might be a very *bad* way to bring it about that *p*, and so might be taken to motivate formulating the kind of belief involved in rationalising psychological explanation along the lines of: *A* believed that φ-ing would be as good a way as any to bring it about that *p*. (I owe this sort of formulation to Michael Morris.) This would make the psychology involved in rationalising psychological explanation doubly evaluative. I ignore this complication in what follows.

(PC) For any *A*, φ, and *p*: For *A* to have φ-ed because she believed that if she φ-ed she would bring it about that *p,* and she desired that *p,* is for *A*'s φ-ing, in its essential nature, to have been such that:

(i) it ought to have been performed only if it was true that (a) *A*'s φ-ing would have brought it about that *p*, and that (b) it would have been good if *p*; *and,*

(ii) if it was not true that (a) *A*'s φ-ing would have brought it about that *p*, or that (b) it would not have been good if *p*, then it ought not to have been performed.[39]

As we saw earlier, since if one believes that *p* it is as if one knows that *p*, it is trivially true that if one knows that *p*, then one believes that *p*, since if one knows that *p* it is as if one knows that *p*. (PC), then, despite describing what it is for *A* to have acted because she believed and desired (which is to say: believed evaluatively) certain things, is neutral between *A*'s knowing these things or merely believing them (without knowing them). Since an understanding of psychological explanation is parasitic upon an understanding of knowledge-explanation, we need to begin by interpreting (PC) for a case where the agent acts because they have knowledge. Since, given the argument of §3.9, to desire something just is to believe that it would be good, there are two bits of knowledge involved in psychological explanation. Not only can one's means-end belief amount to knowledge, but so can one's evaluative belief. Furthermore, given the argument of §§4.5-4.7, *neither* of these bits of knowledge is knowledge of a reason for one to φ, so we need to add that into the mix as well. A reason, I claimed in §4.7, is a thick evaluative fact which makes it the case that it would be good if *p*, and which, ideally, would be one's reason for desiring that *p*, or believing that it would be good if *p*; for one's actions to be responsive to a reason one must have knowledge of it.

The ideal case of explanation in the presence of full knowledge would be one of explanation of the actions of a fully virtuous person, in the Aristotelian sense of the person in full possession of *phronēsis*, or practical wisdom. The virtuous person in possession of *phronēsis* knows what their situation is, and has a special kind of conception of their situation through which they can tell what is of genuine significance

---

[39] There is no reason that the psychological explanation of action must be limited to just one means-end belief and one desire. The characterisation of partial constitution could be expanded indefinitely to accommodate however many attitudes were truly explanatory of one's action.

in it, and so what sorts of actions it demands, from the perspective of virtue.[40] Suppose that one's φ-ing is what the situation demands, and so, from the perspective of virtue, it would be best if one were to φ (perhaps φ-ing is rescuing a child who has fallen in front of oncoming traffic). The virtuous person knows that it would be good (or best) if p (e.g. that the child is rescued), and knows that they can φ by ψ-ing (they can rescue the child by pulling them out of the road), and that their φ-ing would bring it about that p (their rescuing the child would bring it about that the child is rescued).[41] We can formulate an instance of (PC) for this example:

> (PC*) For any A, φ, and p: For A to have pulled the child out of the road because she believed that if she pulled the child out of the road she would bring it about that the child was rescued , and she desired that the child was rescued, is for A's pulling the child out of the road, in its essential nature, to have been such that:
>
> (i) it ought to have been performed only if it was true that (a) A's pulling the child out of the road would have brought it about that the child was rescued, and that (b) it would have been good if the child was rescued; *and*,
>
> (ii) if it was not true that (a) A's pulling the child out of the road would have brought it about that the child was rescued, or that (b) it would not have been good if the child was rescued, then it ought not to have been performed.

Since knowledge is factive, it is guaranteed that it is true that A's pulling the child out of the road would have brought it about that the child was rescued, and that it would have been good if the child was rescued, so (PC*)(i)(a) and (PC*)(i)(b), the *necessary* conditions for it to be true that A's action ought to have been performed, are guaranteed to be met. That knowledge is the ability to be guided by the facts ensures that in acting, A was guided by these facts. By the same token, it is guaranteed that (PC*)(ii)(a) and (PC*)(ii)(b), the *sufficient* conditions for it to be true that A's action

---

[40] See e.g. John McDowell, 'The Role of *Euadaimonia* in Aristotle's Ethics', and 'Some Issues in Aristotle's Moral Psychology'.

[41] By introducing this double instrumental relation between ψ-ing and φ-ing, and between φ-ing and bringing it about that p, we allow for cases in which, first, the φ-ing mentioned in (PC) is not something that one can do directly and, second, cases in which the 'p' in one's belief that it would be good if p, does not concern one's own actions.

ought *not* to have been performed are not met. Furthermore, since this is a case in which, *ex hypothesi*, what *A* did was best, or what virtue required in the situation, this guarantees that a correct evaluative appraisal of *A*'s action would find that it ought to have been performed. But even if it were not a case of a virtuous agent doing what would be best in the situation, their means-end and evaluative knowledge would, plausibly, guarantee that their action was at least permissible; knowledge looks as if it guarantees that actions are justified at least to that extent.[42] The character of the psychological explanation is purely justificatory. But, by accounting for the counterfactual dependence introduced by the psychological 'because' of their actions on the agent's attitudes in terms of a relation of partial constitution, the justificatory relation exhausts the character of psychological explanation, rather than requiring supplementation by a causal relation. This explains how we are to understand the relation of partial constitution between an agent's attitudes and the actions they explain, the way in which it is crucial to the relevant evaluative appraisal, and what this reveals about the character of psychological explanation.

What, then, of the reasons-explanation that knowledge lets us see through to? Well, in our example, there will be several reasons for *A* to pull the child out of the road. The most obvious reason is that the child is in immediate danger. Another seems to be that it would be courageous if one pulled the child out of the path of the oncoming traffic (doing so would be something that involved undertaking a risk to one's own safety in the face of fear). Further reasons might be that one will receive praise, gratitude and attention for rescuing the child. Making use of these different reasons, we can distinguish several importantly different cases. The virtuous person conceives of their situation in such a way that they realise that the relevant feature of it is that the child is in immediate danger, and that nothing else warrants consideration, e.g. the risk to their own safety that would be involved in rescuing the child is not something to be considered, not something to be weighed against the child's being in immediate danger. This is the reason because of which they pull the child out of the road. Furthermore, it is because they see that the child's being in immediate danger is of exclusive relevance that they judge that it would be good if they rescued them, in fact that it would be best, and what the situation demands, from the perspective of virtue.

---

[42] There is, however, a possible complication due to the fact that (PC)(ii) states a sufficient, but not a necessary, condition for it's being the case that one's action ought not to have been performed, which I consider in §5.8.

We may distinguish a range of cases in which the agent falls short of virtue in one way or another, without lacking the relevant knowledge. The virtuous agent is an ideal, but the one who comes as close as possible to this ideal without meeting it is what Aristotle calls the continent person. The continent person is also able to tell that what would be best in the situation, what is demanded by it, from the perspective of virtue, is that they rescue the child. Like the virtuous person, they can tell that the significant thing in the situation is that the child is in immediate danger, and they pull the child from the road for this reason. But for the continent person, in contrast to the virtuous person, other considerations, such as the risk to their own safety, are not, in McDowell's phrase, 'silenced',[43] by their conception of the situation according to which it is the child's being in immediate danger which is of significance, and which determines what the situation demands. That the child is in immediate danger is a consideration that they weigh against competing considerations, such as that rescuing them would involve risking their own safety, and they find it to be weightier.

Another kind of case is one in which the agent is, as it were, an ethical novice. On the Aristotelian view, becoming virtuous is a process of habituation, in which one *practices* acting well. The ethical novice can tell that the child is in immediate danger, and can see that pulling the child from the road would be courageous, and that doing so would involve a risk to their own safety, that they are fearful of taking such a risk, and so on. But, unlike the continent person, they cannot tell straight off that the child's being in immediate danger is the relevant thing in the situation. Instead, they need to work out what the virtuous person would do in their situation, and through, as it were, a process of triangulation from whatever information is available to them, works out that the virtuous person would do the courageous thing, that courage is what is called for. Realising this, they imitate the virtuous person, pulling the child from the road because it would be courageous to do so. But, plausibly, this could not be their only reason; in addition, they would act for the reason that it is what the virtuous person would do.[44]

A final sort of case that I will consider here is that of an agent who knows that they will receive praise, gratitude, and attention for rescuing the child. In a simple version of this case, these are the reasons for which they pull the child from the road, and in this simple case, they would not have acted in an ethical way. In Kant's terms,

---

[43] See e.g. McDowell, 'The Role of *Eudaimonia* in Aristotle's Ethics', p. 18
[44] Cf. L. A. Kosman, 'Being Properly Affected', p. 112, and also Burnyeat, 'Aristotle on Learning to Be Good'.

their action would not have moral worth, or in Kosman's, their action would not merit praise, simply because they would have done the courageous thing, which is what virtue requires, for the wrong sort of reason. In a more complicated version of the case, they might realise that they would receive praise and attention *for doing the courageous thing* (not just for rescuing the child) and so do the courageous thing, pull the child from the road, *instrumentally* for the *reason that it would be courageous to do so*, in order to receive praise and attention. In this case, while the person pulls the child from the road because it would be courageous to do so, like the ethical novice, they act for that reason because doing that – pulling-the-child-from-the-road-because-it-would-be-courageous – instrumentally serves receiving praise and gratitude, and so their action still lacks moral worth. At best, the person's action would be a kind of ethical grandstanding, or, perhaps, an ethically objectionable kind of moral fetishism of the sort that Smith criticises.[45]

Ex hypothesi, it would be courageous if one pulled the child from the road; in all of these cases, therefore, what the person does is courageous, since in all of them the person does in fact pull the child from the road. That they do something courageous, therefore, cannot be the contribution made by their acting for a reason. However, when the virtuous or the continent person pull the child from the road for the reason that they are in immediate danger – when they do so for the ethically right reason, that is – their action, and not just the thing they do, is courageous.[46] When the virtuous person acts for this reason, their action is special in that they not only perform a courageous action, but they do so courageously or virtuously.[47] I will return to this in §5.8. Plausibly, even when the ethical novice does so for the contrasting reason that it would be courageous to pull the child from the road, their action is courageous.[48] But when the person pulls the child from the road for the reason that they themselves will receive praise, gratitude and attention, their action is not courageous; it is unethically selfish, or calculatingly manipulative, or grandstanding. *These* are the sorts of differences which are due to the various reasons for which these different sorts of agents pull the child from the road. Furthermore, since in each case the person rescues the child by pulling them from the road, and therefore their pulling the child from the road was the same action as their

---

[45] Smith, *The Moral Problem*, pp. 74-6; see also Bernard Williams, 'Persons, character and morality'.
[46] Cf. Aristotle, *The Nicomachean Ethics*, 2.4; and Julia Markovits, 'Acting for the Right Reasons'.
[47] Cf. Aristotle, *The Nicomachean Ethics*, 1105a30-1105b10
[48] Even so, however, we might want to distinguish between actions which are e.g. courageous only in a weak sense, and those that are, in a strong or full sense, courageous, as they are when performed by a courageous person; see Kosman, 'Being Properly Affected', p. 114

rescuing them, the virtuous and the continent person's rescuing the child (as well, perhaps, as the ethical novice's) was courageous, while, in the case of the person who does the right thing for the ethically wrong reasons, their rescuing the child, as well as their pulling the child from the road, was selfish, manipulative, grandstanding, or whatever.[49]

We may now consider rational explanation in cases in which one lacks knowledge. One could lack knowledge of three sorts: knowledge of reasons, means-end knowledge, or evaluative knowledge. If one lacks knowledge of reasons, one may still know that it would be good if p, but without knowing what would make it good if p – for example, one may learn this through testimony.[50] This means that one can be guided by this fact, having evaluative knowledge, and also means-end knowledge, without knowing the reasons there are for one to act. In such a case, if what I have been suggesting is on the right track, what one does would have some thick value, due to reasons of which one was ignorant. But the nature of one's action in respect of the partially constitutive conditions (PC)(i) and (PC)(ii) would be unaffected by lacking such knowledge. What we said in considering (PC*) would stand.

If one lacked means-end knowledge, one's belief would not guarantee that (PC)(i)(a) was met, although if it happened to be a true belief, it would be. And if one lacked evaluative knowledge, one's desire would not guarantee that (PC)(i)(b) was met, although if it happened to be a true evaluative belief, it would be. On the other hand, by failing to have means-end knowledge, or evaluative knowledge, there would be no guarantee, respectively, that (PC)(ii)(a) and (PC)(ii)(b) would not be met, and if one's means-end belief or evaluative belief were false, then, respectively, one or the other, or even both, *would* be met, and so one's action ought not to have been performed.

Nevertheless, the understanding of psychological explanations which fell short of knowledge-explanations would be parasitic upon those which did not, for they explain one's actions *as if* by one's knowledge, and purport to justify one's actions in the way that knowledge-explanations would, even if, as it turns out, they fail to do so. This is ensured by the fact that, in virtue of psychological explanation being in terms of

---

[49] This picture, I think, which puts to work the distinction between actions and things done, avoids the criticisms leveled by H. A. Prichard, 'Does Moral Philosophy Rest on a Mistake?', pp. 5-7. But it need not conflict with his main thesis in that paper, namely that it is illegitimate to demand an argument that *proves*, from considerations external to our ethical perspective, that we ought to do what it seems to us we ought to do, or even that we know we ought to do.

[50] This is plausibly an essential part of a proper ethical upbringing. Cf. Aristotle, *The Nicomachean Ethics*, 1095b1-15; and Burnyeat, 'Aristotle on Learning to Be Good'.

belief (and desire, which just is evaluative belief), they are neutral between explanation in terms of *mere belief* and explanation in terms of *knowledge*. Their explanatoriness, therefore, must purport to be that of knowledge-explanation, which is guaranteed to *successfully* justify one's actions.

5.7 Character and Second-Order Explanation of Action

In my discussion of rational explanation in §5.6, I drew on distinctions between agents with different sorts of characters: the courageous or fully virtuous person, the merely continent person, and the person with some defect of character who acts for the wrong sorts of reasons. In this section and the next, I will look more closely at what role traits of character play in the explanation of action. The account of the rational explanation of action proposed in §§5.4-5.6, I will suggest, leaves space for distinctive explanatory roles for an agent's character to play, and it allows us to provide satisfying accounts of what it is, for instance, to act virtuously, continently, *akratically*, or to suffer from *accidie*. I do not aim to provide an account of what traits of character *are*, but to say something about the explanatory roles that traits of character, on an intuitive understanding of them, paradigmatically play. In this section, I shall discuss the first such role, before considering a second in §5.8.

Character, I propose, should be seen as fitting into the account of agency I have been developing by playing a second-order explanatory role, where rational explanations play a first-order explanatory role. The first-order explanatory role played by an agent's psychological attitudes and reasons is the role of providing rational-explanatory answers to the question 'Why did A φ?'; the second-order explanatory role played by character is the role of providing explanatory answers to the question 'Why did A φ because…?', where this question is completed by mentioning the attitudes or reasons which figure in the corresponding first-order rational explanation.[51]

Moreover, there is an intimate connection between one's character traits and the considerations because of which one might act. The connection, roughly speaking, is that having certain traits of character explains one's propensities to be sensitive to certain broad kinds of consideration. For example, having a gluttonous character is to have a trait which explains one's propensity to be particularly sensitive to

---

[51] A similar view of the role of character in the explanation of action is given by Iskra Fileva, 'Two Senses of "Why": Traits and Reasons in the Explanation of Action', §3.2. See also Tenenbaum, *Appearances of the Good*, pp. 43-8; and Korsgaard, 'Kant's Analysis of Obligation', pp. 56-7.

considerations related to gastronomic gratification, while to be shy or to be attention-seeking are, in opposite ways, to be particularly sensitive to considerations related to being the focus of other people's attention.[52]

None of this is to say that one cannot act out of character, or that there will always be an explanation if one does so. It is not even to say that there will *always* be an available answer in terms of character to a request for this sort of second-order explanation, or that character is the only thing that may play this second-order explanatory role (it is plausible to think that certain emotional or even physical states of a person sometimes play such a role). But when we can explain an agent's actions in terms of their character, the resulting explanation is of this second-order kind; it explains why the agent did what they did for the reason that they did.[53]

Such explanation presupposes possible contrasts, which are ruled out by the first-order explanation of the form '*A* φ-ed because *p*', in which '*p*' is substitutable for a clause which cites *A*'s psychological attitudes or reasons. In fact, we can distinguish several different sorts of contrasting alternatives. First, there is the sort of contrast in which the explanatory clause which substitutes '*p*' is different. So this contrast for the explanation '*A* φ-ed because *p*' is the alternative possible (but actually false) explanation '*A* φ-ed because *q*', where '*q*' is substitutable for a clause which cites different attitudes or reasons from that substitutable for '*p*'.[54] A second-order explanation in terms of an agent's character would answer the question: 'why did *A* φ because *p*, rather than because *q*?'.

A second sort of contrast is one in which the verb or verb-phrase which substitutes for 'φ' is different. The contrast with the explanation '*A* φ-ed because *p*' is the alternative possible explanation '*A* ψ-ed because *p*'. There are two varieties of this second sort. (*a*) If *A*'s φ-ing and *A*'s ψ-ing were the same action, then both the original and the alternative explanation would actually be true. In (*a*), the same explanation is given for one and the same action, but the latter is characterised in a different way. (*b*) If

---

[52] Cf. Fileva, 'Two Senses of "Why"', p. 183

[53] Fileva claims that there can be exceptions to this, where a character explanation explains the first-order question of why an agent φ-ed, by indirectly telling us about the general *kinds* of reasons that were involved in a typical first-order explanation, although the latter is not actually provided (ibid. p. 190). But it is more plausible to interpret such cases as ones in which the character explanation still gives an answer to the second-order question, even though it was the first-order (and not the second-order) question that has been explicitly put, and, by doing so, indicates the general kind of first-order explanation we could expect. Fileva can still be correct that this will often satisfy our explanatory interests.

[54] Alternatively, '*q*' might offer a different, but non-competing true explanatory clause to '*p*', but that would just mean that there was another attitude or reason which was partially constitutive of one's action, and so is not relevant for the present discussion; see fn. 39 above.

A's φ-ing was not her ψ-ing, then what the contrast comes to is that *A* φ-ed because *p*, rather than ψ-ed because *p*, and the request for explanation comes to this: 'Why did *A* φ because *p*, rather than ψ because *p*?'. Plausibly, this is only possible at a suitably low level of resolution, as it were, so that what replaced '*p*' might equally explain *A*'s φ-ing or her ψ-ing; in other words, it could not amount to something like a Davidsonian primary reason.[55] Due to their limitations, I will not consider either varieties of this second sort of contrast any further.

Third, there is the sort of contrast yielded when *both* the explanatory clause substituting *p*, and the verb or verb-phrase substituting for φ, are different. The relevant request for explanation would be of the form, 'Why did *A* φ because *p*, rather than ψ because *q*?' Once again, we may distinguish sub-varieties of this third sort. (*c*) If the case is one in which *A*'s φ-ing and *A*'s ψ-ing were the same action, sub-variety (*c*) collapses into the first sort of contrast we considered. (*d*) If *A*'s φ-ing is not the same action as her ψ-ing, then the request for explanation is for why *A* performed one sort of action, explained in one way, rather than a different sort of action, explained in a different way. (*e*) The final variety can be seen as a limiting case of (*d*), and is one in which the contrast is between *A*'s having acted and *A*'s not having acted. So the request for explanation is this: 'Why did *A* φ because *p*, rather than not act at all, because *q*?'

Some examples will help to make these contrasts, and the role of character in their explanations, clearer. A version of Kant's shopkeeper example provides an illustration of the first sort. The shopkeeper, suppose, does not overcharge her inexperienced customer, because she wants to treat him fairly and believes that by charging the same price as for more experienced customers, she would bring it about that she treated him fairly. If we ask, 'Why did the shopkeeper trade at a fair price because she believed and desired these things, rather than because she wanted to maintain steady custom, and believed that by maintaining a reputation for fair trading practice she would bring it about that she maintained steady custom?' a perfectly intelligible answer would be that the shopkeeper is honest. What we have here is, more or less, an explanation in terms of an agent's character of why *A*'s action had moral worth.[56]

---

[55] See Davidson, 'Actions, Reasons, and Causes'

[56] Strictly, Kant would not consider this a case of action with moral worth, since it is not performed from duty: rather it is in conformity with duty, and 'however amiable it may be,' however 'honorable' and 'deserv[ing] praise and encouragement', it does not warrant 'esteem', as would actions with true moral worth (*Groundwork*, p. 11 (4:398), trans. by Mary Gregor).

An important class of cases falling under (*d*) are cases of *akrasia*, or weakness of will. So, suppose that a smoker trying to quit lights up a cigarette because she wants to satiate her unpleasant cravings, and believes that smoking it will do the trick. We can ask, 'Why did *A* light up a cigarette because she believed and desire these things, rather than walk away from the smoking area because she wanted to distract herself with some other engaging task, and believed such distraction would be found elsewhere?' Our answer may well be that she has an incontinent character. Note that this case could go the other way around, so that we could ask, 'Why did *A* walk away from the smoking area, rather than light up a cigarette?' Our answer might be that she is very determined, or that she has a continent character.

It is not entirely clear that there can actually be clear-cut cases falling under the limiting case (*e*), but a case that perhaps comes as close as possible to the limit might be one in which someone gets out of bed in the morning to go out in the sunshine, because they know it would be uplifting to do so, rather than remaining lying torpidly in bed because it would require nothing of them to do so. An intelligible answer to why they did the former rather than the latter would be that they are chirpy. Again, we should note that this sort of case could go the other way around, and this would plausibly yield an example of an important class of cases of the reverse of (*e*): cases of what is sometimes called *accidie*, or listlessness. In such cases, we are told, an agent believes that it would be good to do something, but, due to being in a generally morose state, utterly lacks any motivation to do it. So, in our example, our agent would believe that it would be good to get out into the sunshine, in virtue of the fact that it would be uplifting, but lack any motivation to do so, and as a result remain torpidly in bed. While in some such cases we might explain why someone remains in bed because it would be effortless in terms of their being indolent, in such extreme cases of *accidie* it seems more natural to describe them as suffering from a deep lethargy of spirit, or a profound depression.

Such descriptions do not identify clear-cut traits of character, but possibly emotional conditions, or psychological pathologies. As I suggested earlier, though, it is plausible to think that certain emotional conditions or even physical states can sometimes play the same second-order explanatory role as traits of character, so this

needn't be thought problematic. In fact, it is tempting to just adopt a crude classification, and treat such emotional or physical states as quasi-traits of character.[57]

The phenomena of *akrasia* and *accidie*, which fall under (*d*) and (*e*) respectively, are often thought to present particular potent problems for cognitivist and motivationally internalist views of evaluative judgement, and to be even more problematic for views which accept some form of the Guise of the Good thesis. The account I have been developing accepts a distinctive version of each of these, but in the next section I aim to show how we can still acknowledge that these are genuine phenomena, and accommodate them within the overall account of agency I have been developing.

5.8 Virtue and Continence; *Akrasia* and *Accidie*

The first role I have been suggesting that an agent's character plays in the explanation of action is the second-order role of explaining why an agent did what she did because of the psychological attitudes or reasons that she acted on. Plausibly, it is because an agent's character plays this second-order explanatory role that it also plays a further explanatory role. By its involvement in explaining why an agent did what she did because of her attitudes or reasons, her traits of character explain the *manner* in which she acted. We might call this an adverbial role of character in the explanation of action. For example, a kind person, ordinarily, would help someone lift their heavy bag off the train *kindly* (rather than, say, grudgingly). Of course, an agent may be kind and yet sometimes do things unkindly, and an agent who is not generally kind might on occasion do something kindly nevertheless. We might plausibly suggest that even for such a person, they have, as it were, some kindness in them, but that it is quite undeveloped. This sort of thing, in fact, seems very common: people possess many traits of character only in a minimal and undeveloped form.

In §5.6, we saw that someone could do a courageous thing, but that their *action* would be courageous if and only if they did what they did for the ethically right reasons, such as *that the child is in immediate danger*. Even this, however, leaves open whether they would have performed their courageous action *courageously*, that is, in the manner that a courageous person would do it. To do the courageous thing courageously would be to do it out of a courageous character. If we were to accept some version of the thesis

---

[57] Other examples of the sorts of emotional and physical states of a person that I have in mind might be being grouchy, being elated, having low blood-sugar, or feeling nauseous.

of the unity of the virtues, this would mean that to do something courageously would be to do it out of a virtuous character. Similarly, if one lacks virtue, but one acts continently, having a continent character will explain this manner in which one acts, and likewise if one acts incontinently, performing *akratic* actions. Considering some common features of discussions of character in the Aristotelian tradition may help us to understand what is distinctive about acting virtuously, continently, or incontinently, and how an agent's character is involved in explaining their actions being performed in such a distinctive manner.

To begin, consider L. A. Kosman's comment:

> [A]cts are in a sense virtuous when they are acts of the kind which a virtuous person would perform, but are fully virtuous only when performed by a virtuous person (*one who performs such acts in the way that a virtuous person would perform them – that is, out of the fixed character that is virtue*).[58]

Kosman's use of the term 'acts' here fails to distinguish between things done and actions. Presumably, what he means to claim is that what a virtuous person would do is what would be, in the first, weak sense he mentions, virtuous, and it would, I take it, be virtuous because it would be, say, just, and because, here and now, doing the just thing is what virtue demands. But, of course, if a continent person's doing the thing that the virtuous person would do is to be virtuous even in that first, weak sense, she would have to do it for the right sorts of reasons, and so perform a just action. But even if the continent person did the thing a virtuous person would do, for the right reasons, i.e. those for which the virtuous person would do it, they would still fall short of acting 'fully' virtuously, just because, *ex hypothesi*, their acting as they do for the reason that they do would not be explained by their having a virtuous character.

We can make some sense of what this difference between acting in a virtuous manner (i.e. virtuously) and acting in a merely continent manner (i.e. merely continently) is supposed to be by way of McDowell's interpretation of Aristotle. According to McDowell, both the virtuous and the merely continent person are able to identify, in light of their correct conceptions of *eudaimonia* and the characteristic activity (*ergon*) of a human being (which McDowell glosses as 'what it is the business of [a human being] to do'),[59] what, in the situation they find themselves in, virtue

---

[58] Kosman, 'Being Properly Affected: Virtues and Feelings in Aristotle's Ethics', p. 114, emphasis added.
[59] McDowell, 'The Role of *Eudaimonia* in Aristotle's Ethics', p. 12; cf. Aristotle, *The Nicomachean Ethics*, 1097b20-1098a20

requires of them. Both of them, moreover, would do the thing that virtue requires, and they would both do it for the right sorts of reasons.

The crucial difference between the *way* that the virtuous person does what she does and the way that the merely continent person does what she does for the same sorts of reasons, on McDowell's account, is this. The merely continent person has, or is prone to have, conflicting motivations, due to her sensitivity to the reasons present in her situation which favour doing something else, but, since the sway induced in her motivational propensities by the reasons there are to do what virtue requires is greater than that of those other considerations, she acts as the virtuous person would. For the virtuous person, in contrast, because she has a special sort of conception of her situation, all competing reasons to do something other than that which virtue requires are 'silenced'.[60] Even though she appreciates that, ordinarily, and were it not for the requirement of virtue the situation presents her with, those reasons would favour doing some other thing, she sees such alternative courses of action as conflicting with her conception of 'what a human being, circumstanced as [s]he is, should do',[61] and consequently though 'completely aware of the attractiveness of the competing course […s]he is not attracted by it.'[62]

McDowell himself, following Aristotle, thinks of possession of the special conception of her situation that the virtuous person has as a matter of having, literally, a special *perception* of her situation, which involves a sensitivity to what is really of significance in it. As a cognitive capacity, this is supposed to be suitable to allow that virtue is a matter of having a special sort of ethical knowledge. At the same time, it is supposed to accommodate the view that a conception of how to live well is not codifiable, i.e. it is not possible for it to be formulated in a set of universal rules which, applied to concrete situations, would yield correct answers as to what the virtuous thing to do would be in every possible case.[63] But McDowell's perceptual model does not seem to be necessary. The requirements that the acquisition of virtue is the acquisition of a kind of knowledge, and that the relevant knowledge resists codification, could also be accommodated on a model which took the special knowledge that amounts to

---

[60] McDowell, 'The Role of *Eudaimonia* in Aristotle's Ethics', p. 18
[61] Ibid. p. 10
[62] See John McDowell, 'The Role of *Eudaimonia* in Aristotle's Ethics', pp. 16-9; 'Some Issues In Aristotle's Moral Psychology', pp. 46-9; 'Virtue and Reason', pp. 55-6; and 'Incontinence and Practical Wisdom in Aristotle', esp. pp. 66-9 (the quotation is to be found on p. 68).
[63] See McDowell's papers cited in the previous note, and cf. Aristotle, *The Nicomachean Ethics*, 1142a23-30.

possession of virtue to be a matter of complete grasp of all evaluative concepts. Such grasp of these concepts should not be thought of as codifiable, for example through conceptual analyses of the evaluative concepts, and would be compatible with the knowledge resulting from complete grasp of those concepts carrying the appropriate motivational consequences that McDowell, following Aristotle, insists upon.[64] For present purposes, however, we need not decide how the virtuous person's conception of her situation should be understood.

We are now in a position to understand the adverbial role of virtue and of continence in the conduct of the virtuous and the merely continent agents. The virtuous agent has a character which is such as to explain her propensities to be sensitive to the special sorts of reasons related to the requirements of virtue.[65] But, in addition, as Kosman puts it, 'a given virtue is a disposition with respect to a characteristic set of actions *and* feelings [*pathos*]'.[66] The acquisition of a virtuous character in general, therefore, is a process in which one's sensibilities in general are tutored, with the result that one becomes sensitive to certain sorts of reasons for action so that they engage one's motivational concerns, and at the same time one becomes resistant to having one's motivational concerns engaged when that would be inappropriate, from the perspective of virtue.[67] Because of this, the virtuous agent lacks any temptation to do something other than what virtue requires. The resulting smoothness in performing virtuous actions, without any tendency to be conflicted or reluctant in acting well, is what the distinctive manner of acting virtuously amounts to.

Where the virtuous person acts smoothly, with, as it were, no errant desires, the merely continent person has such errant desires, although she is successful in resisting their sway, for, since she has not yet fully acquired a virtuous character, her motivational propensities have not yet been fully tutored. Consequently, she is prone to conflictedness or reluctance; she needs to struggle with herself, exercising determination and strength of will, to ensure that she does the virtuous thing, perhaps even for the virtuous reasons she can appreciate. Incontinence can be treated along very similar lines. The person of incontinent character is prone to *akratic* action. In cases of

---

[64] See Sarah Sawyer, 'Minds and Morals'. To properly accommodate the smoothness of the virtuous person's actions described below, though, a view like Sawyer's would need to be complicated along the lines I suggest below.

[65] McDowell, 'The Role of *Eudaimonia* in Aristotle's Ethics', p. 10

[66] Kosman, 'Being Properly Affected', p. 109.

[67] Cf. McDowell, 'Some Issues in Aristotle's Moral Psychology', pp. 31-2, 40, 46-9; Aristotle, *The Nicomachean Ethics*, 1119b13-18; Kosman, 'Being Properly Affected', pp. 108-9; J. O. Urmson, 'Aristotle's Doctrine of the Mean', p. 161

akrasia, a person judges that it would be better to φ than to ψ, but despite this they ψ rather than φ. Like the virtuous person, the incontinent person appreciates what acting well would be in her situation, and judges that it would be better to act well than to not. But like the continent person, she has errant desires to do things other than what would constitute acting well. Consequently, the incontinent person is prone to the same internal conflictedness or reluctance as the continent person, but, unlike the continent person, is unsuccessful in her struggle, and does the thing she judges to be less good.

The characteristic manner in which she acts badly is a matter of the presence of her distinctive internal conflict, but, unlike for the continent person, the result of the conflict is that she fails to do what she judges is the better thing. But her doing what she does can readily be explained by the fact that she desires to do something other than what she thinks is best. This is something that must be acknowledged by any account of *akrasia.* The incontinent person's errant desires, together with other relevant beliefs, serve to provide a rational psychological explanation of what she does, even though, as is generally agreed, when someone acts *akratically,* they act irrationally.

That *akrasia* necessarily involves having contradictory desires, however, is controversial, for many accounts of *akrasia* take it that it need only involve a failure of the agent's evaluative beliefs and their desires to line up, without this necessarily involving their having contradictory, inconsistent, or incompatible desires. What they say is essential to the phenomena is that while the agent judges that it would be better to φ than to ψ, and to that extent judges that it would be good to φ, this does not imply that they desire to φ at all, while they do desire, as it happens, and as is shown by their action, to ψ.[68] Because I have argued that desires just are evaluative beliefs, however, I am committed to rejecting such accounts of *akrasia.* The evaluative belief that the incontinent (and the continent) person has that it would be good to do what they judge would be better, implies that they have a contradictory desire to that which they act on, or, in the continent person's case, successfully overcome. I take this to be an advantage of this account, since we can thereby explain the appearance and the phenomenology of such phenomena as involving a genuine internal *contradiction* in the agent. It is not clear that the former sort of account is capable of explaining the genuine irrationality involved in *akrasia,* because it does not allow for any genuine contradiction.

---

[68] Smith, *The Moral Problem*, pp. 119-23; Stocker, 'Desiring the Bad', e.g. pp. 745, 748-9

As we shall see, however, the account I have been suggesting does leave us with a serious puzzle to address. But to see the puzzle we first need to ask how we are to make sense of someone's believing that it would be best if they φ-ed, and yet desiring to ψ (which, of course, would be believing that it would be good if they ψ-ed). This raises the further question of just how judgements of comparative value – that φ-ing would be better than ψ-ing, or that φ-ing would be best – are to be understood. One simple way of understanding such comparative evaluative judgements takes a quantitative conception of value, on which all values are commensurable and measurable by a single standard. On such a view, goodness and badness are taken as relative terms of comparison on a single scale. If one thing measures higher on this scale than another, then the former will be good in relation to the latter, and the latter will be bad in relation to the former. But on this quantitative view of value, if someone believes that it would be better if she φ-ed than if she ψ-ed, then *ipso facto* she believes that it would be good if she φ-ed and bad if she ψ-ed. So how could one believe that it would be better if she φ-ed than if she ψ-ed, and yet still believe it would be good if she ψ-ed, as would have to be the case if she still desired to ψ, given my argument that desire just is evaluative belief? This, in effect, is a version of the argument in Plato's *Protagoras* against the possibility of *akrasia*.[69] But the reply is straightforward: given the quantitative conception of value, someone can believe that it would be better if she φ-ed than if she ψ-ed, and still believe that it would be good if she ψ-ed, if, in addition, she believes that it would be better if she ψ-ed than if she φ-ed. In other words, the additional premise that the argument from the quantitative conception of value requires in order to threaten the possibility of *akrasia* is one which claims that it is impossible to have contradictory beliefs, but this is a premise which we have no reason to accept.[70]

The trouble with this response is that in order to acknowledge the possibility of *akrasia* in the context of the quantitative conception of value, it commits itself to the claim that it is impossible for someone's belief that it would be better if she φ-ed than if she ψ-ed to be *true*, and, at the same time, for her belief that there is something in virtue of which it would be good if she ψ-ed to *also* be true. The response is unable to acknowledge that there could be genuine attractions in doing what one correctly takes to be the less good course of action. To make such an acknowledgement requires that we

---

[69] See Plato, *Protagoras*, 354c-358d.

[70] See Michael Morris, 'Akrasia in the *Protagoras* and the *Republic*', for a treatment of, and response to, the *Protagoras* argument along these lines.

reject the quantitative conception of value, and with it what Wiggins calls '*the principle of compensation in kind*', according to which 'if course *x* is better in respect of eudaimonia than course *y*, then there is no desirable feature that *y* offers that *x* does not offer too, by way of an equal or greater degree of that very feature.'[71] If we do reject these, and embrace a value pluralism, we can acknowledge that 'for any *n*-tuple of courses of action actually available at time *t* to an agent *x* there is some way or other of establishing which of the *n*-tuple is the better course of action',[72] and so secure judgements of comparative value. But, at the same time, we can acknowledge that there may be some value in another course of action in virtue of which it would be good to pursue, which would not be compensated for in kind by the pursuit of the better course of action.[73] Therefore it could be true both that it would be better if one φ-ed, and that it would be good if one ψ-ed; the former would do nothing to impugn the latter.

In the context of the quantitative conception of value, we can easily see what the source of the irrationality involved in *akrasia* would be – one would act irrationally in that one's action would spring from obviously contradictory evaluative beliefs. But, as we just saw, at least one of these beliefs would have to be false, and it would therefore become easy to say what it would be for the continent or the incontinent person to have inappropriate desires. To have an inappropriate desire would be to falsely believe that it would be good if one φ-ed. The virtuous agent would not have such inappropriate desires because, having tutored sensibilities, he only 'desires the things he ought, as he ought, and when he ought'.[74] But to properly accommodate the fact that in cases of *akrasia* an agent can be *right* about the less good course of action still being good, so that it can be true that, as Wiggins puts it, 'the weak-willed man acts not for no reason at all […] but irrationally',[75] we need to explain in some other way the inappropriateness of the incontinent (and equally the continent) person's errant desires, which the virtuous person lacks.

This is where the serious puzzle arises. It seems that there are two possible explanations, and I shall suggest that the second is correct. In order to see the first possible explanation, recall the formulation of the norm of evaluative belief, which

---

[71] Wiggins, 'Weakness of Will, Commensurability, and the Objects of Deliberation and Desire', pp. 256-7

[72] Ibid. p. 256

[73] See ibid. pp. 256-60

[74] Aristotle, *Nicomachean Ethics*, 1119b13-18, as translated in Burnyeat, 'Aristotle on Learning to Be Good', p. 80

[75] Wiggins, 'Weakness of Will, Commensurability, and the Objects of Deliberation and Desire', p. 241

followed from the formulation of the norm of belief in general (NB) that I defended in Chapter 3:

> (NEB*)
>
>> For any A and any p: (i) A ought to (believe that it would be good if p) only if it would be good if p;
>> *and,*
>> (ii) if it would not be good if p, A ought to (not believe that it would be good if p).

(NEB*)(ii) gives a sufficient, but not necessary, condition for its being the case that one ought not to believe that it would be good if p. There is nothing in its formulation which prohibits us from claiming that, if it would be better if one φ-ed than if one ψ-ed, this is another sufficient, though not necessary, condition for its being the case that one ought not to believe that it would be good if one ψ-ed, and *eo ipso* that one ought not to desire to ψ. But this explanation, though technically compatible with the norm of evaluative belief, is not really plausible. First, it intuitively seems wrong that if it really would be good if one ψ-ed, one ought not to believe this. Second, if one did not believe that it would be good if one ψ-ed, one would not be, to use McDowell's phrase again, 'completely aware of the attractiveness of the competing course [of action]',[76] yet on the Aristotelian view, the virtuous person *would* have this awareness, and yet would 'not [be] attracted by it.'[77] So this explanation is not compatible with the Aristotelian view of what would be involved in having a virtuous character.

This seems to make the puzzle even thornier. Having rejected the first possible explanation, we seem to be required to say that possession of a virtuous character involves believing that it would be good if one ψ-ed, even though it would be less good than if one φ-ed (which is what virtue requires), but also requires not desiring to ψ, since one's sensibilities have been tutored so that one does not desire inappropriate things. But how could it be possible that believing that it would be good if one ψ-ed could be appropriate, and even necessary for having a virtuous character, even when desiring to ψ would be inappropriate, and incompatible with having a virtuous character, if desiring to ψ just is believing that it would be good if one ψ-ed?

---

[76] McDowell, 'Incontinence and Practical Wisdom in Aristotle', p. 68
[77] Ibid.

Despite the intractable appearance of this problem, it can be dissolved, and we have already amassed the materials required for its dissolution in earlier chapters. The key is that what would be inappropriate about having a desire to ψ, when ψ-ing would be less good than φ-ing, is its having any tendency to exert, as it were, a motivational sway, or for it to have a tendency to explain one's actions. In the case of both the continent and the incontinent person, their desiring to ψ exerts a motivational sway, which accounts for their characteristic internal struggle, and for the latter, it in fact rationally explains their resultant *akratic* action. Of course, we must also realise that, since they are each at different points on the path towards virtue, and each falls short of that ideal, the continent person is liable, occasionally, to be unsuccessful in her struggle to act as she realises would be best, and instead to act incontinently, while the incontinent person is liable to sometimes be successful, and to act for the best. So the *sort* of inappropriate desire that each has does have the *tendency* to explain their actions, even though particular instances of them may or may not in fact do so.

We could put these points more concisely by saying that what is inappropriate about having a desire to ψ, when it would be less good if one ψ-ed than if one φ-ed, is for the desire to ψ to have the relation to one's agency which we considered in §2.4. It is precisely such a relation to one's agency which is properly severed for the virtuous person, who fully appreciates that it would be good if they ψ-ed, but is utterly unmotivated to ψ because they realise that it would be better if they φ-ed. But, in the course of discussing the significance of the grammar of desire specifications and attributions taking an infinitival phrase to specify their content, we found that the point of this grammar is to emphasise just this relation to one's agency.

In light of this, we can see why it would be natural to describe one's belief that it would be good if one ψ-ed as appropriate, because it would be a true evaluative belief, but also natural to describe one's desire to ψ as inappropriate, because it serves to emphasise a relation between one's psychological attitude and one's agency, where the relation in question amounts to its exerting a motivational sway. This can be so, even though, strictly speaking, the desire to ψ just is the belief that it would be good if one ψ-ed. So, strictly speaking, even the virtuous person would desire to ψ, but it would be misleading to describe them so because it would suggest that they have an attitude which exerts a motivational sway contrary to their correct judgement about what would be best to do, and it is just such a motivational sway contrary to their correct judgement which is properly severed for the person of fully virtuous character.

How can one desire to ψ, though, without this carrying any motivational sway? Is it not of the very essence of desires that they are motivating attitudes? There is something right in this protestation, but as we shall see it does not undermine the claim that one can, strictly, desire to ψ without this carrying any motivational sway, and to think that it does is to relapse into an Humean conception of desire, and a causal conception of psychological explanation.

What it would amount to for it to be of the very essence of desires that they are motivating attitudes can only be understood in terms of the motivating role of desire in the explanation of action. On Hume's conception of the role of desires in the explanation of action, desires supply a fundamentally arbitrary causal-motivational *force*, which is guided by the agent's beliefs, or, on an Humean view like Smith's, supply a causally efficacious goal-directedness which is similarly arbitrary, because the goal in question is provided by the content of a desire which conforms to (FAD). The corresponding conception of psychological explanation takes it to have a *teleologico*-causal explanatory character. From that perspective, it seems unintelligible that an agent could desire to ψ without this carrying any such *teleologico*-causal motivating *force*. But I have argued that both these positions ought to be rejected, and have defended alternatives in their place.

We see the significance of the wholesale rejection of the Humean conception of desire and the causal conception of the character of psychological explanation when we realise that we cannot, having rejected them, accept any literal reading of the apparently causal metaphors with which we are prone to describe the role of desires. We often speak, for example, of desires having motivating *force*, of desires *pushing* or *pulling* someone in one direction or another, or even, as I have been speaking, of their *exerting* motivational *sway*. But these are nothing but metaphors, and not to be taken in any literal causal sense.

Here again is the formulation of the relation of partial constitution between an agent's beliefs, desires and actions which is the core of my proposed non-causal account of the character of psychological explanation:

    (PC)   For any *A*, φ, and *p*: For *A* to have φ-ed because she believed that if she φ-ed she would bring it about that *p,* and she desired that *p,* is for *A*'s φ-ing, in its essential nature, to be such that:

        (i) it ought to have been performed only if it was true that (a) *A*'s φ-ing would have brought it about that *p*, and that (b) it would have been good if *p*; *and,*

        (ii) if it was not true that (a) *A*'s φ-ing would have brought it about that *p,* or that (b) it would not have been good if *p,* then it ought not to have been performed.

The role that desire has in the explanation of action, I have argued, is a constitutive role of supplying the perceived good for the sake of which someone's action is undertaken, or towards which it is directed. This can be seen in the contribution it makes, via its constitutive norm (ND), to (PC)(i)(b) and (PC)(ii)(b). This provides a non-Humean construal of the way in which rationalising psychological explanation is teleological explanation; it is explanation of something directed towards what one takes to be the good, where the good is the end of action, and action is that which is, in Aristotle's phrase, 'towards [*pros*] the end'.[78] Desire's contribution to (PC), which provides the perceived good for the sake of which one's action is undertaken, *fully exhausts its motivating role.*

The belief that it would be good if one ψ-ed supplies the perceived good, which, *if* the belief explains an action, the action would be performed for the sake of. So the evaluative belief that it would be good if one ψ-ed *essentially* has the capacity to serve its motivating role in the explanation of action, even if in fact it happens to play no such explanatory and motivating role at all. Since the desire to ψ just is this evaluative belief, this means that we *can* make sense of it being of the essence of desires that they are motivating attitudes, in the sense that they essentially supply the perceived good which an action would be performed for the sake of *if* they explain an action at all, even in the absence of actually having such a motivating influence. That would be how it was in the ideal case of the virtuous agent who would be utterly unmotivated by their inappropriate desires to do something other than what they judge to be best.

---

[78] Aristotle, *The Nicomachean Ethics*, 1111b25-30; cf. McDowell's comments on the translation of Aristotle's Greek in 'Some Issues in Aristotle's Moral Psychology', pp. 25-6

With this explanation in hand as to how it can be possible, metaphorically speaking, for desires to be insulated from exerting any motivational sway, we can easily account for the phenomenon of *accidie*. The person suffering from *accidie* believes that it would be good, or even best, if they φ-ed, but, due to depression, despair, mental exhaustion or whatever, they are completely uninfluenced by their evaluative belief, and so lack any motivation to φ. We can sympathise with certain Humean accounts of *accidie*, which claim that the person suffering it believes that it would be best if they φ-ed, although they lack any desire to φ.[79] Claiming that they desire to φ would be misleading, because it would emphasise a relation to the person's agency which, in these cases of *accidie*, is precisely what is lacking. Strictly, though, the claim that the person suffering from *accidie* desires to φ would be true, since the desire to φ is the evaluative belief that it would be good if they φ-ed, and *ex hypothesi* they have such an evaluative belief. What is missing from these cases is the tendency for their desires to be explanatory of their actions, something which is registered by speaking of them as having the evaluative beliefs, but not of having the desires with which those beliefs are identical.

That the desires of the person suffering from *accidie* lack this tendency to be explanatory of their actions is adequately explained by the fact that they are depressed, or despairing, or mentally exhausted and so on, because these are just the sorts of states of character or emotional conditions which serve the second-order explanatory role discussed in §5.7. These states or conditions explain why, for instance, they remain lying torpidly in bed because they believe that doing so would require nothing of them, and believe it would be good if they did something which required nothing of them. Their torpor explains their propensity to be sensitive to considerations relating to effortlessness and inactivity, and insensitive to all other evaluative considerations such as that going out in the sunshine would be uplifting.

Through this discussion, we can see that the cognitive account of desire as evaluative belief (§3.9), together with the non-causal account of the rational explanation of action (§5.6), have the resources to do justice to the phenomena of *akrasia* and *accidie*. It should also have clarified the conception of motivation as such that is in play in the non-causal account of rational explanation. Having rejected the view that

---

[79] See e.g. Smith, *The Moral Problem*, pp. 120-1; Stocker, 'Desiring the Bad', esp. p. 744

psychological explanation has a causal-explanatory character, we cannot still try to cling to a corresponding causal conception of motivation.

5.9 Conclusion

Chapter 4 defended the view that actions are a categorial species of event, which are distinctive in being the occurrences of an *agent's* doing something intentionally, by which the agent causes the consequences of her actions. It thus rejected theories of action according to which the relation between an agent's actions and her attitudes which explain them is a causal relation. This chapter has defended an alternative, constitutive conception of psychological explanation, according to which the character of the explanation provided is justificatory, setting constraints by determining the necessary conditions for its being the case that the action ought to have been performed, and sufficient conditions for its being the case that the action ought not to have been performed. This alternative conception of the character of psychological explanation, moreover, answers Davidson's challenge, thus subverting the primary motivation for a causal conception.

The conception of the nature of actions that this yields is one according to which actions are particulars which are essentially apt for ethical evaluation. In addition to the justificatory constraints set by the partially constitutive psychological attitudes which explain them, actions admit of rich ethical evaluation as, for instance, being, courageous, attention-seeking, or, more commonly, simply mundanely useful. Such rich evaluative status is constituted by an agent acting for the reasons that they do, and such acting for reasons is made possible by the agent's having knowledge of the reasons that favour performing the relevant sort of action, i.e. the *thing they do* in acting.

Actions, as we have seen, are an *agent's* doing something, and they are rationally explained by *the agent's* psychological attitudes or reasons. By rationally explaining an agent's actions, we learn something about its nature, but, since it is essentially *hers*, and explained by attitudes which are essentially *hers*, we also learn something about the agent herself.[80] We learn that she was sensitive to certain kinds of considerations, that she did, or did not, struggle over whether to be swayed by those or other kinds of considerations, we learn what good she saw in doing what she did, and so on. We thereby learn the sort of person she is: we learn something about her character.

---

[80] Cf. Hornsby, 'Agency and Actions', p. 8

In Chapter 1, we began our enquiry into moral psychology and reasons for action with Hume's claim that 'Reason is, and ought only to be the slave of the passions'.[81] We can now see that this is exactly wrong. Desires, which Hume thought of as passions, themselves belong to reason, for they just are evaluative beliefs. Desires are not, as Hume thought, fundamentally arbitrary non-representational causal forces which, guided by our beliefs, propel our bodies to move in various ways, but are rather essentially rationally evaluable attitudes which are partially constitutive of *our actions*, including when what we do is *move our bodies*. But desires being rationally evaluable is not, as Kant thought, a matter of their being evaluable in terms of the structure of Reason itself, but rather of their being constrained by the 'evaluative contours of our world'.[82] The evaluative facts themselves, which are independent of our desires, are what constitute our reasons to act.

In *The Moral Problem*, Smith asks: 'How much of Hume's theory should we keep, how much should we reject?'[83] If the argument of this thesis is on the right track, the answer is that we should maintain Hume's insistence that belief and desire are central to the psychological explanation of our actions, but reject just about everything else.

---

[81] Hume, *Treatise*, III.i.1, p. 415
[82] McDowell, 'Might There Be External Reasons?', p. 109
[83] Smith, *The Moral Problem*, p. 91

Bibliography

Ackrill, J. L., 'Aristotle on Action' in *Essays on Aristotle's Ethics*, ed. by Amélie Oksenberg Rorty, (Los Angeles: University of California Press, 1980), pp. 93-101

— 'Aristotle on *Eudaimonia*', in *Essays on Aristotle's Ethics*, ed. by Amélie Oksenberg Rorty, (Los Angeles: University of California Press, 1980), pp. 15-33

Allison, Henry E., *Kant's Theory of Freedom*, (New York: CUP, 1995)

— *Kant's Groundwork for the Metaphysics of Morals: A Commentary*, (Oxford: OUP, 2011)

Altham, J. E. J., 'The Legacy of Emotivism', in *Fact, Science, and Morality*, ed. by G. MacDonald and C. Wright, (Oxford: Blackwell, 1986), pp. 275-288

Alvarez, Maria, *Kinds of Reasons*, (Oxford: OUP, 2010)

Alvarez, Maria and John Hyman, 'Agents and their Actions', *Journal of Philosophy*, 73 (1998), 219-245

Anscombe, G. E. M., *Intention*, (Oxford: Basil Blackwell, 1963)

— 'Under a Description', in G. E. M. Anscombe, *Metaphysics and the Philosophy of Mind*, (Oxford: Basil Blackwell, 1981), pp. 208-219

— 'Practical Inference', Essay 10 in *Human Life, Action, and Ethics: Essays by G. E. M. Anscombe*, ed. by Mary Geach and Luke Gormally, (Exeter: Imprint Academic, 2005)

Aristotle, *The Nicomachean Ethics*, trans. by David Ross and Lesley Brown, (New York: OUP, 2009)

Ayer, A. J., *Language, Truth and Logic*, (London: Victor Gollancz, 1946)

Ben-Yami, Hanoch, 'Against Characterising Mental States As Propositional Attitudes', *The Philosophical Quarterly*, 47 (1997), 84-89

Blackburn, Simon, *Ruling Passions*, (Oxford: Clarendon, 2000)

Brandom, Robert, 'Action, Norms, and Practical Reasoning', in *Varieties of Practical Reasoning*, ed. by Elijah Millgram (Cambridge MA: MIT Press, 2001), pp. 465-480

Bricke, John, *Mind and Morality: An Examination of Hume's Moral Psychology*, (Oxford: Clarendon, 1996)

Broome, John, 'Reasons and Motivation', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 71 (1997), 131-146

— 'Normative Requirements', *Ratio*, 12 (1999), 398-419

Burnyeat, M. F., 'Aristotle on Learning to Be Good', in *Essays on Aristotle's Ethics*, ed. by Amélie Oksenberg Rorty, (Los Angeles: University of California Press, 1980), pp. 69-92

— 'Idealism and Greek Philosophy: What Descartes Saw and Berkeley Missed', *The Philosophical Review*, 91 (1982), 3-40

Bykvist, Krister, and Anandi Hattiangadi, 'Does Thought Imply Ought?', *Analysis,* 67 (2007), 277-285

Coleman, Mary Clayton, 'Directions of Fit and the Humean Theory of Motivation', *Australasian Journal of Philosophy*, 86 (2008), 127-139

Dancy, Jonathan, *Moral Reasons*, (Oxford: Blackwell, 1993)

— 'In Defense of Thick Concepts', *Midwest Studies in Philosophy*, 20 (1995), 263-279

— 'Why There Is Really No Such Thing as the Theory of Motivation', *Proceedings of the Aristotelian Society*, 95 (1995), 1-18

— *Practical Reality*, (Oxford: OUP, 2000)

— *Ethics Without Principles*, (Oxford: Clarendon, 2004)

— 'Practical Concepts', in *Thick Concepts*, ed. by Simon Kirchin, (Oxford: OUP, 2013), pp. 44-59

— 'On Knowing One's Reason', in *Epistemic Norms: New Essays on Action, Belief, and Assertion*, ed. by Clayton Littlejohn and John Turri, (Oxford: OUP, 2014), pp. 81-96

Davidson, Donald, *Essays on Actions and Events*, (Oxford: Clarendon Press, 2001)

— 'Actions, Reasons and Causes', in Donald Davidson, *Essays on Actions and Events*, (Oxford: Clarendon Press, 2001), pp. 3-19

— 'Agency', in Donald Davidson, *Essays on Actions and Events*, (Oxford: Clarendon Press, 2001), pp. 43-61

— 'The Individuation of Events', in Donald Davidson, *Essays on Actions and Events*, (Oxford: Clarendon Press, 2001), pp. 163-180

— 'The Logical Form of Action Sentences', in Donald Davidson, *Essays on Actions and Events*, (Oxford: Clarendon Press, 2001), pp. 105-148

— 'Mental Events', in Donald Davidson, *Essays on Actions and Events*, (Oxford: Clarendon Press, 2001), pp. 207-227

— 'Reply to Quine on Events', in Donald Davidson, *Essays on Actions and Events*, (Oxford: Clarendon Press, 2001), pp. 305-311

— 'Problems in the Explanation of Action', in Donald Davidson, *Problems of Rationality*, (Oxford: Clarendon, 2004), pp. 101-116

Dent, N. J. H., *The Moral Psychology of the Virtues*, (Cambridge: CUP, 1984)

Engel, Pascal, 'Is truth a norm?', in *Interpreting Davidson,* ed. by P. Kotatko, P. Pagin, and G. Segal, (Stanford: CSLI Publications, 2001), pp. 35-51

— 'In Defense of Normativism About the Aim of Belief', in *The Aim of Belief*, ed. by Timothy Chan, (Oxford: OUP, 2013), pp. 32-63

Enoch, David, 'On Mark Schroeder's Hypotheticalism: A Critical Notice of *Slaves of the Passions*', *Philosophical Review*, 120 (2011), 423-446

Fileva, Iskra, 'Two Senses of "Why": Traits and Reasons in the Explanation of Action', in *Questions of Character*, ed. by Iskra Fileva, (Oxford: OUP, 2016), pp. 182-202

Finlay, Stephen, 'The Obscurity of Internal Reasons', *Philosophers Imprint*, 9:7 (2009), 1-22

Foot, Philippa, 'Morality as a System of Hypothetical Imperatives', *The Philosophical Review*, 81 (1972), 305-316

Ford, Anton, 'Action and Generality', in *Essays on Anscombe's* Intention, ed. by Anton Ford, Jennifer Hornsby and Frederick Stoutland, (Cambridge, MA: Harvard University Press, 2011), pp. 76-104

Frankfurt, Harry G., 'Identification and wholeheartedness', in Harry G. Frankfurt, *The Importance of What We Care About*, (New York: CUP, 1998), pp. 159-176

Gibbard, Alan, *Wise Choices, Apt Feelings,* (Cambridge, MA: Harvard University Press, 1990)

— *Thinking How to Live,* (Cambridge, MA: Harvard University Press, 2003)

Glüer, Kathrin, and Åsa Wikforss, 'The Truth Norm and Guidance: a Reply to Steglich-Petersen', *Mind*, 119 (2010), 757-761

— 'Still No Guidance: Reply to Steglich-Petersen', *Theoria*, 81 (2015), 272-279

Greenberg, Alexander, and Christopher Cowie, 'Is the Norm on Belief Evaluative? A Response to McHugh', *Pacific Philosophical Quarterly*, 97 (2016), 1-18

Harcourt, Edward, and Alan Thomas, 'Thick Concepts, Analysis, and Reductionism', in *Thick Concepts*, ed. by Simon Kirchin, (Oxford: OUP, 2013), pp. 20-43

Hare, R. M., *The Language of Morals*, (Oxford: Clarendon, 1961)

Hornsby, Jennifer, *Actions*, (London: Routledge, 1980)

— *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind,* (Cambridge, MA: Harvard University Press, 1997)

— 'Agency and Causal Explanation' in Jennifer Hornsby, *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind,* (Cambridge, MA: Harvard University Press, 1997), pp. 129-153

— 'Physicalism, Events, and Part-Whole Relations', in Jennifer Hornsby, *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind,* (Cambridge, MA: Harvard University Press, 1997), pp. 46-62

—'Which Physical Events are Mental Events?', in Jennifer Hornsby, *Simple Mindedness: In Defense of Naïve Naturalism in the Philosophy of Mind,* (Cambridge, MA: Harvard University Press, 1997), pp. 63-77

— 'Anomalousness in Action', in *The Philosophy of Donald Davidson*, ed. by Lewis Edwin Hahn, (La Salle: Open Court, 1999), pp. 623-635

— 'Agency and Actions', *Royal Institute of Philosophy Supplement*, 55 (2004), 1-23

— 'A Disjunctive Conception of Acting for Reasons', in *Disjunctivism: Perception, Action, Knowledge*, ed. by Adrian Haddock and Fiona Macpherson, (Oxford: OUP, 2008), pp. 244-261

— 'The Standard Story of Action: An Exchange (2)', in *Causing Human Actions*, ed. by Jesús H. Aguilar and Andrei A. Buckareff, (Cambridge, MA: MIT Press, 2010), pp. 57-68

— 'Actions in Their Circumstances', in *Essays on Anscombe's* Intention, ed. by Anton Ford, Jennifer Hornsby and Frederick Stoutland, (Cambridge, MA: Harvard University Press, 2011), pp. 105-127

— 'Causality and "the mental"', *Humana.Mente Journal of Philosophical Studies*, 29 (2015), 125-140

Hubin, Donald C., 'What's Special About Humeanism?', *NOÛS*, 33 (1999), 30-45

Humberstone, Lloyd, 'Directions of Fit', *Mind,* 101 (1992), 59-83

Hume, David, *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. by L. A. Selby-Bigge (Oxford: Clarendon, 1957)

— *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*, ed. by L. A. Selby-Bigge and P. H. Nidditch, (Oxford: Clarendon, 1978)

Hursthouse, Rosalind, 'Virtue Theory and Abortion', *Philosophy & Public Affairs*, 20 (1991), 223-246

Hyman, John, *Action, Knowledge, and Will,* (Oxford: OUP, 2015)

Joseph, H. W. B., *Some Problems in Ethics*, (Oxford: OUP, 1933)

Kant, Immanuel, *Groundwork of the Metaphysics of Morals*, trans. by Mary Gregor, (Cambridge: CUP, 1998)

— *Groundwork of the Metaphysics of Morals*, trans. by Mary Gregor and Jens Timmermann, (Cambridge: CUP, 2011)

Kirchin, Simon, 'The Shapelessness Hypothesis', *Philosophers' Imprint*, 10:4 (2010), 1-28

— 'Introduction: Thick and Thin Concepts', in *Thick Concepts*, ed. by Simon Kirchin, (Oxford: OUP, 2013), pp. 1-19

— 'Thick Concepts and Thick Descriptions', in in *Thick Concepts*, ed. by Simon Kirchin, (Oxford: OUP, 2013), pp. 60-77

Korsgaard, Christine M., *The Sources of Normativity*, (Cambridge: CUP, 1996)

— 'Kant's analysis of obligation: The argument of *Groundwork I*', in Christine M. Korsgaard, *Creating the Kingdom of Ends*, (Cambridge: CUP, 1996), pp. 43-76

— 'Skepticism about practical reason', in Christine M. Korsgaard, *Creating the Kingdom of Ends*, (Cambridge: CUP, 1996), pp. 311-334

— 'Two distinctions in goodness', in Christine M. Korsgaard, *Creating the Kingdom of Ends*, (Cambridge: CUP, 1996), pp. 249-274

— 'The Normativity of Instrumental Reason', in *Ethics and Practical Reason*, ed. by Garrett Cullity and Berys Gaut, (New York: Clarendon, 1998), pp. 215-254

Kosman, L. A., 'Being Properly Affected: Virtues and Feelings in Aristotle's Ethics', in *Essays on Aristotle's Ethics*, ed. by Amélie Oksenberg Rorty, (Los Angeles: University of California Press, 1980), pp.103-116

Kripke, Saul, *Wittgenstein on Rules and Private Language*, (Cambridge, MA: Harvard University Press, 1982)

Lewis, David, 'Desire as Belief', *Mind*, 97 (1988), 323-332

— 'Desire as Belief II', *Mind*, 105 (1996), 303-313

Little, Margaret Olivia, 'Virtue as Knowledge: Objections from the Philosophy of Mind', *NOÛS*, 31 (1997), 59-79

MacBride, Fraser, 'Lewis's Animadversions on the Truthmaker Principle', in *Truthmakers: The Contemporary Debate,* ed. by Helen Beebee and Julian Dodd, (Oxford: Clarendon Press, 2005), pp. 117-140

Mackie, J. L., *The Miracle of Theism: Arguments for and against the existence of God*, (Oxford: Clarendon, 1982)

— *Ethics: Inventing Right and Wrong*, (London: Penguin, 1990)

Markovits, Julia, 'Acting for the Right Reasons', *Philosophical Review*, 119 (2010), 201-242

McCarty, Richard, *Kant's Theory of Action*, (Oxford: OUP, 2009)

McDowell, John, *Mind and World*, (Cambridge, MA: Harvard University Press, 1996)

— 'Criteria, Defeasibility, and Knowledge', in John McDowell, *Meaning, Knowledge, and Reality*, (Cambridge, MA: Harvard University Press, 1998), pp. 369-394

— 'In Defence of Modesty', in John McDowell, *Meaning, Knowledge, and Reality*, (Cambridge, MA: Harvard University Press, 1998), pp. 87-107

— 'Might There be External Reasons?', in John McDowell, *Mind, Value, and Reality*, (Cambridge, MA: Harvard University Press, 1998), pp. 95-111

— 'Non-Cognitivism and Rule-Following', in John McDowell, *Mind, Value, and Reality*, (Cambridge, MA: Harvard University Press, 1998), pp. 198-218

— 'Some Issues In Aristotle's Moral Psychology', in John McDowell, *Mind, Value, and Reality*, (Cambridge, MA: Harvard University Press, 1998), pp. 23-49

— 'The Role of *Eudaimonia* in Aristotle's Ethics', in John McDowell, *Mind, Value, and Reality*, (Cambridge, MA: Harvard University Press, 1998), pp. 3-22

— 'Virtue and Reason', in John McDowell, *Mind, Value, and Reality*, (Cambridge, MA: Harvard University Press, 1998), pp. 50-73

— 'Incontinence and Practical Wisdom in Aristotle', in John McDowell, *The Engaged Intellect*, (Cambridge, MA: Harvard University Press, 2009), pp. 59-76

McHugh, Conor, 'The Truth Norm of Belief', *Pacific Philosophical Quarterly*, 93 (2012), 8-30

McHugh, Conor, and Daniel Whiting, 'The Normativity of Belief', *Analysis Reviews*, 74 (2014), 698-713

Melden, A. I., *Free Action,* (London: Routledge and Kegan Paul, 1961)

Mele, Alfred, *Motivation and Agency,* (New York: OUP, 2003)

Milliken, John, 'In a Fitter Direction: Moving Beyond the Direction of Fit Picture of Belief and Desire', *Ethical Theory and Moral Practice,* 11 (2008), 563-571

Moran, Richard, and Martin J. Stone, 'Anscombe on Expression of Intention: An Exegesis', in *Essays on Anscombe's* Intention, ed. by Anton Ford, Jennifer Hornsby and Frederick Stoutland, (Cambridge, MA: Harvard University Press, 2011), pp. 33-75

Morris, Michael, *The Good and the True*, (Oxford: Clarendon, 1992)

— 'Mind, World and Value', in *Current Issues in Philosophy of Mind*, ed. by Anthony O'Hear, (Cambridge: CUP, 1998), pp. 303-320

— 'Akrasia in the *Protagoras* and the *Republic*', *Phronesis*, 51 (2006), 195-229

Nagel, Thomas, *The Possibility of Altruism*, (Oxford: Clarendon Press, 1970)

Parfit, Derek, 'Reasons and Motivation', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 71 (1997), 99-130

— *On What Matters: Volume 1*, (New York: OUP, 2011)

Pettit, Philip, 'Humeans, Anti-Humeans, and Motivation', in *Mind*, 97 (1988), 530-533

Pettit, Philip and Michael Smith, 'Backgrounding Desire', *Philosophical Review*, 99 (1990), 565-592

Plato, *Protagoras*, trans. by Stanley Lombardo and Karen Bell, in *Plato: Complete Works*, ed. by John M. Cooper, (Indianapolis: Hackett, 1997)

Platts, Mark, *Ways of Meaning: An Introduction to Philosophy of Language*, (Cambridge, MA: MIT Press, 1997)

Price, Huw, 'Defending Desire-as-Belief', *Mind*, 98 (1989), 119-127

Prichard, H. A. 'Does Moral Philosophy Rest on a Mistake?', in H. A. Prichard, *Moral Obligation* (Oxford: OUP, 1968), pp. 1-7

Quinn, Warren, 'Putting rationality in its place', in *Value, Welfare, and Morality*, ed. by R. G. Frey and Christopher W. Morris, (New York: CUP, 1993), pp. 26-50

Radcliffe, Elizabeth S., 'Hume's Psychology of the Passions: The Literature and Future Directions', *Journal of the History of Philosophy*, 53 (2015), 565-605

Radford, Andrew, Martin Atkinson, David Britain, Harald Clahsen, and Andrew Spencer, *Linguistics*: *An Introduction,* (Cambridge: CUP, 2009)

Railton, Peter, 'On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action', in *Ethics and Practical Reason*, ed. by Garrett Cullity and Berys Gaut, (Oxford: Clarendon Press, 1998) 53-80

Raz, Joseph, 'Agency, Reason, and the Good', in Joseph Raz, *Engaging Reason: On the Theory of Value and Action,* (Oxford: OUP, 1999), pp. 22-46

Reath, Andrews, 'Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination', in Andrews Reath, *Agency and Autonomy in Kant's Moral Theory: Selected Essays*, (Oxford: OUP, 2006), pp. 8-25

Roberts, Debbie, 'Shapelessness and the Thick', *Ethics*, 121 (2011), 489-520

— 'It's Evaluation, Only Thicker', in *Thick Concepts*, ed. by Simon Kirchin, (Oxford: OUP, 2013), pp. 78-96

Russell, Bertrand, *An Inquiry into Meaning and Truth,* (London: Unwin, 1940)

— 'The Philosophy of Logical Atomism', in Bertrand Russell, *Logic and Knowledge*, (London: Routledge, 1956), pp. 177-281

Sawyer, Sarah, 'Minds and Morals', *Philosophical Issues*, 24 (2014), 393-408

Scanlon, T. M., *What We Owe To Each Other*, (Cambridge, MA: Harvard University Press, 2000)

— *Being Realistic About Reasons*, (New York: OUP, 2016)

Schmidtz, David, 'Choosing Ends', in *Varieties of Practical Reasoning,* ed. by Elijah Millgram, (Cambridge MA: MIT Press, 2001), pp. 237-258

Schroeder, Mark, *Slaves of the Passions*, (Oxford: OUP, 2007)

Schuler, G. F., *Desire: Its Role in Practical Reason and the Explanation of Action*, (Cambridge MA: MIT Press, 1995)

— *Reasons and Purposes*, (Oxford: OUP, 2003)

Searle, John, *Expression and Meaning: Studies in the Theory of Speech Acts*, (New York: CUP, 1979)

— *Intentionality*, (Cambridge: CUP, 1983)

Sehon, Scott R., 'An Argument Against the Causal Theory of Action Explanation', *Philosophy and Phenomenological Research,* 60 (2000), 67-85

Shah, Nishi, and J. David Velleman, 'Doxastic Deliberation', *The Philosophical Review,* 114 (2005), 497-534

Sinhababu, Neil, 'The Humean Theory of Motivation Reformulated and Defended', *Philosophical Review*, 118 (2009), 465-500

— *Humean Nature: How desire explains action, thought and feeling*, (New York: OUP, 2017)

Skorupski, John, *The Domain of Reasons*, (Oxford: OUP, 2010)

Smith, Michael, 'The Humean Theory of Motivation', *Mind*, 96 (1987), 36-61

— *The Moral Problem*, (Oxford: Blackwell, 1994)

— 'The Standard Story of Action: An Exchange (1)', in *Causing Human Actions*, ed. by Jesús H. Aguilar and Andrei A. Buckareff, (Cambridge, MA: MIT Press, 2010), pp. 45-56

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

Stefánsson, H. Orri, 'Desires, Beliefs, and Conditional Desirability', *Synthese,* 191 (2014), 4019-4035

Steward, Helen, *The Ontology of Mind: Events, Processes, and States*, (Oxford: Clarendon, 1997)

Stocker, Michael, 'Desiring the Bad: An Essay in Moral Psychology', *The Journal of Philosophy*, 76 (1979), 738-753

Stroud, Barry, *Hume*, (London: Routledge, 1977)

— 'Sense-Experience and the Grounding of Thought', in *Reading McDowell: On Mind And World*, ed. by Nicholas H. Smith, (London: Routledge: 2002), pp. 79-91

Tappolet, Christine, 'Through Thick and Thin: *good* and its determinates', *Dialectica*, 58 (2004), 207-221

Tenenbaum, Sergio, 'Directions of Fit and Motivational Cognitivism', in *Oxford Studies in Metaethics*, ed. by Russ Shafer-Landau, (Oxford: OUP, 2006), pp. 235-264

— *Appearances of the Good: An Essay on the Nature of Practical Reason*, (New York: CUP, 2007)

— 'Guise of the Good', in *The International Encyclopedia of Ethics*, ed. by Hugh LaFollette, (Oxford: Blackwell, 2013), pp. 2262-2271

Timmermann, Jens, *Kant's Groundwork of the Metaphysics of Morals: A Commentary*, (Cambridge: CUP, 2007)

Urmson, J. O., 'Aristotle's Doctrine of the Mean' in *Essays on Aristotle's Ethics*, ed. by Amélie Oksenberg Rorty, (Los Angeles: University of California Press, 1980), pp. 157-170

Van Roojen, Mark, 'Humean Motivation and Humean Rationality', *Philosophical Studies*, 79 (1995), 37-57

Väyrynen, Pekka, 'Thick Concepts: Where's Evaluation?', in *Oxford Studies in Metaethics: Volume 7,* ed. by Russ Schaffer-Landau, (Oxford: OUP, 2012)

Velleman, J. David, *The Possibility of Practical Reason*, (Oxford: Clarendon, 2000)

— 'The Guise of the Good', in J. David Velleman, *The Possibility of Practical Reason*, (Oxford: Clarendon, 2000), pp. 99-122

— 'What Happens When Someone Acts?', in J. David Velleman, *The Possibility of Practical Reason*, (Oxford: Clarendon, 2000), pp. 123-143

— 'On the Aim of Belief', in J. David Velleman, *The Possibility of Practical Reason*, (Oxford: Clarendon, 2000), pp. 244-282

Wallace, R. Jay, 'How to Argue about Practical Reason', *Mind*, 99 (1990), 355-385

Watson, Gary, 'Free Agency', *The Journal of Philosophy*, 72 (1975), 205-220

Wedgwood, Ralph, 'The Aim of Belief', *Philosophical Perspectives,* 16 (2002), 267-297

Whiting, Daniel, 'Does Belief Aim (Only) at the Truth?', *Pacific Philosophical Quarterly*, 93 (2012), 279-300

Wiggins, David, 'Weakness of Will, Commensurability, and the Objects of Deliberation and Desire', in *Essays on Aristotle's Ethics*, ed. by Amélie Oksenberg Rorty, (Los Angeles: University of California Press, 1980), pp. 241-266

— 'Deliberation and Practical Reason', in *Varieties of Practical Reasoning*, ed. by Elijah Millgram (Cambridge MA: MIT Press, 2001), pp. 279-300

— *Sameness and Substance Renewed*, (Cambridge: CUP (Virtual Publishing), 2003)

Williams, Bernard, 'Consistency and Realism', in Bernard Williams, *Problems of the Self: Philosophical Papers 1956-1972*, (Cambridge: CUP, 1973), pp. 187-205

— 'Replies' in *World, Mind, and Ethics: Essays on the philosophy of Bernard Williams*, ed. by J. E. J. Altham and Ross Harrison, (New York: CUP, 1995), pp. 185-224

— 'Internal and external reasons', in Bernard Williams, *Moral Luck: Philosophical Papers 1973-1980*, (Cambridge: CUP, 1997), pp. 101-113

— 'Persons, character and morality' in Bernard Williams, *Moral Luck: Philosophical Papers 1973-1980*, (Cambridge: CUP, 1997), pp. 1-19

— 'Internal reasons and the obscurity of blame', in Bernard Williams, *Making sense of humanity and other philosophical papers 1982-1993*, (Cambridge: CUP, 1998), pp. 35-45

— 'Postscript: Some Further Notes on Internal and External Reasons', in *Varieties of Practical Reasoning*, ed. by Elijah Millgram (Cambridge MA: MIT Press, 2001), pp. 91-97

— *Ethics and the Limits of Philosophy*, (Abingdom: Routledge, 2006)

Williamson, Timothy, *Knowledge and Its Limits,* (Oxford: OUP, 2000)

Zangwill, Nick, 'Directions of Fit and Normative Functionalism', *Philosophical Studies*, 91 (1998), 173-203

Appendix 1

List of Numbered Arguments in Alphabetical Order

(A) Alvarez's Argument

(A1)   An attitude is cognitive only if it has propositional content.

(A2)   Having an attitude with propositional content requires a capacity to reason.

(A3)   Some sorts of desire may be had both by people and some non-human animals.

(A4)   Non-human animals lack the capacity to reason.

So,

(A5)   Some desires do not have propositional content.

So,

(A6)   Some desires are non-cognitive attitudes.

(A*) Alvarez Argument (Modified)

(A1)   An attitude is cognitive only if it has propositional content.

(A2)   Having an attitude with propositional content requires a capacity to reason.

(A3*)   Some sorts of belief may be had both by people and some non-human animals.

(A4)   Non-human animals lack the capacity to reason.

So,

(A5*)   Some beliefs do not have propositional content.

So,

(A6*)   Some beliefs are non-cognitive attitudes.

(DC) Davidson's Challenge

(DC1) An account of the explanatory character of psychological explanation must be able to explain the force of the psychological 'because'.

(DC2) Causal accounts of the explanatory character of psychological explanation are able to explain the force of the psychological 'because'.

(DC3) The explanation of the force of the psychological 'because' given by causal accounts fits attitudes and actions into an explanatory pattern that we understand as well as any.

So,

(DC4) Psychological explanation is a species of causal explanation.

(G) Grammatical Argument for Desire as Non-Cognitive

(G1)   The content of a desire is not specifiable with an indicative sentence.

(G2)   Only indicative sentences are truth-apt.

(G3)   The content of any cognitive attitude is specifiable with a truth-apt sentence.

So,

(G4)   The content of a desire is not the content of a cognitive attitude.

So,

(G5)   Desires are non-cognitive attitudes.

(H) Hyman's Argument

(H1)   Dispositions generally are specifiable with an infinitival phrase

So

(H2)   If desires were dispositions, then they would be specifiable with an infinitival phrase.

(H3)   The content of desire is specifiable with an infinitival phrase.

(H4)   If (H2) and (H3), then (H3) is best explained by conceiving of desires as dispositions.

So,

(H5)   (H3) is best explained by conceiving of desires as dispositions.

(HA) Hume's Anti-Realist Argument

(NA1) The nature of a subject is revealed by the nature of the impressions from which our thoughts about it are derived.

(NA2) The impressions from which our thoughts about morality are derived are impressions of pleasure, pain, and passions, not impressions of moral qualities.

So,

(NA3) Morality concerns pleasure, pain, and passions, not moral qualities.

(HM) Hume's Moral Argument

(HM1) Moral 'judgements' are the grounds of some actions.

(HM2) Nothing belonging to the power of reason is the ground of any action (from HP6).

So,

(HM3) Moral 'judgements' do not belong to reason.

So,

(HM4) Moral 'judgements' are not really judgements at all (from Reason).

(HP) Hume's Psychological Argument

> (HP1)   Something can be the ground of an action only if it is its cause.
>
> (HP2)   Something can be the cause of an action, only if it is a *forceful impulse*.
>
> (HP3)   Reason is only the power of judging how things are, and how they are related.
>
> So,
>
> (HP4)   Nothing belonging to the power of reason is a forceful impulse.
>
> So,
>
> (HP5)   Nothing belonging to the power of reason is the cause of any action.
>
> So,
>
> (HP6)   Nothing belonging to the power of reason is the ground of any action.
>
> So,
>
> (HP7)   Reason is not a power that can provide the ground of any action.

(MQ) Mackie's Argument from Queerness

> (MQ1)   If values were objective, they would be entities, properties, or relations of a queer sort
>
> (MQ2)   There is nothing of such a queer sort.
>
> So,
>
> (MQ3)   Values are not objective.

(Tr) Transparency Argument

> (Tr1)   Knowledge-explanations are psychological explanations.
>
> (Tr2)   Psychological explanations identify psychological attitudes in the causal aetiology of actions.
>
> (Tr3)   If knowledge-explanations are equivalent to reasons-explanations, then either they do not identify psychological attitudes in the causal aetiology of actions, or reasons are (standardly) facts about psychological attitudes in the causal aetiology of actions.
>
> (Tr4)   Knowledge-explanations are equivalent to reasons-explanations.
>
> So
>
> (Tr5)   Either knowledge-explanations do not identify psychological attitudes in the causal aetiology of actions, or reasons are (standardly) facts about psychological attitudes in the causal aetiology of actions.
>
> (Tr6)   Reasons are not (standardly) facts about psychological attitudes in the causal aetiology of actions.
>
> So

(Tr7)   Knowledge-explanations do not identify psychological attitudes in the causal aetiology of actions.

But

(Tr8)   If (Tr7), then knowledge-explanations are not psychological explanations

So

(Tr9)   Knowledge explanations are not psychological explanations.

(TA) Smith's Teleological Argument

(TA1)   Having a motivating reason is, *inter alia*, having a goal

(TA2)   Having a goal *is* being in a state with which the world must fit

(TA3)   Being in a state with which the world must fit is desiring.

So:

(TA4)   Anything that constitutes a motivating reason consists, at least in part, in being in a state of desiring something.

Appendix 2

List of Abbreviations in Alphabetical Order

(B) Belief

What it is to believe that *p* is to have a view as to how things are, namely that the way things are is that *p*

(BEC) Belief Explanation to Causal Aetiology

If *A* φ-ed because she believed that *p*, her believing that *p* was in the causal aetiology of her φ-ing

(BRE) Belief is Rationally Evaluable

Belief is rationally evaluable

(BS) Blindspot Proposition

It's raining and no-one believes that it is raining

(CGN) Conditional Goodness Norm

For any *A* and any *p*: if *A* considers whether to desire that *p*, then *A* ought to (desire that *p*) if and only if it would be good if *p*

(CGN*) Conditional Goodness Norm*

For any *A* and any *p*: if *A* considers whether to desire that *p*, then *A* ought to (desire that *p*) if and only if it would be good if *p*, and were *A* to desire that *p* they would not have any incompatible desires

(CIB) Constitutive Intention of Belief

For any *p* (I intend that (I do not believe that *p* when not-*p*))

(CIB*) Constitutive Intention of Belief*

I intend that (for any *p* (I do not believe that *p* when not-*p*))

(CID) Constitutive Intention of Desire

For any *p* (I intend that (it be the case that *p* if I desire that *p*)

(CID*) Constitutive Intention of Desire*

I intend that (for any p (it be the case that p if I desire that p)).

(CTN) Conditional Truth Norm

For any A and any p: if A considers whether p, then A ought to believe that p if and only if it is true that p

(CTN*) Conditional Truth Norm*

For any A and any p: if A considers whether p, then A ought to believe that p if and only if p would be true were A to believe that p.

(DRE) Desire is Rationally Evaluable

Desire is rationally evaluable.

(DS) *De Se* Proposition

I believe that a conspiracy of mice secretly controls the world.

(E$_o$)Truth Operator Equivalence Schema

It is true that p if and only if p

(EC) Explanatory Constraint

R is a reason for A to φ only if (i) A could be motivated to φ for R, and (ii) were A to φ for R, R would figure in a true explanation of A's φ-ing.

(EED) Exclusive Efficacy of Desire

Desires, and only desires, are motivationally efficacious psychological states or attitudes.

(F) Falsity as a Defect

False belief is faulty or defective.

(FAD) Fundamental Arbitrariness of Desire

Fundamentally, at least, desire is not rationally evaluable.

(FN) Falsity Norm

For any A and any p: if it is false that p, then A ought to (not believe that p).

(G) Goodness Obligation

> The good is what one ought to desire.

(GN) Goodness Norm

> For any *A* and any *p*: one ought to desire that *p* if and only if it would be good if *p*.

(GN$_n$) Goodness Norm (narrow scope)

> For any *A* and any *p*: one ought to (desire that *p*) if and only if it would be good if *p*.

(GN$_w$) Goodness Norm (wide scope)

> For any *A* and any *p*: one ought to (desire that *p* if and only if it would be good if *p*).

(K1)-(K5) Kantian Theses

> (K1)   A motive is rationally evaluable only if it is a moral motive.
>
> (K2)   It is possible for a motive to be rationally evaluable
>
> (K3)   If something is a moral motive, then it is rationally evaluable.
>
> (K4)   If a motive is rationally evaluable, then it is derived from the structure of Reason.
>
> (K5)   If something is a moral motive, then it is derived from the structure of Reason.

(KE) Knowledge-Explanation Schema

> *A* φ-ed because she knew that *p*.

(NB) Norm of Belief

> For any *A* and any *p*: (i) *A* ought to (believe that *p*) only if it is true that *p*;
>
> *and,*
>
> (ii) if it is not true that *p*, *A* ought to (not believe that *p*)

(NB$_{sch}$) Norm of Belief (school funding)

> For any *A*: (i) *A* ought to (believe that it would be good if schools were properly funded) only if it is true that it would be good if schools were properly funded;
>
> *and,*
>
> (ii) if it is not true that it would be good if schools were properly funded, *A* ought to (not believe that it would be good if schools were properly funded).

(NC) Normative Condition

> An account of directions of fit must make sense of the normativity of the original metaphorical characterisations.

(ND) Norm of Desire

> For any *A* and any *p*: (i) *A* ought to (desire that *p*) only if it would be good if *p*;
>
> *and*,
>
> (ii) if it would not be good if *p*, *A* ought to (not desire that *p*).

(ND!) Norm of Desire (redacted)

> For any *A* and any *p*: (i) *A* ought to (…) only if it would be good if *p*;
>
> *and*,
>
> (ii) if it would not be good if *p*, *A* ought to (not…)

(NEB) Norm of Evaluative Belief

> For any *A* and any *p*: (i) *A* ought to (believe that it would be good if *p*) only if it is true that it would be good if *p*;
>
> *and,*
>
> (ii) if it is not true that it would be good if *p*, *A* ought to (not believe that it would be good if *p*).

(NEB*) Norm of Evaluative Belief*

> For any *A* and any *p*: (i) *A* ought to (believe that it would be good if *p*) only if it would be good if *p*;
>
> *and,*
>
> (ii) if it would not be good if *p*, *A* ought to (not believe that it would be good if *p*).

(NEB*!) Norm of Evaluative Belief* (redacted)

> For any *A* and any *p*: (i) *A* ought to (…) only if it would be good if *p*;
>
> *and,*
>
> (ii) if it would not be good if *p*, *A* ought to (not…).

(P1) Problem One

> What could ground the Humeans' acceptance of (EED), interpreted so as to support (RDB)?

(P2) Problem Two

>What could ground the Humeans' acceptance of (FAD)?

(PC) Partial Constitution

>For any A, φ, and p: For A to have φ-ed because she believed that if she φ-ed she would bring it about that p, and she desired that p, is for A's φ-ing, in its essential nature, to have been such that:
>
>(i) it ought to have been performed only if it was true that (a) A's φ-ing would have brought it about that p, and that (b) it would have been good if p; *and*,
>
>(ii) if it was not true that (a) A's φ-ing would have brought it about that p, or that (b) it would not have been good if p, then it ought not to have been performed.

(PC*) Partial Contitution*

>For any A, φ, and p: For A to have pulled the child out of the road because she believed that if she pulled the child out of the road she would bring it about that the child was rescued , and she desired that the child was rescued, is for A's pulling the child out of the road, in its essential nature, to have been such that:
>
>(i) it ought to have been performed only if it was true that (a) A's pulling the child out of the road would have brought it about that the child was rescued, and that (b) it would have been good if the child was rescued; *and*,
>
>(ii) if it was not true that (a) A's pulling the child out of the road would have brought it about that the child was rescued, or that (b) it would not have been good if the child was rescued, then it ought not to have been performed.

(PE) Psychological Explanation Schema

>A φ-ed because she believed that p and desired that q,

(RDB) Reasons as Desire-Based

>Reasons are *desire-based*.

(RE) Reasons-Explanation Schema

>A φ-ed because p.

(RN) The Normativity of Reasons

>Reasons are essentially normative.

(SN-SA) Same Norm-Same Attitude

> The constitutive norms of attitude A and attitude B are the same if and only if A just is
> B

(T) Anti-Pyrrhonist Truth Obligation

> The truth is what one ought to believe.

(TN) Truth Norm

> For any A and any p: A ought to believe that p if and only if it is true that p.

(TN$_n$) Truth Norm (narrow scope)

> For any A and any p: A ought to (believe that p) if and only if it is true that p.

(TN$_{nec}$) Truth Norm (necessary condition)

> For any A and any p: A ought to (believe that p) only if it is true that p.

(TN$_{suf}$) Truth Norm (sufficient condition)

> For any A and any p: if it is true that p, then A ought to (believe that p).

(TN$_w$) Truth Norm (wide scope)

> For any A and any p: A ought to (believe that p if and only if it is true that p).

(UC) Universality Condition

> An account of directions of fit must hold universally for all possible instances of states
> with one or the other direction of fit.

(Z) Zangwill's Condition of Adequacy

> The dispositions one has when one believes or desires that p ought to be *explained by*
> the account of what it is to have a belief or desire that p.

(~TN) Non-Truth Norm

> For any A and any p: if it is not true that p, then A ought to (not believe that p).