Habits in action: a corrective to the neglect of habits in contemporary philosophy of action

William James Pollard

HABITS IN ACTION

A Corrective to the Neglect of Habits in Contemporary Philosophy of Action

William James Pollard, PhD Thesis, 2002

Abstract

I propose that if we pay proper attention to habits, we can correct distortions in prevailing accounts of action, and make progress in a number of contemporary debates. First I describe the everyday phenomenon of habit, and sketch the context as we find it within contemporary analytic philosophy. I then develop a notion of habit which has its origins in Wittgenstein, Ryle and Aristotle. The generic notion upon which all three thinkers draw is that of a kind of behaviour which is *repeated; automatic,* in the sense that it does not involve deliberation or trying; and *responsible,* since it is under the agent's control. I call such behaviour habitual action.

Third, I reject the widely held view that the class of rational actions and the class of actions which we perform "for reasons" are equivalent. This view, made popular by Davidson, distorts our conception of rational actions by taking deliberated actions to be the sole paradigm. I suggest that this is an "intellectualist" error, which gives too prominent a place to our deliberative capacity in our picture of rational actions. I argue, against this, that on many of the occasions that we act habitually, we do not act for reasons, although we do act rationally, in ways that I spell out.

Fourth and finally I outline how broadening our conception of rational actions to include many of those we perform habitually allows us to make progress in contemporary debates. I focus on the debate in meta-ethics between Humean (Smith, Blackburn) and anti-Humean (McDowell) accounts of moral motivation. I argue that properly understood, habits form a crucial part of the anti-Humean argument - one which has hitherto been obscure in McDowell. I suggest other debates to which an understanding of habits could contribute, such as the project of "naturalising" rational action.

HABITS IN ACTION

A CORRECTIVE TO THE NEGLECT OF HABITS IN CONTEMPORARY PHILOSOPHY OF ACTION

William James Pollard

PhD Thesis

University of Durham
Department of Philosophy
2002

TABLE OF CONTENTS

DECLARATION

No part of this thesis has previously been submitted for a degree in this or any other university. Where the information contained in this thesis is not the original work of the author, the sources are acknowledged within the text, in footnotes, and in the Bibliography.

STATEMENT OF COPYRIGHT

ACKNOWLEDGEMENTS

I am indebted to a number of people for the completion of this thesis. Most obviously I am grateful to my supervisor Jonathan Lowe, for his rigorous comments, his guidance, and his patience with my Wittgensteinian tendencies.

I thank members of EIDOS, the Durham postgraduate philosophy group, who have listened to endless draft material, and been consistently constructive in what they have said in response. In particular I thank three philosophical friends: Alan Brown, for his detailed comments and conversations on drafts which hardly deserved it; Dawn Philips, for making me think I can do analytic philosophy; and Geraldine Coggins, for reminding me, in the nicest possible way, that often I can't.

For helpful philosophical pointers and encouragement at critical points in my research I thank Paddy FitzPatrick, Chris Hookway, Al Mele and Bas Van Fraassen. And I am particularly grateful to my examiners, Rowland Stout and Andy Hamilton, not only for giving me an extremely enjoyable viva, but also for giving me a sense of the work which lies ahead.

Thanks also to Chloe and Mahalia, for whom I have too often been Mr. Grumpy when I should have been Mr. Tickle. And of course to Mum and Dad, for their constant support, and providing sanctuary.

My greatest debt is to Soran Reader, my fiancée, who has lived with my habits for four years, who has supported me financially since my AHRB funding ran out, and without whom I would still be swimming with the anchovies. It will convey something of my intellectual debt to her if I say that whilst some of the ideas within these pages are much better for her input, she shares responsibility for those that are not.

W.J.P.

Durham, 14th June 2002

1

CREATURES OF HABIT

> The aspects of things that are most important for us are hidden because of their simplicity and familiarity. (One is unable to notice something - because it is always before one's eyes.)
>
> Wittgenstein, *Philosophical Investigations*, §129

1. Noticing Habits

We are creatures of habit. We have ways of going about our lives that are so simple and familiar that we do them automatically, that is, without thinking about them. These are habits. Exercising them comes as second nature to us. Habits are involved in just about every human activity. Our daily routine of getting up in the morning, our walk to our workplace, our polite social interactions, our use of language, and even our patterns of thought, all depend upon habits. Unless we could rely on habits to take care of the familiar parts of what we do, such activities would simply be impossible. Habits free us to think about more important things, or about novel features of our situation.

We have a natural capacity to acquire habits during upbringing. Some of them, such as the habits involved in using a language, arise at least partly as a result of contact with other people with those habits. Other habits, such as mannerisms, arise without the assistance of others. Thankfully our parents saw to it that the habits we cultivated were by and large good ones, keeping bad habits to a minimum.

Habit acquisition does not of course end when adulthood arrives. If we practise a certain kind of behaviour for long enough, there is a good chance that it will become habitual. That is, we will come to do it automatically. That is how new ways of doing things are learned and kept, and how we become accustomed to new environments. In this way our persisting capacity for habit acquisition is extremely useful. But we also acquire habits unintentionally, and this can sometimes be a

nuisance. We have all found ourselves with habits that, on reflection, we would rather not have. For habits, once acquired, can be hard to lose.

In being repeated and automatic, exercises of habits seem similar to reflexes and bodily processes such as the heartbeat and digestion. But they are also importantly distinct from such phenomena. We agents cannot do much about our reflexes or bodily processes. Even if I wished to stop my heart from beating, I doubt that I could (though perhaps I could slow it down). And it is impossible for me to resist the extension of my leg if my knee is struck in the right place. These are things that happen to us. It would not normally make sense to hold each other responsible for these occurrences, or their results.

Exercising habits, in contrast, seems to be something we do. We can decide not to exercise a given habit on a particular occasion. We can override our habits for good reason, or on a whim. That is how we are able to lose them (or at least attempt to). And unlike reflexes and bodily processes, we can properly be held responsible for the exercises of our habits, or for their results. If I continue to drive on the left when I leave the ferry in Calais, the fact that it is a habit of mine does not free me from blame. We are personally responsible for our exercises of habits. In this way, habitual behaviours are unlike reflexes and bodily processes.

These statements about habits are platitudes. The fact is that habits are so familiar that in the normal run of things, we tend not to notice them, and they escape such remark. They form part of the background to our lives.

Of course habits do attract our attention from time to time. One sort of context in which we notice them is when things around us aren't as they usually are. For instance, when my keys are not on their usual hook, I still reach for them on my way out, "through force of habit", we might say by way of explanation. To our amusement my partner calls me by the name of her ex-partner, through a well-worn habit. We also refer to habits in order to say something about an individual's distinctive characteristics: "Geraldine is extremely sociable"; "George is always late for seminars"; "I go sculling regularly". These sorts of expression pick out idiosyncrasies - we could call them "individual habits" or "character traits", which contribute to making people who they uniquely are.

But although we tend to mention habits only to refer to unusual situations or features of people, they are nevertheless at work in the background all the time. When my keys are in their normal place, a habit helps me to locate them without effort, and allows me to keep thinking about the day ahead. My partner, in the main, gets my name right without having to remind herself. Thanks to habit. This harmonious conduct, remarkable though it is, will rarely merit comment in everyday life.

In a similar way, we all share innumerable habits that resist remark simply because we all share them. Many habits of etiquette, habits of language, and habits of reasoning are like this. Again, this commonality might be thought something of a miracle. But in our daily lives, our other preoccupations mean that we must take such things for granted. Referring to widely shared habits does not help to distinguish individuals, or serve any other obvious explanatory purpose in the usual run of things. As a result, shared habits, like our well-functioning individual habits, go unnamed and unnoticed. Habits of both sorts are subject to Wittgenstein's homily quoted above: we may be unable to notice these habits because they are always before our eyes.

But are habits important, as the homily also suggests? And if so, how? This thesis gives a positive answer the first of these questions, and provides something of an answer to the second.

2. *Philosophy and Habits*

Philosophers don't notice habits very much either. At least that is true of those who work within the philosophical tradition we call "analytic", a term which roughly marks out the philosophy that has been practised in Great Britain and America since around 1900. If we look to the places that we might expect to see serious treatments of habit, such as the philosophy of action (a subject which tends to be subsumed under the philosophy of mind), or to moral philosophy, if habits are mentioned at all, their treatment is cursory.

In his introduction to *The Philosophy of Action,* a collection of influential essays spanning the past 40 years, Al Mele (1997) notices that "Our ordinary explanations of

human actions draw upon a rich psychological vocabulary" (p. 1). These include "beliefs, desires, reasons (construed as psychological states by some), intentions, decisions, plans, and the like". No mention of habits here. But there is a glimmer. "Occasionally," he notices, "we advert as well to finely distinguished traits of character." Perhaps these are habits. But neither his lucid 26 page introduction, nor any of the otherwise excellent essays he collects, make anything of these latter items. Philosophers of action tend to be preoccupied with the debates begun by Donald Davidson (1980), who argued in 1963 that reasons are the causes of actions. Is Davidson right? If not, and reasons are not the causes, what are they? And then, what, if anything, does cause actions? In these discussions habits have played no role. The temptation to understand actions as originating from some faculty of thought remains part of our Cartesian heritage; one in which habits have no clear place.

Moral philosophy fares only slightly better. Hume, Kant and Mill remain influential figures in this field, and habits do not feature in their accounts of the moral life at all. However, the recent revival of the ethics of Aristotle, sparked by the publication of Elizabeth Anscombe's "Modern Moral Philosophy" (1958), promised to change all this. For Aristotle, the moral virtues, usually understood as "stable character traits", such as patience, temperance and charity, are essential for moral action. And in Book 2 of the *Nicomachean Ethics*, we find Aristotle proclaiming that the virtues are acquired through nothing less than (what it translated as) "habit".[1] But even in this burgeoning neo-Aristotelian tradition, which includes such thinkers as Philippa Foot (1978), Alasdair MacIntyre (1981) and John McDowell (1998), the significance of the specifically habitual nature of virtue remains under-appreciated. At best, habit is taken to provide a story about how we acquire the capacity for moral action. And this story is generally assumed to be compatible with some kind of psychological account of those actions.

What might explain the current attitude to habits? Philosophers like Gilbert Ryle (1949) and Peter Winch (1958) were influential in the middle of the 20th century and both were highly dismissive of habits. They argued that habits had no role in a

---

[1] 1103a14-19.

proper understanding of intelligent or meaningful behaviour. We could also speculate about whether the support for these views was partly due to historical forces in play at the time, and in particular the associations of habits with the disparaged "behaviourism" in psychology proposed by J. B. Watson (1919) and B. F. Skinner (1953). But whatever the historical explanation, it is now usual for philosophers, when they talk of habits at all, to talk of them as "mere" and associate them with behaviour which is "blind" or "non-rational".

Not all analytic philosophers have this disparaging attitude to habits. For example, Philip Pettit (1993) has recently tried to incorporate habits into his account of "rule following" behaviour (a topic about which I shall say much more in Chapter 2). But Pettit is exceptional in recent literature. A search of the *Philosopher's Index* reveals that those who think habits might have a role in understanding intelligent actions are few and far between. Nathan Brett wrote an article called "Human Habits" in 1981, in which he rejects the assumption, traced to Ryle and Winch, that habits are "neither intelligent nor purposeful" (p. 357). And R. S. Peters defends a role for habits in intelligent behaviour in a number of places (1958, 1963). But already we are going back a number of decades. Other recent literature on habits tends to be exegetical (Mixon 1980; Wright 1994; Crocker 1998) or to be found in journals of education (Hamm 1975; Götz 1989). Whilst some of these sources are helpful in providing suggestions for the investigation I have in mind, none provide detailed treatments.

For the purposes of this thesis, the gap in the current philosophical literature is a mixed blessing. On the positive side, it means that there is something to investigate. Something that may have escaped the notice of contemporary philosophers, to which we might profitably address ourselves. On the negative side, of course, the lack of awareness of habits means that the resources for prosecuting such a task are limited.

But two writers from the middle of the 20th century will prove helpful here. I think Ryle (1949) is one of them, despite his, on the face of it disparaging, attitude to habits mentioned above. For I believe he is working with an impoverished understanding of the idea. And if we correct for this we can see that Ryle has important things to say about how certain kinds of habit are philosophically important. I also think Wittgenstein (1953) in his later work, relied on an

understanding of habits, though the translation brings with it little in the way of explicit indications that this should be the case. I shall assess the work of these two writers, as well as Aristotle's contribution to our understanding of the notion, in Chapter 2. This will provide the material for a critique of some contemporary analytic debates, whose neglect of habits is to their cost. Or so I shall argue.

Since time and space is limited, the breadth of material covered must also be limited, however arbitrarily. I here restrict myself to mainstream analytic sources. It is with some regret that I do not explore the work of founding "pragmatist" philosophers such as William James, John Dewey and Charles Peirce, despite the fact that they may all have more or less important places for habits in their thinking. And neither do I offer a treatment of any writer from the "continental" tradition, though I am aware that the work of Jean-Paul Sartre, Maurice Merleau-Ponty, Hans-Georg Gadamer and Pierre Bourdieu could provide rich resources. The contribution of these traditions to the present topic will have to be assessed elsewhere.

3. *Intellectualism*

Philosophers are creatures of habit too. And philosophical habits, like others, are hard to break. Ironically, ignoring the philosophical significance of habits may be one of them. But there is another habit of thought to which today's analytic thinkers are introduced in their philosophical infancy. And that is the habit I call *intellectualism*. Intellectualism, as I understand it, is the unwarranted privileging of our capacity to reason. Like any unwarranted move, intellectualism is something to be avoided in philosophy, even on a single occasion. But calling intellectualism a habit captures two of its most pernicious features. We do it repeatedly; and we do it automatically, that is, without thought. And philosophers should know better. Intellectualism, then, is a bad philosophical habit.

Intellectualism pervades current philosophy of action. We saw something of this above in Davidson's thought. Davidson thinks that the reasons for which agents act are made up of two elements from their psychology: a positive attitude towards actions of a certain kind (a "pro attitude"), and a belief that the action in prospect is of that kind. And Davidson thinks that this model can work for all actions which are

done, as he says, "intentionally". I think Davidson's intellectualism consists in the fact that all intentional actions are correctly modelled on those upon which we actually perform some process of reasoning, ending with a decision to act in a certain way, that is, deliberation. And habitual actions just don't seem to be like that.

Davidson's ideas in this respect are hardly unusual. His ideas about reasons for actions have their roots in Hume, who, in his *Treatise* divides human psychology into "reason" and "passion", and claims that "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" (Hume 1740, p. 415). Since Humean "reason" is instrumental, in the sense that it cannot by itself motivate our actions, though for the action to be rational it must "guide" them, every action proper requires a contribution from both sides: a belief and a desire (the term "desire", like Davidson's "pro attitudes", being rather liberally understood to cover a whole range of attitudes and emotions). Davidson is not the only contemporary writer to defend this kind of view. Other recent defences have been given by David Lewis (1988, 1996), Simon Blackburn (1984a, 1994, 1998) and Michael Smith (1987, 1994, 1998). But more often than not, this Humean structure for the explanation of actions is simply assumed to be correct.

Why should we accept it? If the phenomenology of habits is any kind of guide, attributing beliefs and desires to an agent who is exercising a habit seems simply to be a prejudice. Habits *spare* us from feeling desires and entertaining beliefs, at least as they concern particular actions. And if anything like a "decision" is made when we exercise a habit, it appears quite unlike the conscious decisions which we all make from time to time. Are we really doing all these things as it were "behind the scenes"? Most philosophers will be ready with replies to such questions, which will require detailed treatment. But it looks at least possible that the Humean picture has an intellectualist case to answer.

Humean accounts of action have come under attack in recent years. John McDowell, in many influential papers (collected in his 1998), has been a prominent anti-Humean. Others include Arthur Collins (1987, 1997), Jonathan Dancy (1992, 2000), Warren Quinn (1993) and Rowland Stout (1996). Anti-Humeans have rejected many aspects of the Humean view. For instance, McDowell (1978), Dancy (1992) and

Quinn (1993) reject the claim that desires are necessary for action. McDowell (1995b) and Collins (1997) reject the idea that beliefs are essential. Stout (1996) has rejected both views. A promising feature of some of these arguments, such as those of Collins (1987, Ch. 6) and Stout (1996), is that the anti-Humean moves are based on a novel understanding of actions as *teleological* phenomena, that is, as processes that aim at some given outcome. Thus a belief-desire explanation is useful insofar as it specifies an outcome, but it does not form part of a proper teleological explanation of the action at all.

But even these relatively radical writers retain a central assumption. The assumption is that all actions, or at least all of the "rational" ones, are done, as it is said, "for the agent's reasons," or equivalently, "for reasons." And I think that given the way the idea of a reason is currently understood, even anti-Humean philosophers are guilty of intellectualism. For I don't think that habitual actions are done for reasons *at all*. They do not result from desire, from thought, and neither do they aim at anything in particular. But I don't think that disqualifies them from being rational. This is potentially an even more radical rejection of the Humean view than any yet proposed. If I am right, this thesis will be a significant contribution to our understanding of rational actions in general, and the idea of actions for reasons in particular.

There is an obvious reply to this intellectualist charge, which is to deny that exercises of habits are *bona fide* actions. Exercises of habits, the reply goes, are not sufficiently "intentional" as the preferred (and to my mind rather unhelpful) terminology has it, for them to count as such. But I think that this cannot be sustained without either distorting the idea of a habit so that it is little more than a reflex, or by presupposing an already intellectualised conception of what an action is. Either way, largely due to these sorts of prejudice, exercises of habits stand little chance of being actions proper.

My diagnosis is that intellectualism has become so habitual in our current thinking about action, that characterising actions without being intellectualist has become if not unthinkable, then unthought. What I want to suggest is that once cured of our intellectualist prejudices about the kinds of things actions, rationality and

habits are, we should admit exercises of habits into the class of rational actions. Redefining these things in non-intellectualised ways will take work. But that is the task at the heart of this thesis, and it will occupy Chapters 3 to 6.

Ridding philosophy of prejudice is worth doing in its own right, and the prejudice of intellectualism is no exception. But I think the pay-offs may extend far beyond admitting habitual actions into the class of rational actions. For the emerging anti-intellectualist philosophy of action promises to provide novel solutions to entrenched problems. Some of these problems, as they occur in the field of meta-ethics, will be addressed with the clarifications in hand, in Chapter 7. Others will have to be pursued elsewhere, and I shall mention some in Chapter 8.

4. Naturalism

This thesis is conducted against a background in which philosophers want to be naturalists, and I am no exception. That is to say, when we propose an account of some phenomenon, a requirement of that account is that it should be possible to offer a plausible story about how that phenomenon is a part of the same world that is revealed to us by the natural sciences, which, for the sake of argument, we can take to include physics, chemistry, biology and zoology. The assumption is that the disclosures of these sciences are straightforwardly natural, whereas for the phenomenon at hand, this needs to be shown. Crucial to the plausibility of such a story is that it should not depend upon any mysterious entities, leaps of faith or otherwise problematic moves. At least this much is Darwin's legacy to philosophy.

Action naturalism is the view that we can construct a naturalistic story about actions in particular. I intend the argument of this thesis not only to be compatible with action naturalism, but also to strengthen that view. However, naturalism comes in a number of varieties, and in saying this, I have in mind a particular sort of naturalism which many will not find acceptable. I briefly describe what sort of naturalism this is not, before describing what it is, and how the present project might strengthen it.

Popular kinds of naturalism attempt some kind of *reduction* of a given phenomenon to some other sort of thing, which is assumed to be less problematic.

Paradigm reductions include the claim that water is H_2O, or that heat is mean kinetic energy. If a reductive strategy is adopted to actions, a popular strategy is to reduce them to *bodily movements* of a certain kind.[2] Davidson (1980) has been more influential than most in defending this sort of reduction. Bodily movements are described in terms of the spatio-temporal location and trajectory of the agent's body (which need not actually be moving to count as such). Bodily movements are thought to be less problematic than actions because that puzzling phenomenon of agency is absent from such descriptions. Hence they are naturalistically respectable. All that remains is to distinguish those bodily movements which are also actions, from those which are not - pushings, reflexes, bodily processes, and so forth. This is most often done in terms of the movement's having a certain kind of etiology (e.g. Davidson 1980; Goldman 1970; Dretske 1988). According to these standard reductions, then, an action is nothing but a bodily movement that has been caused in the right sort of way.

Whatever the reductive base, however, a reduction violates what might be called *Butler's law.* Bishop Butler (1726) wrote in the preface to his *Sermons*, "Everything is what it is, and not another thing" (p. 14). And this is something that the reductive naturalist, it seems to me, flouts. In saying that x *really is* y, or x *is nothing but* y, even with the laudable aim of naturalism, is to deny the reality of x in favour of y. Reducing x to y claims more than to have provided one sort of explanation of x. There is an *identity* between x and y being posited, and that is what Butler would, I think rightly, object to.

A non-reductive naturalism, in contrast, finds continuity elsewhere. And in the variety that I endorse, that might be termed *continuity in forms of life.* I find this idea in the later work of Wittgenstein (1953), and in the writings of McDowell (1995a, 1996), who is himself influenced by Wittgenstein, as well as Aristotle. The idea is roughly that one kind of entity can be understood to be related to entities of another kind, by tracing the natural history of the lives in which those entities occur. So, for

---

[2] Jennifer Hornsby (1980) has pointed out that the phrase "bodily movement" is ambiguous and has both transitive (John moves his arm) and intransitive (his arm moves) senses. It is the intransitive sense I mean here.

instance, a naturalising account of water (if such be needed) would trace the life in which the term "water" is used, and would relate to the term "H_2O" presumably via some account of the scientific investigations into the structure of the stuff, and their historical development, culminating in the common conviction that "H_2O" is a proper way of formulating what we hitherto knew only as "water". Those who are optimistic about such a naturalism may go on to suggest that the motivation to ask whether water *really is* H_2O, and posit identities, should now seem less pressing. But the success of non-reductive naturalism does not depend on that motivational matter.

The same thing goes for action naturalism. Our first naturalising move should be to trace the life in which the term "action" occurs back to a life in which our preferred naturalistic materials occur. McDowell's "naturalism of second nature" does just this.[3]

McDowell begins with the assumption that our capacity for "mere behaviour", as opposed to action proper, is naturalistically respectable. This seems reasonable since it is a capacity we share with other animals. The continuity between our capacity for action and the capacity for mere behaviour is, for McDowell, supplied by the natural process of upbringing, during which we acquire a "second nature".

McDowell finds the germ of this idea in Aristotle's writings on ethics:

> The notion [of second nature] is all but explicit in Aristotle's account of how ethical character is formed. Since ethical character includes dispositions of the practical intellect, part of what happens when character is formed is that the practical intellect acquires a determinate shape. .... human beings are intelligibly initiated into this stretch of the space of reasons by ethical upbringing, which instils the appropriate shape into their lives. The resulting habits of thought and action are second nature. (McDowell 1996, p. 84)[4]

Though the above remarks concern specifically "ethical" upbringing and character, McDowell goes on to say that they apply quite generally.

---

[3] See especially McDowell (1996, pp. 78-86).

[4] "The space of reasons" is a metaphor McDowell borrows from Wilfrid Sellars (1956).

> The point is clearly not restricted to ethics. Moulding ethical character, which includes imposing a specific shape on the practical intellect, is a particular case of a general phenomenon: initiation into conceptual capacities, which includes responsiveness to other rational demands besides those of ethics. ... If we generalise the way Aristotle conceives the moulding of ethical character, we arrive at the notion of having one's eyes opened to reasons at large by acquiring a second nature. (McDowell 1996, p. 84)

Thus, according to McDowell, we have the materials for a naturalism for rational actions as such, without reductive methods, and hence without reference to bodily movements of any kind.

Now I do not want to say that McDowell's naturalism is unproblematic.[5] Indeed I think some of the details of just how we acquire our capacity for action are left if not wholly mysterious, at least shrouded in metaphor. What is it, for instance, to "have one's eyes opened to reasons at large"? And how is upbringing supposed to confer the possession of psychological states like beliefs and desires, onto an individual? But spelling out these and other details aside, I think that as an overall strategy for action naturalism, McDowell's is the best on offer. It is this kind of action naturalism that I shall assume to be right.

The above passage also suggests how a thesis on habits could contribute to such a naturalism. For it is striking that McDowell explicitly says that second nature comprises "*habits* of thought and action". Indeed the very phrase "second nature" captures something distinctive about habits already mentioned, that is, that we perform them automatically. So it begins to look as though the phenomenon of habit is going to be a key resource in a naturalism of second nature. If further motivation for an investigation into habits were needed, we have it here. And in the final chapter I shall be in a position to say how this kind of naturalism is strengthened by what I present in the intervening pages.

---

[5] For criticism see for example Wright (1996) and Williams (1996).

5. *Outline of the Thesis*

The thesis is divided into six central chapters, sandwiched between this introduction and a concluding chapter.

Chapter 2 is exegetical, and locates the project of anti-intellectualism in a historical tradition. I consider the work of three thinkers each of whom oppose a distinctive variety of intellectualism in a distinctive way. The thinkers are Aristotle, Ryle and Wittgenstein. What is notable about these three thinkers is that they each oppose intellectualism by drawing on notions which can be seen as kinds of habit. The notions in question share the three features of habitual actions of being at once repeated, automatic and responsible. Thus the provenance of the present anti-intellectualist project will be secured.

In Chapters 3 and 4 I offer a definition of habitual action, thus filling a notable gap in current analytic philosophy of mind and action. I begin in Chapter 3 by defining the idea of habitual behaviour, that is a behaviour which is both repeated and automatic. In Chapter 4 I develop the suggestion that we are personally responsible for at least some of our habitual behaviours. I do this by arguing that we have a certain kind of control over them. It is because we are personally responsible for our exercises of habits that they deserve to be called habitual *actions*. This marks a departure from existing intellectualist accounts of action.

Chapters 5 and 6 deal with the issue of whether habitual actions can be regarded as rational. In Chapter 5 I argue that according to the orthodox view of rational actions, according to which all rational actions are actions for reasons, habitual actions are not rational. However, I suggest that since this criterion for rational action is infected by intellectualism, we need not accept it. In Chapter 6 I consider an alternative conception of rational action which is not intellectualist. I develop suggestions made separately by Bas Van Fraassen (1989) and Bernard Gert (1998) which I term *the permissive conception of rational action*. I argue that this notion of rational action allows us to admit a large class of habitual actions into the class of rational actions. I also suggest that there is a class of habitual actions that can be regarded as rational in another sense, namely, that they are *reasonable* things to do.

In Chapter 7 I describe one way in which the anti-intellectualist philosophy of action which admits habitual actions, can solve problems in current debates. The debate in question is in the field of meta-ethics, and takes place between Humeans Smith and Blackburn on the one hand, and anti-Humean McDowell on the other. I suggest that as things currently stand the Humeans have the best of the arguments. But I argue that equipped with the idea of an action which is at once rational and habitual, the anti-Humean can offer new replies to Humean arguments. McDowell's own intellectualism has hitherto prevented these replies from emerging. This resolution has implications for important issues in meta-ethics, such as that of moral cognitivism.

In Chapter 8 I sketch out some of the implications for the philosophy of action, and directions in which this research may profitably go forward. In particular I say how attention to normal cases is promoted in importance, and how action naturalism can be significantly enhanced.

2

THREE CURES FOR INTELLECTUALISM

1. Curing Intellectualism

Intellectualism, the unwarranted privileging of reason and reasoning, is a perennial philosophical malady. Different strains of intellectualism arise in each generation of thought, each requiring a distinctive anti-intellectualist medicine, in the form of argument. But if the maladies they treat have something in common, it would be surprising if the medicines themselves did not share some features. And in the treatment of intellectualism, this is what we find.

In this chapter I examine the work of three historical thinkers who oppose forms of intellectualism. Those thinkers are Aristotle, Gilbert Ryle and Ludwig Wittgenstein. I do not claim that these thinkers exhaust the history of anti-intellectualism, or even that they exhaust the history of a certain kind of anti-intellectualism. But I do think that in the context of a thesis on habits, the similarities between their projects is both striking and significant.

For each thinker I do three things. First, I describe the particular intellectualism to which the thinker responds. Second, I describe their anti-intellectualist argument. Finally I link together this anti-intellectualist argument with those of the other two, by showing that they employ the notion of a behaviour which is repeated, automatic and responsible. The conclusion will be that for all three thinkers, the secret ingredient to the anti-intellectualist medicine is the notion of a habit.

This result will serve to motivate, and the discussion will inform, the detailed explication of the idea of a habit, a task which will be undertaken in Chapters 3 and 4.

2. *Aristotle*

In "Aristotle on Learning to be Good" Myles Burnyeat (1980) suggests that "Aristotle knew intellectualism in the form of Socrates' doctrine that virtue is knowledge" (p. 70).[1] It would, however, be misleading to say that Aristotle rejected the equation of virtue with knowledge wholesale. For we shall see shortly that Aristotle was content with an non-intellectualist interpretation of that doctrine. What Aristotle objected to was an intellectualist interpretation.

McDowell (1979) spells out this intellectualist interpretation as follows:

> We tend to assume that the knowledge [in which virtue consists] must have a stateable propositional content (perhaps not capable of immediate expression by the knower). Then the virtuous person's reliably right judgements as to what he should do, occasion by occasion, can be explained in terms of interaction between this universal knowledge and some appropriate piece of particular knowledge about the situation at hand; and the explanation can take the form of a "practical syllogism", with the content of the universal knowledge, or some suitable part of it, as major premise, the relevant particular knowledge as minor premise, and the judgement about what is to be done as deductive conclusion. (McDowell 1979, p. 57)

The above interpretation of the Socratic doctrine is intellectualist because it requires that the virtuous person's conception of how he should behave can be "codified" in such a form that it could play its supposed role of in the above mentioned practical syllogism. By any standards, such knowledge requires that the virtuous agent has considerable powers of reasoning. This I call *Socratic intellectualism.*

There are a number of reasons why Aristotle might have found Socratic intellectualism to be problematic. McDowell, for instance, suggests that it is difficult to see how the practical knowledge required for virtue could formulated:

---

[1] In this section I rely on interpretations of Aristotle's *Nicomachean Ethics* by Myles Burnyeat (1980), John McDowell (1979, 1995a) and John M. Cooper (1986) which I take to be compatible with one another. Where I refer to the primary text I employ Ross's revised (1984) translation. However, I substitute the term "virtue" for the term "excellence" (*areté*) throughout, consistent with the commentaries consulted. *Nicomachean Ethics* is abbreviated *NE* followed by book number, section number and Bekker reference.

> If one attempted to reduce one's conception of what virtue requires to a set of rules, then, however subtle and thoughtful one was in drawing up the code, cases would inevitably turn up in which a mechanical application of the rules would strike one as wrong - and not necessarily because one had changed one's mind; rather, one's mind on the matter was not susceptible of capture in any universal formula. (McDowell, 1979, p. 58)[2]

What is more, it seems that even if we could come up with such a universal formula for a certain virtue, there is a problem of accounting for how individuals come to learn it, exposed as they are to only a finite sample of experiences. There are other difficulties with Socratic intellectualism, but we need not pursue them here.[3]

Aristotle's solution was to offer an interpretation of the Socratic idea that virtue is knowledge which did not suffer from the problems of the intellectualist reading. Here is how McDowell presents the idea that the particular virtue of kindness can be thought of as "knowledge":

> A kind person has a reliable sensitivity to a certain sort of requirement that situations impose on behaviour. The deliverances of a reliable sensitivity are cases of knowledge; and there are idioms according to which the sensitivity itself can appropriately described as knowledge: a kind person knows what it is like to be confronted with a requirement of kindness. The sensitivity is, we might say, a sort of perceptual capacity. (McDowell 1979, p. 51)

Rather than understanding the possession of a virtue as having knowledge of some universal principle about how to act kindly in any given situation, that is a knowledge-*that* such-and-such is to be done, McDowell suggests that it is better understood as a "sensitivity", or a kind of knowledge-*how*. This latter is a disposition to act in appropriate ways, which need not be explained in terms of the apprehension

---

[2] In the paper quoted, McDowell goes on to use a Wittgensteinian argument against the idea of codifiability, an argument which we shall encounter in §6. Though the argument may support Aristotle's view, there is no suggestion that it was actually proposed by Aristotle.

[3] McDowell points out two. First, we can only explain a person's acting non-virtuously as "ignorance", which means that "paradoxically, failure to act as a virtuous person would cannot be voluntary, at least under that description" (1979, p. 54). Second (pp. 55-6) the intellectualist interpretation leaves us without a distinction between virtue and continence, a distinction which Aristotle seeks to preserve.

of a proposition. Assuming that this distinction is a good one, the suggestion is that if we interpret the Socratic thesis in the way McDowell suggests, we can overcome the difficulties with the intellectualist interpretation. For we can drop the idea that being virtuous depends upon the codifiability of a virtuous person's conception of how to act. Somebody can be said to have the knowledge-how - say, how to act kindly - without having implausibly to attribute to them the knowledge-that, in this case, that kindness consists in nothing more than knowledge of the proposition to the effect that they should do such-and-such things in so-and-so situations.

To conceive of virtue in this non-intellectualised way is not to give up on the idea of explaining the knowledge in which the particular virtue of kindness consists, because it is still possible to formulate "generalisations" about how kind people behave, for instance, "they help others in adversity". The mistake would be to think of such generalisations as attempts to reduce the particular virtue to universal principles. As Aristotle says, "the best generalisations about how one should behave hold only for the most part" (*NE* 1.3 1094b13-27; quoted by McDowell 1979, p. 58). Conceiving of virtue as knowledge-how also facilitates a plausible account of how people become virtuous: not by somehow grasping universal principles, but by practice at the right kinds of behaviour, which in turn equips them with the relevant capacities.

3. *Three Features of Virtues*

I want now to draw attention to three of the features of an Aristotelian virtue that, I shall later argue, the notion of a virtue shares with notions deployed by Ryle and Wittgenstein in their arguments against intellectualism. Virtues, I want to say, represent behaviours that are repeated, automatic and responsible.

First, it is not difficult to show that virtues are repeated. Here is Aristotle on the acquisition of virtues:

> Again, of all the things that come to us by nature we first acquire the potentiality and later exhibit the activity ...; but virtues we get by first exercising them, as also happens in the case of the arts as well. For the things we have to learn before we can do, we learn by doing, e.g. men become builders by building and lyre-players by playing the lyre; so too

> we become just by doing just acts, temperate by doing temperate acts, brave by doing brave acts. (*NE* 2.1 1103a26-b2) ...
>
> Thus, in one word, states arise out of like activities. This is why the activities we exhibit must be of a certain kind; it is because the states correspond to the differences between these. It makes no small difference, then, whether we form habits [*ethike*] of one kind or of another from our very youth; it makes a very great difference, or rather *all* the difference. (*NE* 2.1 1103b20-6)

For Aristotle, then, we acquire the "states" which are the virtues, in just the same way as we acquire other practical abilities, that is, by practice.[4] But it would be a mistake to think that for Aristotle, repetition is merely a way of acquiring virtues, which are from then on unchanging states of the agent. Rather, just as to remain a lyre-player one must continue to play the lyre, to remain virtuous, an agent must continue to be practise the virtues. There is no requirement in Aristotle that there is an end-point to the process.[5]

There may, however, seem to be a problem at the beginning. In the above passage Aristotle writes "virtues we get by first exercising them". But how can somebody exercise a capacity that she does not already possess? Certainly, if all the information about a person that we have is that they seem to have acted in the way a virtuous person would have done just once, there may not be a fact of the matter about whether this really is an exercise of virtue, and we may not know what to say. Perhaps we could justify our calling the action virtuous by referring to the similarity with the actions of other agents which we know to be virtuous. If we were to allow this, we would thereby concede that the particular individual herself need not have repeated the action for it to be an exercise of virtue, though others must have done. I am not sure if Aristotle would allow this. But even if he would, the issue only affects the very first apparent exercise of a virtue, whose status may anyhow best be thought

---

[4] Burnyeat (1980) reminds us that such remarks should be read not just as embodying his conception of virtue, but also as Aristotle's spelling out the prerequisites for his lectures, i.e. that the students should already have acquired the virtues.

of as indeterminate. First attempts aside, it seems that repetition can be established with reference only to the behaviour of a single individual. I shall follow this individualistic understanding of repetition from here on. Let us accept, then, that Aristotelian virtues require the repetition of virtuous actions, both to acquire them, and to maintain them.

In what sense are virtuous actions "automatic"? Here we can return to McDowell's suggestion described in Chapter 1, §4, that virtues are part of our "second nature". The thought is that the phrase not only captures the way our virtues are acquired naturally (a process consistent with action naturalism as described); but it also suggests something about the way in which virtues are exercised once they are acquired, and that is "naturally" in the sense that such actions come easily to the agent.

We use the term "second nature" to pick out this proficiency when we describe people who are particularly skilled at some activity. Pulling an oar is second nature to Steve Redgrave; running down the wing comes as second nature to Ryan Giggs. This chimes with Aristotle's analogy between the acquisition of virtues with the acquisition of "arts" such as building and lyre-playing. What marks out the expert builder or lyre-player from the novice is that they do it "naturally", which is to say effortlessly. This characteristic proficiency, one would like to say, is a result of the practice (though it may require other things - like natural talent - in addition). But whatever the relation to practice, we have here a sense in which virtuous actions can be said to be "automatic".

What marks this effortlessness? One suggestion would be that no deliberation is involved in acting virtuously. However, those who have only a rudimentary understanding of Aristotle may well balk at this suggestion. For isn't virtuous action supposed to be a matter of judging the mean between two vices, a process which must be deliberative? In support of this view consider a plausible reading of some passages in Aristotle, where virtues do seem to involve both reasoning and choice:

---

[5] See for example Sabina Lovibond's (1996) treatment of how the ongoing process of moral education works.

> Virtue, then, is a state concerned with choice, lying in a mean relative to us, this being determined by reason and in the way in which the man of practical wisdom would determine it. (*NE* 2.6 1106b36-1107a1-2)[6]

But I think it would be a mistake to conclude from this that for Aristotle virtuous action involves deliberation. John M. Cooper (1986) has suggested that this and other similar passages in Aristotle, need not be read as implying that any actual process of deliberation or choice goes on before an agent acts virtuously. Cooper finds the suggestion that deliberation and choice precede every moral action unsatisfactory because of the many common experiences we have of acting morally, without any apparent prior deliberation. Cooper writes,

> Deliberation is called for principally or only in difficult or delicate cases, where the facts are complicated, or their bearing on the interests or principles in question is not easy to assess, or where a very great deal hangs on the decision's being correct, and so on. Certainly, cases of routine action do not call for deliberation.[7] Furthermore, many overt decisions are reached without going through any sort of calculation (at least none that is explicit). (Cooper 1986, p. 6)

Rather confusingly, Cooper here uses the term "decision" to be applicable equally to those actions upon which we deliberate, and those for which - he wants to argue - we do not. So we should not read "decision" as entailing any intellectual endeavour.

Cooper then goes on to say how this interpretation need not be thought to be in tension with Aristotle's other views.

> Aristotle's insistence that moral decisions are all of them "choices", and therefore supported by deliberation, can be defended ... provided one understands by this only

---

[6] Another passage on virtuous acts has a similar implication: "The agent must be in a certain condition when he does them; in the first place he must have knowledge, secondly he must choose the acts, and choose them for their own sakes, and thirdly his action must proceed from a firm and unchangeable character" (*NE* 2.4 1105a30-2).

[7] Here Cooper references Aristotle: "in the case of exact and self-contained sciences there is no deliberation, e.g. about the letters of the alphabet (for we have no doubt how they should be written)" (*NE* 3.3 1112a34-b2).

> that moral decisions are always backed by reasons which, when made explicit, constitute a deliberative argument in favour of the decision. Insofar as those are one's reasons it is *as if* one had deliberated and decided accordingly (even if one actually did no deliberating at all). (Cooper 1986, pp. 9-10)

If this is a correct reading of Aristotle, then it follows that the exercise of a virtue need not be understood as involving deliberation.[8]

Of course, the fact that virtuous actions may not involve deliberation does not yet show that they always lack deliberation, or that they must. But I think we can rule out various sorts of deliberation nonetheless. For whilst it may be plausible to say that a virtuously acting agent might deliberate about some of the *details* of how to conduct herself in a particular case - and Cooper suggests the sorts of details that might detain her in the passage quoted earlier, such as complexity or importance - she will not presumably have to decide *whether* to exercise the virtue in question. Should that latter question arise, we would have reason to doubt that the agent has the virtue at all. Presumably too, in the sorts of context with which the agent is most familiar, that is, those which are similar to or identical with the contexts in which she has practised the virtue, we can expect no deliberation whatever to take place, simply because it will be unnecessary, and waste time.[9] So it seems plausible that many virtuous actions, perhaps including those which we would regard as paradigms, will, for good reason, lack any deliberation whatever, and in that sense can be said to be automatic. And nothing Aristotle says suggests that such actions would thereby forgo their status as virtuous, nor indeed, as I shall later suggest, as rational.

Finally we must ask in what sense agents are "responsible" for their exercises of virtue. Implicit here is a distinction between the kind of responsibility that persons

---

[8] In a recent book Rosalind Hursthouse (1999) argues for a similar point about the irrelevance of one particular kind of reflection in virtuous action. She claims that her argument entails "not only that occurrent thoughts somehow equivalent to 'This is right, virtuous, noble, my duty', or what have you, are not necessary for moral motivation; they are not sufficient either" (p. 160).

[9] I am indebted here to some of Christopher Hookway's remarks in his (2000), though they do not directly concern Aristotle, and in conversation.

have, as a result of which they are praise or blameworthy, and the kind of responsibility which we might attribute to a merely physical cause, such as the rainfall being responsible for the river's filling, for which attributions of praise or blame make no sense. Call them respectively *personal* and *causal* responsibility.[10] Are we personally responsible for virtuous actions? The materials for an answer are found in the following passage from Aristotle:

> For all men think that each type of character belongs to its possessors in some sense by nature; for from the very moment of birth we are just or fitted for self-control or brave or have the other moral qualities; but yet we seek something else as that which is good in the strict sense - we seek for the presence of such qualities in another way. For both children and brutes have the natural dispositions to these qualities, but without thought these are evidently hurtful. Only we seem to see this much, that, while one may be led astray by them, as a strong body which moves without sight may stumble badly because of its lack of sight, still, if a man once acquires such thought that makes a difference in action; and his state, while still like what it was, will then be virtue in the strict sense. (*NE* 6.13 1144b1-14).

Here Aristotle distinguishes between "natural virtue", which is possessed by children and non-human animals, and "virtue in the strict sense", which is the sense which we have been discussing. He also gives us an explanation of that distinction. So the courageous behaviour of a lioness defending her cubs is only regarded as a natural virtue because the lioness lacks "thought that makes a difference in action". I understand this as the capacity to deliberate upon, and act in the light of, reasons. By contrast, a rational agent courageously defending her children would count as exercising virtue in the strict sense, because she does have this rational capacity. This is why I think we can say that Aristotle did not think of the virtues as mere dispositions to respond to certain situations in certain ways, like reflexes, or animal behaviour, to which (at least arguably in the case of the lion) only causal

---

[10] Writers who discuss free will often refer to a distinction between causal and "moral" responsibility, see e.g. Fischer & Ravizza (1998, pp. 1-2). Since nothing for them turns on the nature of moral, as opposed to any other sort of responsible human activity, I take it they can be read as meaning the same thing by "moral" as I mean by "personal".

responsibility attaches, but as something more. That something more is explained in terms of what he shortly afterwards calls the "presence" of the capacity to act for reasons.[11] That this capacity should be present in virtuous actions is consistent with what I have said about their being automatic, since we need not read this in an intellectualist way, as the *operation* of the capacity to act for reasons. The capacity's being there, as it were, in the background, is enough. In the Chapter 4 I shall say more about how the presence of this rational capacity gives us the kind of control over our actions that we need to be personally responsible for them. For now, we can conclude that for Aristotle, agents are indeed personally responsible for virtuous actions.

In sum, whatever else it is, the notion of a virtue that Aristotle employs to undermine Socratic intellectualism, is a notion of a kind of behaviour which can be understood as being repeated, automatic and responsible. This is not to say that it is *only* by employing a concept with these three features that Socratic intellectualism could be avoided. But if Aristotle's project is successful, as many think it is, it owes much of its success to using these familiar features of human behaviour, and not problematic entities such as universal principles. Aristotle for one, then, saw the anti-intellectualist potential of the idea of a habit.

4. Ryle

A second philosopher who I think employs a notion akin to that of a habit to combat intellectualism is Gilbert Ryle in *The Concept of Mind* (1949).[12] In this work Ryle expresses his conviction that even though we often speak about mental phenomena as being "inside" people's heads, if we interpret such talk literally, we thereby commit ourselves to the "Cartesian" picture of the mind. According to this picture, there are two kinds of substance: that which is extended in space, and that which is not.

---

[11] "[Virtue] is not merely the state in accordance with right reason, but the state that implies the *presence* of right reason, that is virtue; and practical wisdom is right reason about such matters. Socrates, then, thought the virtues were forms of reason (for he thought they were, all of them, forms of knowledge [that]), while we think they *involve* reason." (*NE* 6.13, 1144b26-9)

[12] Page references in this section and the next are to this work unless otherwise stated.

"Mental" terms are supposed to refer to substances of the second kind, the existence of which we can know with certainty by introspection; whereas "physical" terms, which include those which refer not only to tables, chairs and the like, but also to the living human body, are supposed to refer to substances of the first kind, the existence of which we can coherently doubt. This is a clear case of intellectualism, since it is highly optimistic about reason's capacity to operate on itself.

According to Ryle, this dualistic picture brings with it a set of intractable philosophical problems (pp. 13-17). These include the problem of interaction: how can "the mental" be causally related to the "non-mental", as it seems, at least sometimes, to be? And then there is the epistemological problem of other minds: though it seems I can be fairly confident that other people have "bodies", how can I know that they also have "minds" like mine? Ryle's suggestion is that the only way of avoiding such difficulties is to drop the Cartesian picture of mind which they presuppose. The aim of his book is to show this picture to be an "intellectualist legend" (p. 30), or as he famously puts it, "the dogma of the Ghost in the Machine" (p. 17). As with other dogmas, Ryle thinks that we can get on much better without it.

It need hardly be said that the Cartesian intellectualism that Ryle attacks is not one which concerned Aristotle. As a matter of historical fact, in Aristotle's time that account of mind had yet to be conceived. What is more, the Cartesian ghost does not haunt only moral agents, but agents *per se*. Ryle's target is thus broader. In view of the differences in their aims, one would expect Ryle's rejection of intellectualism to differ from Aristotle's. But whilst there are differences, they are in some respects remarkably similar.

To exorcise the Cartesian ghost, Ryle attempts to show how many of our references to supposedly "inner" items can instead be understood as indirect references to overt behaviour. This is not the naïve and implausible suggestion that a reference to a mental item, such as my belief that the windows need painting, is just reference to my present behaviour. Obviously I can have the belief without doing anything at all at the present time. Rather, it is a reference to what my future behaviour will, or might, be, given various possible future circumstances. My belief that the windows need painting just is my being *disposed* to say that they need

painting when the topic comes up, my being disposed to buy paint and brushes in preparation for painting them, to begin painting them in fine weather, or perhaps being disposed to ring up decorators, direct them to the windows in question, and so on. Ryle's recommendation, then, is to understand mental terms as references not to what I am now doing, but to what I am disposed to do, that is, *behavioural dispositions.*[13]

Ryle agrees with the Cartesian premise that the mental is something "not present" in a particular piece of behaviour, but rejects the Cartesian conclusion that it must therefore be "present" in some other "place", such as in the Cartesian "inner theatre". Rather, Ryle thinks that the mental is "not present" in the sense that what somebody will (or might) do is "not present" in what they now do. The idea of a disposition is the idea of a set of possible and actual behaviours which extend beyond any particular instance of behaviour. Ryle writes "a disposition is a factor of the wrong logical type to be seen or unseen, recorded or unrecorded" (p. 33). That is not to deny the reality of the mental, as some have suggested, though it is to deny that looking at or behind a single example of present behaviour is our only access to the real.[14] If the mental is understood to be logically inseparable from behaviour thus construed, the intractable problems associated with dualism simply do not arise, or at least not in that form. This bold claim about the meaning of mental terms is often termed *logical behaviourism.*[15]

---

[13] In this section I follow Ryle in his liberal use of the term "disposition", to apply to both physical substances and objects as well as persons. I shall restrict the use of this term in subsequent chapters, because I think calling habits "dispositions" is liable to give an unwarranted sense that a reductive naturalism has been sought, and to obscure a number of the unique features of habits.

[14] This is to disagree with a recent reading of Ryle by Stephen Mumford (1998). According to Mumford, Ryle thinks that, "to ascribe a disposition is not to ascribe a property but merely to say how something will behave in certain circumstances. Dispositional properties are, on this view, logical fictions; what really exist are regularities of events" (p. 4). But nothing Ryle says prevents him from saying that how something will behave in certain circumstances *just is* to ascribe a property to it. Certainly, Ryle has no theory of properties which would prevent this.

[15] The "logical" part distinguishes it from other forms of behaviourism, such as the *methodological behaviourism* of J. B. Watson (1919) and B. F. Skinner (1953).

When we compare Ryle's opposition to intellectualism with Aristotle's, then, we find many differences. What of the similarities?

5. *Three Features of Behavioural Dispositions*

What Aristotle and Ryle have in common is that they both use notions of a behaviour which is at once repeated, automatic and responsible. In Ryle's case the notion is that of a behavioural disposition, which he wants us to understand as a variety of knowledge-how, as opposed to knowledge-that, something that I earlier (§2) said of Aristotelian virtue.[16] But even given that behavioural dispositions are knowledge-how, we cannot simply conclude that they will be subject to the same treatment as virtues. For within the class of behavioural dispositions Ryle distinguishes between "mere habits" and "intelligent capacities" (pp. 41-4), and he is dismissive about the former. He writes that "the common assumption that all second natures are mere habits obliterates distinctions which are of cardinal importance for the inquiries in which we are engaged" (p. 42).[17] This may signal problems not only for my interpretation of Ryle, but also for my broader claims about habits. But there is a way of understanding Ryle which leaves both projects intact. How, then, does Ryle draw the distinction?

"Mere habits" are distinct from intelligent capacities in four respects (pp. 42-3): (i) in the amount of "care, vigilance, or criticism" needed to exercise them - none is

---

[16] The claim that Ryle's distinction is a good one has recently been questioned by Stanley and Williamson (2001), who argue that "knowledge-how is simply a species of knowing that" (p. 411). However, not only do they rely on extremely weak (and thus contentious) criteria of propositional knowledge, and deploy the (to this reader) opaque idea of a "practical mode of presentation", I think their argument against Ryle's positive account (p. 416) may trade on an equivocation between the knowledge required to *describe* how to do something, and the practical skill of doing it. It is the latter interpretation, and not the former, that I take it that Ryle intends.

[17] I take the "common assumption" to be akin to that of the (influential at the time) methodological behaviourists such as Skinner and Watson (see n. 15). Ryle's thought is that a convincing case against the Cartesian picture can only be made by explicating the more "sophisticated" dispositions that Ryle terms "intelligent capacities", because it is generally only the exercise of these capacities that tempts us into Cartesian myths.

required for mere habits; (ii) in the way in which each performance affects future performances - exercising a mere habit has no such effect: "one performance is a replica of its predecessors"; (iii) in the way they are learned - mere habits are learned by "drill", that is, "the imposition of repetition", in contrast to "training", which involves both drill and "the stimulation by criticism and example of the pupil's own judgement"; and finally (iv) in the range of performances that can arise from one of them - mere habits are "single-track", whereas intelligent capacities are "heterogeneous".[18]

Ryle's distinction has come in for criticism.[19] However, whilst the distinction may have its problems, it is still possible to argue that both mere habits and intelligent capacities have our three features, so problems with the distinction do not count against my claim.

First, given Ryle's distinction it is easy to settle the question of whether behavioural dispositions are repeated. The acquisition of both mere habits and intelligent capacities involves drill. There is also evidence that Ryle thinks that dispositions of both kinds need to be practised continually to be correctly attributed, as was the case for the virtues.[20]

However, when we turn to the questions of whether the exercises of behavioural dispositions are automatic and whether they are responsible, things become much less straightforward. For Ryle mere habits certainly seem to be automatic. He writes:

---

[18] This last distinction originates in Aristotle, see Makin (2000).

[19] See for instance Nathan Brett (1981), who attack's Ryle's distinction in order to argue that there is no inconsistency between "rule following" behaviour and acting habitually.

[20] For instance, when discussing whether a marksman's bull's eye was a fluke or a good shot Ryle writes that, amongst other things, "we should take into account his subsequent shots" (p. 45). Later on, "When we say that someone acts in a certain way from sheer force of habit, part of what we have in mind is ... that in similar circumstances he always acts in just this way" (p. 106).

> When we describe someone as doing something from pure or blind habit, we mean that he does it automatically, and without having in mind what he is doing. He does not exercise care, vigilance, or criticism. (p. 42)

But then it seems to follow that when an agent manifests an intelligent capacity, he does exercise "care, vigilance, or criticism", and therefore does not act automatically. So intelligent capacities seem to provide a counter example to the claim that behavioural dispositions are automatic.

But I think this conclusion would be a mistake. In his criticism of Ryle's distinction, Nathan Brett (1981) gives an example of a dentist who washes his hands before each appointment (pp. 363-5). Whilst there is undoubtedly good reason for the dentist to do this, it has become a habit, so he does not deliberate about whether to do it each time. In this sense it is automatic. This activity also counts as an intelligent capacity because (i) hand washing requires care and vigilance (if not also criticism); (ii) in view of this care, it is open to the dentist to improve upon his habit, say, by being more thorough; (iii) the learning of the habit would not typically have involved mere drill, but instruction too; and (iv) there are no good grounds for thinking of hand-washing as "single-track", for there may be infinite small variations between successive exercises.[21] Brett's point, then, is that none of these features of intelligence are incompatible with the fact that the agent acts automatically. I am not sure whether this conclusion would be satisfactory to Ryle, but I speculate that it would not be fatal to his anti-Cartesian project. So both mere habits and intelligent capacities can be automatic.

Another difficulty arises when we ask if we have personal responsibility for manifesting both sorts of dispositions. Whilst Ryle would no doubt agree that we are personally responsible for our intelligent capacities, I doubt that he would say the same of mere habits. He writes, for instance:

---

[21] Brett writes "a close look at either the facts about human habits or the language we use in talking about them would lead us away from the view that these must be 'single-track' dispositions yielding responses which are 'exact replicas' of one another" (1981, p. 369).

> The practices are not learned until the pupil's responses to his cues are automatic, until he can "do them in his sleep", as it is revealingly put. (p. 42)

> An action done from pure habit is one that is not done on purpose and is one that the agent need not be able to report having done even immediately after having done it; his mind may have been on something else. (p. 127)

These quotes suggest that for Ryle the agent is not responsible for her merely habitual behaviours, but rather her body, through drill, becomes "programmed" to react in a certain way. The reaction is little more than a reflex.[22] On any sensible view we are not personally responsible for the manifestations of our reflexes.

However, when we look at the examples of mere habits that Ryle gives, it seems that we would say that we are responsible for their exercise, and so the assimilation with reflexes would be a mistake. Ryle's examples of mere habits include the recruit who learns to slope arms, and the child who learns the alphabet and multiplication tables (p. 42). Even though we might learn these things by drill, we do not thereby relinquish responsibility for their exercise. We can see this by reflecting on the fact that we have a certain amount of control over such behaviours. Specifically, it is always open to an agent who has been drilled in these things not to exercise a mere habit on a particular occasion. The rebellious recruit could defy the orders given to him; and there may be occasions on which sloping arms might actually cause harm (doing it in a confined space, for example). In such cases, the recruit would be culpable for the merely habitual behaviour because he should have been sensitive to other factors. This shows that a "merely" habitual behaviour is not an uncontrolled response, like a reflex, but is under the agent's control and could therefore count as something that the agent does. This thought will be developed in Chapter 4.

In sum, even if we accept Ryle's distinction between mere habits and intelligent capacities, they both represent the idea of behaviours that are repeated, automatic and responsible. In these respects at least, then, the idea of a behavioural

disposition which Ryle uses to combat intellectualism, can be thought of as kindred to Aristotle's idea of a virtue, which is used to an analogous end.

6. *Wittgenstein*

The third and final philosopher whose work I want to consider is the Wittgenstein of the *Philosophical Investigations*. Like Aristotle and Ryle, Wittgenstein had his own intellectualist ghosts to exorcise, and in doing so, he, like them, employed a number of notions which share the same three features.

One of Wittgenstein's aims in the *Investigations* is to explicate the phenomenon of linguistic meaning: to say what it is for a speaker to "mean" something by a word or sentence, or, as it is sometimes put, to use a word or sentence "with understanding".[23] Early on in the book Wittgenstein writes:

> For a *large* class of cases - though not for all - in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language. (*PI* §43)

It is not until later in the book that we see how this might be part of an anti-intellectualist solution to a problem that has hitherto been addressed in intellectualist ways.

---

[22] Ryle explicitly groups mere habits with reflexes, for instance, when he talks of them both as being "single-track dispositions" (p. 46).

[23] I say that Wittgenstein sought to "explicate" meaning in order to distinguish his aspirations from another twentieth century project of constructing a "theory" of meaning. A theory (or "analysis") of meaning explains what it is for any word or sentence to be meaningful in terms which do not themselves draw on the idea of meaning. Examples of such theories include those which understand meaning in terms of "truth conditions" (e.g. Davidson), or "assertability conditions" (e.g. Dummett). Wittgenstein notoriously did not offer theoretical accounts of this kind for any concept, (ironically) precisely because of what he took meaning to be. According to his own conception of meaning as "use", he thought that the proper approach to giving an account of any concept is to examine the actual uses to which the term expressive of that concept is generally put, and to describe those uses as accurately as possible, that is, to *explicate* it. Such explication displays the concept's meaning, and is prior to any theory construction which is thereby shown to be an optional exercise at best. The concept of linguistic meaning itself is no exception to this approach.

The passages which present the intellectualist conception of meaning to be rejected are found within the portion of the Investigations known as "the rule following considerations" (*PI* §§138-242). Here, Wittgenstein identifies a feature of all meaningful word or sentence use. Namely, all meaningful word or sentence uses are not merely subject to "factual" judgements of the form "A uttered such-and-such". In that respect they resemble any happenings, irrespective of meaningfulness. Rather, in addition to being subject to such factual judgements, in virtue of being meaningful, these uses are also subject to judgements to the effect that "such-and-such" is in some sense an "appropriate" or "correct" thing to say in the circumstances; as we might call them, "normative" judgements. We can only make sense of the use of a word being a "mistake" against a background of there being a "correct" way to use it. Wittgenstein notices that in respect of having this normative feature, meaningful word or sentence uses are akin to many other performances, all of which we might call instances of "following a rule". Wittgenstein's example of a non-linguistic rule is that of continuing a number series (*PI* §143). Thus for Wittgenstein, in respect of the normativity of rules, the investigation into linguistic meaning expands into the more general investigation of what it is to follow *any* rule.

Wittgenstein illustrates the normativity of mathematical rules with his famous example of the pupil who applies the rule "+2" in the usual way for numbers up to 1000 and then writes down "1000, 1004, 1008, 1012" (*PI* §185). We can certainly conceive of individuals who might apply the rule in this way, but Wittgenstein is concerned to discover in what the correctness we find to be lacking consists. This then is the problem of rule following.

It is at this point that Wittgenstein considers an account of rule following which deserves the name "intellectualist". This is the idea that understanding a rule consists in having some "interpretation" of the rule which determines how the rule is to be properly applied. So having such an interpretation is the agent having some item of "knowledge" which discloses to the agent how the rule should actually be applied on any future occasion. According to this view, the item of knowledge features in the explanation of any instance of the rule's correct application (§§186-8). Such knowledge might be the kind of thing that comes to us "in a flash" when we

study a series of numbers and suddenly see how to go on (§§191, 197).[24] On this view, the distinction between correct and mistaken applications of a rule consists in whether the performance is appropriately "guided" by the supposed knowledge, or not.

This intellectualism about rule following has clear affinities with the intellectualist interpretation of the Socratic thesis that Aristotle attacked, discussed in §2.[25] Both the intellectualist notion of virtue, and the intellectualist account of rule following suppose that what explains these capacities is knowledge of some "universal principle". But like Aristotle before him, Wittgenstein did not think that such notions were required to make sense of the phenomenon in question.

> "But how can a rule show me what I have to do at this point? Whatever I do is, on some interpretation, in accord with the rule?" - That is not what we ought to say, but rather: any interpretation still hangs in the air along with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meaning.
>
> "Then can whatever I do be brought into accord with the rule?" - Let me ask this: what has the expression of a rule - say a sign-post - got to do with my actions? What sort of connexion is there here? - Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it.
>
> But that is only to give a causal connexion; to tell how it has come about that we now go by a sign-post; not by what this going-by-the-sign really consists in. On the contrary; I have further indicated that a person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom (*PI* §198).

I follow McDowell (1984) in thinking that the above passage contains an outline of Wittgenstein's response to intellectualism. But rather than justifying that reading of

---

[24] Another metaphor for the same idea is that of the rule as a "super-rigid machine" which determines all the steps (§§193-4). See also Wittgenstein's discussion of the "cube" (§140).

[25] McDowell (1979) implicitly connects these two kinds of intellectualism because he employs Wittgenstein's argument (below) to combat Socratic intellectualism (see n. 2 above).

Wittgenstein, I want simply to defer to McDowell on this matter, whom I read as follows.[26]

Wittgenstein's argument against the intellectualist account of rule following has two parts: one negative; the other positive. The negative argument rejects two ideas. First, if having an interpretation of the rule were all there was to following it, then we generate a vicious regress. For even if we concede that such an interpretation were available, there would still be a question of how the rule follower should follow the interpretation. If the intellectualist were to invoke his account of rule following at this point, then any such guidance must in turn be determined by a further interpretation, whose correct application must in turn be guided by some further interpretation, and so on *ad infinitum*. The regress is vicious because the question of how the rule follower should act is never settled. Hence the capacity to follow a rule cannot be what the intellectualist takes it to be.

In rejecting the intellectualist account, however, Wittgenstein wanted to avoid a second, and equally unattractive account. Here is how McDowell puts things:

> Wittgenstein's problem is to steer a course between a Scylla and a Charybdis. Scylla is the idea that understanding is always an interpretation ... Charybdis [is] the picture of a basic level at which there are no norms. (McDowell 1984, p. 242)

We have already seen both the attractiveness and the perils of Scylla. What then of Charybdis, "the picture of a basic level at which there are no norms"?[27] This idea is that following a rule is, at its most basic, a mere mechanical response to some stimulus. This view might seem attractive to anybody with naturalistic sympathies, because it dispenses with spurious items of knowledge, and instead promises to explain our capacity to follow rules with reference to nothing more than the dispositions of physical phenomena to behave according to causal laws. But this kind of reduction creates a problem, because at the "basic level" we no longer have any

---

[26] In particular this is to follow McDowell in his rejection of Kripke's (1982) "sceptical" reading of the rule following passages.

[27] This is a position which, according to McDowell, Crispin Wright (1980) mistakenly occupies. One of the purposes of McDowell's paper is to reject Wright's interpretation.

grounds for saying that a given response is "correct" or "incorrect". To borrow McDowell's example (pp. 234-5), if we insist on understanding an utterance of "This is yellow" as a mechanical response to a yellow flower, we thereby forfeit the right to say that the utterance was "meaningful". We have not explained the normativity of meaning, but denied it. Hence the reduction of rule following to the "basic level" cannot succeed.

Having considered and rejected two wrong accounts of rule following, Wittgenstein makes his positive proposal, that it is possible to steer a course between Scylla and Charybdis. McDowell asks:

> How can a performance both be nothing but a "blind" reaction to a situation, not an attempt to act on an interpretation (avoiding Scylla); and be a case of going by a rule (avoiding Charybdis)? The answer is: by belonging to a custom, practice, or institution. (McDowell, 1984, p. 242).

The idea is that when we understand an instance of rule following as part of a "custom", "practice" or "institution", this gives us a way of understanding a performance as subject to normative judgements, whilst avoiding both Scylla and Charybdis. By understanding a putative instance of rule following as part of a "custom" (practice, institution), we identify a sense in which that performance can be judged "correct" or "incorrect", that is, by making reference to the way the rule is normally applied within that "custom" (practice, institution). This is not to say that reference to custom delivers certainty about the status of a given performance, because there is still contingency in any judgement of what counts as "more of the same". But it is to say that if we are to look anywhere to understand the sense in which we count one performance as "correct", another as "incorrect", customs are the place to look. The solution avoids Scylla because the notions of "custom" (practice, institution) are not conceived of as "inner" states of agents, but, on the contrary, express what is plain for us all to see in our everyday ways of doing things. And we avoid Charybdis by refusing to identify a customary behaviour with a mere causal

response. As Wittgenstein writes "To use a word without a justification does not mean to use it without right" (*PI* §290).[28]

This in very broad outline is what I take to be Wittgenstein's solution to the problem of rule following. The implications are extremely wide ranging, not just for philosophical accounts of meaning and normativity, but also for our conception of the kinds of thing that philosophy itself can deliver. The question for now, however, is this: do the notions that Wittgenstein uses in his solution to the problem of rule following have the three features of being repeated, automatic and responsible? To answer this question we shall need to find our way through a controversy which has sprung up around Wittgenstein's comments on rule following.

7. The Spurious Role of the Community

There is much disagreement about how the notions of "custom", "practice" and "institution" should be understood. Specifically, there is controversy over how these ideas are related to the idea of a "community". Some claim that Wittgenstein's central point is that they are. If this is indeed his point, then the claim that Wittgenstein's anti-intellectualism depends upon employing notions which share the same features as Aristotle and Ryle's projects, could be weakened, since we have had no need to refer to anybody but the individual agent in characterising their solutions. We must therefore consider this debate if we are to argue convincingly that Wittgenstein's anti-intellectualism is of a piece with theirs.

In the 1980s and 1990s many philosophers were polarised on the issue of whether Wittgenstein uses the notions of "custom", "practice" and "institution" in order to draw our attention to the fact that the normativity inherent in rule following in general and meaning in particular is essentially social.[29] One group of philosophers, who became known as communitarians, took Wittgenstein's use of

---

[28] Although this passage lies outside the portion identified as the "rule following considerations", I follow McDowell (1984, p. 241) in thinking it to be relevant here.

[29] Though the debate was initiated in an exchange between A. J. Ayer (1954) and Rush Rhees (1954), it was given new life by Kripke's (1982) reading of Wittgenstein.

terms like "custom" to mean "social custom".[30] They tried to argue that the point of Wittgenstein's rule following passages was to advance a theory according to which rule following is "essentially social", and hence the idea of a "solitary" rule follower made no sense (though there is disagreement upon what being "solitary" could amount to).[31]

Communitarianism has been rejected by individualists, who deny that Wittgenstein meant to implicate the community, and argue that the idea of "solitary" rule following makes perfectly good sense.[32] For instance, Baker and Hacker write:

> The pivotal point in Wittgenstein's remarks on following rules is that a rule is *internally* related to acts which accord with it. The rule and nothing but the rule determines what is correct. This idea is incompatible with defining "correct" in terms of what is normal or standard practice in a community. (Baker & Hacker, 1986, pp. 171-2).

Baker and Hacker here propose that what is objectionable about the communitarian interpretation is that thinking of correctness in terms of communal practice is "incompatible" with another, better, definition of correctness. However, I do not think that we need to read Wittgenstein as advancing a positive claim like the one suggested here in order to oppose the communitarian view.

One source of opposition to communitarianism comes from the rule following passages themselves. Although McDowell is right that Wittgenstein uses the terms "custom" (*PI* §198), "practice" (*PI* §202) and "institution" (*PI* §199) to give his account of rule following, Wittgenstein also uses the terms "use" (*PI* §199) and "technique" (*PI*

---

[30] Communitarians include Rush Rhees (1954), Saul Kripke (1982), Norman Malcolm (1986, 1989), Robert Fogelin (1987), Meredith Williams (1991), T. S. Champlin (1992) and David Bloor (1997).

[31] All agree that "solitary" does not here mean mere physical isolation. Aside from this, interpretations differ. Kripke thinks that an individual "considered in isolation" (1982, p. 110) could not be said to follow rules. Malcolm thinks that an individual's lack of teaching (1986, p. 159), or her not being subject to other people's corrections (1986, p. 175) are the marks of solitariness. Implicit here are views about what it is to be a member of a community.

[32] Individualists include Baker & Hacker (1984, 1985, 1990), Simon Blackburn (1984b) and Colin McGinn (1984).

§199). Of the five terms, these latter two do not obviously imply any connection to a community, but are more naturally thought of as applicable to individuals.[33] However, this textual argument is not decisive.

There are two further difficulties for the communitarian. First that the communitarian has yet to say exactly how the community is implicated in rule following. As McDowell puts it,

> if regularities in the verbal behaviour of an isolated individual, described in norm-free terms, do not add up to meaning, it is quite obscure how it could somehow make all the difference if there are several individuals with matching regularities. (1984, pp. 252-3)

In the absence of such an articulation, the communitarian claim is weakened.[34] Unfortunately this leaves us without an interpretation of Wittgenstein's use of apparently communal notions such as "custom", and leaves the individualist view looking more plausible than perhaps it should.

A second objection affects both communitarians and individualists alike. The point is simple. It is a mistake to read Wittgenstein as offering any general account of what the correctness of rule following comes to. This is supported by the fact that Wittgenstein uses the range of terms that he does to explicate the phenomenon, and is consistent with a broader view of how Wittgenstein thinks philosophy should be done.

---

[33] It is perhaps because McDowell underplays such terms that Blackburn (1984b) reads McDowell as advancing a version of communitarianism. In favour of the attribution McDowell writes, for instance, "it cannot be denied that the insistence on publicity in Kripke's reading corresponds broadly with a Wittgensteinian thought" (1984, p. 243). However, in a later paper McDowell goes on to distance himself from any "constructive" account of how meaning is possible, which any version of communitarianism would represent: "I now think [the paper "Wittgenstein on Following a Rule"] is too hospitable to this kind of reading" (1993, p. 275, n. 6). McDowell rejects Blackburn's attribution explicitly at 1984, p. 253, n. 49.

[34] This would include the notions of community membership implicit in Kripke and Malcolm; see n. 31 above.

Wittgenstein is well known for stressing the importance for our understanding of concepts of what he calls "family resemblance" terms (*PI* §66-77). His examples of such terms include "game", "good" and "number". For Wittgenstein, it would be a mistake to try to identify the common features of all things to which we apply such terms, and think of these features as representing the "essence" of the notion, or the necessary and sufficient conditions for the term's correct application. Even if we could identify a set of common features, this would not tell us what the term in question "really means". It is plausible that terms such as "rule", and hence "rule follower" would fall into the same category.[35] Therefore, in view of his general opposition to the idea that the primary role of philosophy is to advance "theories" (*PI* §109) or "theses" (*PI* §128), the idea that Wittgenstein is making any general claim about the correctness of rule following, whether individualist or communitarian, requires defence.

If this interpretation is right, Wittgenstein is not committed to any view about the normativity of particular rules. For any given rule, it is quite open to Wittgenstein to say that the criteria for its correct application are "communal" or "individual" in some sense. The sense, or senses, of correctness associated with any given rule may be attributed according to communal or individual standards, or indeed both. We need to look in detail at the ways in which the rule is, or is not, related to other people, by looking at all aspects of how the rule features in the life of the rule follower. This will include considering how the rule is learned: is it taught, picked up by copying, or is it in some sense self-taught? We need to consider whether following the rule essentially requires the presence of others, as in the rules of conversation, or those of team games. And we should take into account whether others have a role in correcting misapplications of the rule, or whether mistakes will be picked up by the

---

[35] Wittgenstein mentions a diverse range of "rules": not just mathematical rules like "+2" (*PI* §185), and "rules" of word use (*PI* §80-2), but also sign-posts (*PI* §198); the rules of chess (*PI* §197); rules which work like natural laws which are inferred from watching others follow them (*PI* §54); rules which we make up as we go along (*PI* §83); and rules which we alter as we go along (*PI* §83).

rule follower herself.[36] These are all facts about particular rules that can be determined empirically. The mistake is to think that such facts are available for rules in general.

If this is right, it is reasonable to conclude that Wittgenstein's employment of five different terms to explicate rule following, some of which suggest links to a community, others which do not, is no oversight. He is rather drawing our attention to the diversity of examples of rule following. It follows that the suggestion that Wittgenstein is making a general claim about the community's essential involvement or non-involvement in rule following is incorrect.

8. Wittgenstein's Actual Solution and Three of its Features

Nevertheless, I think that there is one general point that Wittgenstein has to make about rule following, though in comparison with the issue considered above it has received little attention.[37] The general point is not that rule following is essentially connected with the community, or indeed essentially unconnected with it. It is rather that rule following is an activity that an individual learns, and subsequently goes on to perform. Whether the activity is learned from others, executed amongst others, or corrected by others, as we would expect when the activities in question are "customs" or "institutions", are factors which we can reasonably expect to vary from activity to activity. And others may not be involved in the acquisition of rules at all. What Wittgenstein suggests is central to all rule following activities, whether "techniques", "customs", "uses", "institutions" or "practices", is that they are all *learned reactions.*[38] And it is in this respect that rule following action has a close affinity with the notions employed by Aristotle and Ryle.

---

[36] It is no coincidence that these roles for the community correspond to those suggested by communitarians such as Malcolm (see n. 31 above).

[37] A notable exception is Meredith Williams (1994).

[38] We might speculate on the reasons for the neglect of this aspect of Wittgenstein's solution. Perhaps it is symptomatic of the same prevailing philosophical blindness that leads philosophers to neglect habits themselves.

Wittgenstein repeatedly emphasises the importance of "learning", "teaching" and "training". Here are some examples from the rule following sections themselves:

> How does [the pupil] get to understand this notation? - First of all the series of numbers will be written down for him and he will be required to copy them ... And here already there is a normal and an abnormal learner's reaction. (*PI* §143)

> What is the criterion for the way the formula is meant? It is, for example, the kind of way we always use it, the way we are taught to use it. (*PI* §190)

> "Then can whatever I do be brought into accord with the rule?" - Let me ask this: what has the expression of a rule - say a sign-post - got to do with my actions? What sort of connexion is there here? - Well, perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it. (*PI* §198)[39]

In this last remark, recall, Wittgenstein is offering the idea of a learned reaction as an alternative to the intellectualist picture that McDowell calls Scylla. If this is right as a reading of Wittgenstein, and there is plenty of support for this view from Meredith Williams (1994), we can now ask how Wittgenstein's idea of a learned reaction can be seen to be the idea of a behaviour which is repeated, automatic and responsible.

Wittgenstein is clear about the role of repetition. He writes "It is not possible that there should have been only one occasion on which someone obeyed a rule" (*PI* §199). So there is no difficulty with the status of the initial attempt at rule following, as there may have been for Aristotle's virtues (see §3). The repetition in learned reactions consists in the individual's past performances of the behaviour in question, and particularly those which were part of the initial training.

Is the exercise of a learned reaction automatic in the sense that it is marked by the absence of deliberation? We saw above that on McDowell's reading the whole point of Wittgenstein's solution to the rule following problem is that it is an alternative to a picture according to which "inner" items are supposed to "guide" the rule follower. This generated McDowell's Scylla. So if we understand "deliberation" as the consultation and weighing up of "inner" representations which might speak for

---

[39] See also *PI* §206.

or against acting, Wittgenstein's regress applies equally to this (*PI* §211). For Wittgenstein, following a rule cannot merely be an intellectual matter, rather it must be practical. That is not to say that an agent could not deliberate before following a rule on some particular occasion. It is rather to say that the question of whether or not deliberation takes place is independent from the question of whether or not she actually follows the rule. It is no coincidence that all five of the terms that Wittgenstein uses in his account of rule following are practical notions, i.e. they refer to things that are overtly done, not things that are merely thought. The idea of learning can be seen as a replacement for deliberation because it shows how the kind of reaction in question is not accidental, but a reaction which is constrained by, and naturally follows from, the learning process.

Are agents personally responsible for their learned reactions? Wittgenstein was not directly concerned with this question, but there are a number of points in the text that suggest not only that Wittgenstein does not want to demote the notion of rule following to the status of a mere mechanical response (avoiding Charybdis), but also wants to say that the agent is implicated in exercises of rule following. For instance, Wittgenstein asks "what has the expression of a rule got to do with my actions?" (*PI* §198), to which he answers, "I have been trained to react to this sign in a particular way, and now I do so react to it". Wittgenstein's use of the term "action", as well as the personal pronouns "my" and "I" suggest that rule following behaviour can, for Wittgenstein, count as nothing less than an expression of agency.

This is not conclusive. For all of the notions that Wittgenstein employs, we would be wise to express reservations about the individual agent's responsibility for exercising them. A technique can be picked up unintentionally, so the individual may be thought not to be wholly responsible for its exercise. And as we saw in §5, when a technique is acquired from others, as when a child copies her parents, we have strong intuitions that the child is not personally responsible when she later practises that technique. These issues are complex and will be taken up in Chapter 4. For now we can rely on the *prima facie* case for thinking that Wittgenstein's notion of a learned response is something for which agents are personally responsible. We can conclude that Wittgenstein too cures intellectualism with the idea of a habit.

It is worth mentioning that the view that rule following is a sort of habitual behaviour is not entirely new. Philip Pettit (1993) has proposed a version of this idea in what he calls his "ethocentric" account of rule following (pp. 76-108), though to my mind Pettit's view of these things has other weaknesses.[40] Others have explicitly rejected the suggestion. Peter Winch, for example, (1958, pp. 57-65) denies that rule following behaviour is habitual on the grounds that a pupil "has to learn not merely to do things in the same way as his teacher, but also *what counts* as the same way" (p. 59). I think Winch makes two mistakes. First, I think his reading of Wittgenstein is questionable. And second, I think he, like Ryle, is working with an impoverished notion of habit.[41] My suspicion is that insofar as people reject my view of rule following, they are likely to be making one or both of these mistakes.

9. *The Curative Power of Habits*

In this chapter I have argued that Aristotle, Ryle and Wittgenstein have all tried to offer remedies for distinctive strains of intellectualism. These are respectively the intellectualist interpretation of the Socratic thesis, the Cartesian picture of the mind, and the idea that rule following in general and meaning in particular is always an interpretation. Though each thinker proposes distinctive practical notions to effect their cure - respectively a virtue, a behavioural disposition, and a learned response - these notions have much in common. All three represent kinds of behaviours which are repeated, in that they are all learned by repetition; automatic in the sense that they do not involve a process of deliberation; and they are all behaviours for which the agent is personally responsible. It is in virtue of sharing these features that we can term all of these three notions "habits", in a sense of that term which is familiar enough from its ordinary use.

If we understand habits, then, as all and only those behaviours which are repeated, automatic and responsible in the senses described, we can see the idea of a

---

[40] In Chapter 3, §7 I take issue with the conception of habits Pettit uses in his account.

[41] Here I follow Brett (1981) in rejecting the presupposition embodied in Winch's claim, that we cannot know "what counts" as the same way when we exercise a habit.

habit as a notion which is both simpler than, and broader than, those employed by the three thinkers surveyed. Unlike Aristotelian virtues, habits need not be associated with specifically moral behaviour; unlike some Rylean dispositions, habits need not be thought of as particularly "intelligent" or "unintelligent"; and unlike some Wittgensteinian learned reactions, habits need not be thought of as essentially connected to a community. But Aristotelian virtues, behavioural dispositions and learned reactions are all kinds within the class of habits. The short history presented in this chapter can therefore be understood as the judicious deployment of kinds of habit in arguments against intellectualism.

This interpretation of the history suggests two profitable lines of thought. First, and most straightforwardly, habits, or some varieties of them, might be employed to cure other intellectualisms. That may well be a fruitful exercise in a number of areas of philosophy, and I shall make some suggestions in that respect in Chapters 7 and 8.

But there is a second implication specifically for the philosophy of action. For if we ask why habits have this power to combat intellectualism, a plausible explanations is this. For the historical anti-intellectualisms to have succeeded, habits must not themselves be vulnerable to intellectualist interpretations, but must rather represent a class of human behaviour which is not beset by intellectualist problems. Whilst Aristotle, Ryle and Wittgenstein all noticed the possibility of this kind of behaviour, habits, as I noted in Chapter 1, have no distinctive place in contemporary philosophy of action. Not only does this invite an appropriate correction to current accounts of action, but if the first suggestion made above is to be pursued, such a correction will also be important outside that field. For an argument which relies upon an unsubstantiated and controversial claim about actions is less likely to persuade than one that does not. The prior task, then, is to substantiate the claim that habits have a proper and distinctive place in the philosophy of action. That is the job of the next four chapters.

3

HABITUAL BEHAVIOUR

1. Defining Habits

In Chapter 2 we saw that Aristotle, Ryle and Wittgenstein all employed varieties of habit in order to combat intellectualist trends of their times. In that chapter we worked with a simple definition of a habit, that is, as a behaviour which is repeated, automatic and responsible. Whilst some progress was made in that chapter towards saying how these terms might be understood, the primary task there was exegetical. We can now turn our attention to the job of describing in detail what it means to say that a given piece of behaviour is "habitual", by explicating each of these three features. The result will be a more precise definition for use later in the thesis. Even though I shall make some stipulations in giving this definition, the hope is that it will still capture a large portion of the behaviours that we would naturally call "habitual" in our everyday discourse.

In this chapter I shall deal with the first of the two features of habits - repetition (§§2-4) and automatism (§§5-7). These two features alone give us an understanding of what it means to say that a habit is learned, which I shall also spell out (§8). This will yield a definition of a habitual *behaviour*, though not yet a definition of habitual *action*. In the next chapter we shall turn our attention to the question of how such behaviour can be regarded as responsible, and hence as action proper.

It may seem that there are two jobs here - one of defining a habit and another of defining a particular piece of habitual behaviour - and we only need to do the second. But the two jobs are interdependent. We cannot understand what a habit is without understanding its manifestations; and we cannot understand what a habitual behaviour is without appreciating that it is just one manifestation of a habit. So whilst in this chapter the primary goal is to define particular habitual behaviours, largely because we shall later be engaging with debates about particular actions, we shall

implicitly be developing a corresponding notion of habit, of which those behaviours are exercises.

A note on some of the terminology used in this chapter. First, when I use the term "agent" I shall mean the kind of being that it makes sense to hold personally responsible for what she does. Healthy, adult humans are agents in this sense. Very young children, people in comas, non-human animals, plants, corkscrews and stones are not. Second, I mean the term "behaviour" to refer to happenings which involve the agent's body (though may extend beyond it, as for example when we use equipment). In particular, I mean the term "behaviour" to be neutral between actions, roughly, things agents do, and what we might call "mere behaviour", things that happen to agents. Because I shall not argue that habitual behaviours are actions until the next chapter, I shall not in this chapter refer to them as actions. Lastly, although I shall not be proposing a definition of the term "habit" until the end of the next chapter, I shall use the term frequently in the meantime. When I do so I mean to pick out an agent's temporally extended pattern of behaviours which conform to our working definition of a habitual action, by being repeated, automatic and responsible.

I should also comment on the fact that I do not draw on the writings of others very much in this chapter. There is a good explanation for this. Whilst writers such as Peters (1963), Hamm (1975) and Brett (1981) have noticed the importance of habits, and specified them as having roughly the three features I have identified, they have not gone into the details of what each of these features amounts to. I seek here a detailed account, which may have implications for their claims.

2. Repetition

Habitual behaviours are repeated. We have seen that for Aristotle, Ryle and Wittgenstein, repetition is important because it is the means by which doing a given sort of thing becomes automatic. However, repetition of a behaviour is not merely a means to acquiring a habit (whatever that may mean); it is part of what it is to have a habit. If somebody said of a given piece of behaviour that it is habitual, not knowing whether the agent had done nothing like it before, their competence in using the term would be called into question. Repetition, then, is built in to the concept of a habit. In

this and the next two sections I want to clarify this idea of repetition, its limits, and how particular repetitions might be specified. I shall return to the question of what it means to say that a habit has been "acquired" by repetition in §7.

To say, as I almost did above, that for a piece of behaviour to be habitual, something like it must have been done before, and is in that sense repeated, is both too vague and too liberal. It is too vague because it is quite unclear what "something like it" determines; and too liberal, because unless this is specified, it does not seem to rule out many behaviours at all. If we look more closely, we find that we can be more specific about the sort of repetition involved in habitual behaviour.

The following are all examples of what I want to call *repeated* behaviours (though they may not also be habitual):

a) Tom goes to the pub this evening, and has been to the pub on previous evenings this week.

b) Pamela is cleaning the floor, as she often does when she's nervous.

c) My blinking.

d) Paul goes sculling on Wednesday, as he does most Wednesdays.

e) My biting my nails when I'm thinking.

f) My regular attendance at a Durham night club on Mondays.

The first thing to notice about the repetition in these behaviours is that they have been repeated by the individual agents themselves. We saw in Chapter 2 (§3) that there is a possible case for saying that a piece of behaviour could be regarded as repeated on the sole grounds that people other than the agent had done this sort of thing before, and indeed that fact might explain why the agent does it herself. (In particular we saw how this might be a tempting move if we wanted to count the very first execution of an apparently virtuous behaviour as virtuous). But whilst we might be able to make some sense of the idea of a habit whose repetitions are spread across individuals in this way, I think that would not be our usual way of understanding the notion. Intuitively, if we know that an individual has never performed a given sort of behaviour before, we will not say that it is *her* habit, whatever the explanation for her

doing this now might be. It is habits as they pertain to individuals, and their particular histories, in which we are interested.

The number of repetitions required for an individual to be said to have a particular habit can vary enormously. There is no upper limit to the number of repetitions required for habitual behaviour. The more times a behaviour has occurred, the more confident we get that it is repeated, and that in turn reinforces our sense that the individual has the habit. We say that the habit becomes "entrenched". There is, however, a lower limit on the number of repetitions.

For a behaviour to be repeated at all we can say for sure that they must have done something like it at least once before. Without at least two behaviours to consider, the present one and at least one previous one, the idea of repetition has no application. But two occasions may not be enough. We can see this by looking at how our everyday talk about repetition tends to go. For we would seldom say of a kind of behaviour that it happened "repeatedly" unless it occurred considerably more than twice. I may have gone to a particular night-club in Durham twice, but it would be misleading for me to say that I had gone "repeatedly", which suggests rather more past attendances. Of course, the second occurrence might rightly be termed a "repetition" of the first, but it doesn't follow from this that the term "repeatedly" applies to this kind of behaviour without strain. Similarly with "habitually". These latter terms would be more at home had I been to the night-club on, say, a dozen occasions, and perhaps with a certain regularity (of which more shortly). We could of course resolve these uncertainties by stipulating a lower limit on repetition, say at two, or some other small number. But in view of what is to follow, there is no need. For when we come to consider the automatic feature of habits in §5, we have there a criterion for what counts as sufficient repetitions for habit attribution. At this stage, we can just say that for a behaviour to be repeated it must have occurred at least once before.

3. *Specifying Repetition*

To say of a behaviour that it is repeated is to say that it has something - perhaps many things - in common with previous behaviours, of which, we have established, there is

at least one. But that will not suffice to specify a repeated behaviour. For how we describe the behaviour is crucial to making any commonality across behaviours visible. How does this work?

Consider one of the examples from the previous section. When we describe what Paul is doing as "going sculling on Wednesday", we are in a position to say that he has engaged in behaviour which fits that same description, before. If we had described what he is doing in other ways, the repetition could be obscured. So we might describe what Paul is doing as "going sculling in the afternoon", but as a matter of fact, Paul has never actually sculled in the afternoon before, so he is not repeating anything under that description.

Because repetitions can be described in many different ways the links between particular repetitions may otherwise be quite tenuous. So if I say that Paul won at rowing on Saturday, and won a game of chess on Sunday, one could say that Paul is doing a similar thing, namely, winning repeatedly. Had these activities been described in any other way it is unlikely that any kind of similarity could have been discerned, and hence there would be no grounds for saying that he was repeatedly doing anything. So if we say of any behaviour that is repeated, we imply that there is some description which would connect it to past behaviours.

This means that in describing a behaviour as repeated, we don't merely draw attention to common features. We also specify which aspects of the agent's past are to count as the behaviour's ancestors. So in saying that Paul's sculling on Wednesday is something he does on most Wednesdays, a whole tract of Paul's earlier career becomes salient. If I say that Paul sculled very well, as he usually does, another aspect of Paul's past comes into view.

The importance of description in defining repetition does not imply that the agent herself has any grasp whatsoever of what that description might be. Identifying similarities between behaviours can be done from a purely third-personal perspective. And I think the same goes for habit attribution. One need not have any idea that one has a given habit, for it to be true that one does. For that reason I do not think we can specify any limit on the kinds of similarities that should be permitted in specifying repetition.

Even given a specification of a repetition, however, the notion can tolerate exceptions within various limits. So for example, if I go to the Durham night-club every Monday night for five weeks, and then miss a Monday, before resuming my routine once more, we would probably want to say that I had repeated this behaviour sufficiently often for "going to the night-club on Mondays" to be regarded as a habit of mine. If the number of exceptions became too great, however, we would no longer say that I repeated this enough for it to count as habitual. So if I went clubbing for less than half the available Mondays, it would seem wrong to say that I had this particular habit. Given this different history, for us to be able to say that I have a habit of clubbing at all, it would have to be described in a different way.

In our natural understanding of habit, there seem also to be more nuanced requirements, not just concerning the proportion of times that I do it, but also concerning when the particular historical repetitions take place. There seems, for instance, to be a requirement for a certain sort of regularity in the behaviour, which rules out what we might call "sporadic" histories. For example, had I gone to the club more than half of the available Mondays, but in fits and starts, that would be sporadic. My going on four consecutive Mondays, then not at all for four more, then three Mondays in succession, would be such a sporadic history. In such cases I take it we would be less inclined to call this a habit than we would had I gone on exactly the same number of occasions, but with no gaps of more than a week.

In addition to this, there seems to be a further requirement concerning the most recent behaviour of the agent. If I had been to the club every Monday night for five weeks, and then did not go at all for another five, it would seem right to say that I had not repeated the behaviour sufficiently recently to say that I have the habit. Though we may have been willing to say that I once had the habit on grounds of repetition alone, we will probably not wish to say that I have the habit now. We might indeed want to say that I have "lost the habit" or "got out of the habit".[1]

The tolerance for exceptions, then, is quite a complicated business, and specifying some behaviour as a repetition will not generally be to specify precisely

---

[1] I shall offer my own understanding of these terms in §7.

how much play will be allowed in these matters. Nevertheless I think we can be more prescriptive about the way in which we specify repetitions, in a way that allows us to define which past behaviours should count as "exceptions", and which ones should not.

4. A Suggested Schema

I want to propose that in saying of a behaviour that it is repeated, we don't just convey a sense of *what* occurs, but we also imply an understanding of *when* it occurs. When it comes to specifying the behaviour, that is, in saying what it is to be the particular repetition that it is, we can distinguish between a description of the *kind of behaviour* on the one hand, and the description of what I call the *normal circumstances* on the other, that is, the circumstances or context in which that behaviour usually (statistically) takes place. So for instance, "my going to a Durham night-club" describes my behaviour, and "on Mondays" describes, at least partially, some of the circumstances in which that usually takes place. To take the earlier example, "Paul going sculling" specifies the behaviour, and "its being a Wednesday" captures the circumstances in which he usually does it. Table 1 (overleaf) summarises how, using this schema, we might capture this information for the six examples of repeated behaviours given earlier.

	Particular Behaviour ϕ	Kind of Behaviour	Particular Circumstances C	Normal Circumstances
a)	Tom going to the pub this evening	Tom going to the pub	It's being an evening of this week	Evenings this week
b)	Pamela cleaning the floor now	Pamela cleaning the floor	Pamela is nervous now	Pamela's being nervous
c)	My blinking now	My blinking	3 seconds elapsed since last blink	Short time elapsed since last blink, dust in the eye, strong wind
d)	Paul's sculling outing this Wednesday	Paul going sculling	It's being Wednesday	Wednesdays
e)	My biting my nails at time t	My biting my nails	My thinking at time t	My thinking
f)	My going to a Durham night-club this Monday	My going to a Durham night-club	It's being Monday	Mondays

Table 1

Adopting the above schema does not only force us to make explicit some features of the repetition which might otherwise be implicit in our descriptions. It also allows us to make explicit the tolerance of exceptions described in §3 above.

In the earlier example of clubbing on Mondays, given a certain history, it was easy to distinguish those parts of the history which were genuine ancestors to this repetition, from those parts which were exceptions. The reason for this is that in the description of the repetition in question the kind of behaviour (going clubbing) was clearly distinguished from the normal circumstances (Mondays). But it is possible to specify a repetition without making such things explicit, as in example (c), which was originally specified as "my blinking". Given only this description, although we can identify which parts of my history are past occurrences, we cannot locate any "exceptions", because we cannot make good sense of the idea of an "exception". This

is because we have no specification of when blinking should take place. Exceptions, in other words, presuppose an expectation of something happening, which doesn't, and without distinguishing the kind of behaviour from normal circumstances, we cannot specify when an absence is significant.

In allowing us to make sense of exceptions in every case, then, the schema also facilitates our making sense of what can count as sporadic and recent repetitions. For determining both sporadic and recent repetition depends upon our having some grasp on when our given behaviour normally takes place.

Given, then, that the distinction between kinds of behaviour and normal circumstances allows us to make sense of some of the aberrations from repeated behaviour, we could, if we wished, now answer our previous questions about where to draw various lines. We could, for instance stipulate that a behaviour is repetitious only if, in less than 10% of occurrences of normal circumstances, it did not occur. Or we could say that it is repeated only if, on the last three occurrences of normal circumstances, the normal behaviour occurred. Whilst such stipulations are possible, I think that they only open up possible discrepancies between ordinary usage and a technical notion of habit, and tiresome debates concerning counter-examples. So that we can instead focus on the more important aspects of habits, then, I am going to leave these matters vague, and the lines blurred.

One last point I want to make about specifying repetition concerns how we understand a specification of normal circumstances. This matter will become significant both later in this chapter (§7), and in later ones, since some arguments will turn on how exactly we understand the idea of normal circumstances. The problem is that in many cases, normal circumstances are very difficult, if not impossible, to capture linguistically. One reason for this is that there many be no particular point in repeating a behaviour. Repeating things is something we cannot help but do, simply because we encounter similar circumstances in our everyday lives, and our responses tend to be similar. Another reason is limitations in our descriptive powers. Capturing what is similar about a number of sets of circumstances is not something we should expect to be able to do very precisely, since there is only a finite stock of terms with which to describe such similarities, and even the terms we have may not be well

suited for the job. Of course sometimes agents repeat things on purpose, and in such cases asking the agent may be the best way to establish what sorts of circumstances she is intending to do such things in ("I'm hitting this drum 80 times a minute"). But even when we can ask such questions, it is still possible for the agent's actual behaviour to depart from what she intends, or will admit to, and it is her actual behaviour in which we are interested.

There may in fact be an argument for thinking that specifying normal circumstances will always be impossible. I adapt this from Dancy's (1993, Ch. 6) argument against the assumption that there are "general moral truths". The proposal is this. Given any finite specification of the normal circumstances, we shall always be able to think of some circumstances which do not fall under that description in which the behaviour does occur, and it would still naturally be counted as a relevant repetition. Perhaps more contentiously, we may also be able to think of some circumstances which fall under that description in which the relevant behaviour in such circumstances would not naturally be thought of as a repetition.

To illustrate this consider behaviour (a) from Table 1, for which the normal circumstances for Tom's pub visit are described as "evenings this week". First, suppose Tom's only pub visit on Tuesday was between 4 and 5pm. We might intuitively not want to count this as an evening this week, since it is too early to be evening, yet we intuitively would want to say that this repetition was significant in Tom's history, since it seems to support our saying of him that his pub-going is in this way repetitious. Intuitions may of course pull in other directions here, but the examples can be modified so that they don't.

Second, suppose that on Wednesday evening Tom was captured by a desperate criminal and forced at gun-point to go into the pub to be held hostage. Here he is intuitively going to the pub on an evening this week, but equally intuitively, he is hardly well described as doing anything "repeated". He has not been into the pub in anything like that *way* before, or since. Again, if intuitions slide, modify the example.

The thought is that any attempt to anticipate such exceptions by specifying them in a revised definition of normal circumstances will be doomed, because there

will always be other exceptions which we had not anticipated. I cannot see how to avoid this conclusion. This is not to say that the characterisation of normal circumstances is of no use, however, for a good characterisation will capture most repetitions, and that is how I intend the table to be read. But it is to say that we must read the specifications of normal circumstances as having an implicit *ceteris paribus* qualifications. In other words, a description of the behaviour plus the normal circumstances are together to be read as conveying a rough sense of what would be normal, and implicitly what would be abnormal, behaviour for that agent in those sorts of circumstances, which may leave some cases indeterminate. If, for some reason, we must classify such cases, it will require judgement and a decision.

We now have the materials for a definition of repeated behaviour.

(R) An agent A's behaviour ϕ in circumstances C is *repeated* if and only if A has on numerous prior occasions encountered circumstances similar to C ("normal circumstances"), and when in such circumstances, A usually, and on most of the recent occasions, behaved in a similar way to ϕ.[2]

The terms "numerous prior occasions", "usually", and "most of the recent" are here left deliberately vague for the reasons given above. This then will suffice for our definition of what the repetition involved in habitual behaviour amounts to.

5. *Automatic Behaviour*

There seems to be a consensus in the writing on habits that habitual behaviour cannot be characterised only in terms of repetition. R. S. Peters (1963) writes that "'Habit' also carries with it the suggestion not only of repetition but also of the ability to carry out the action in question 'automatically'" (p. 60). C. M. Hamm (1975) writes that "automatism is an essential ingredient of 'habit'" (p. 418). Neither writer tells us how this idea is to be understood. Nathan Brett (1981) gives us only slightly more, writing that "Habitual behaviour must in some respect be 'automatic'", which he glosses as "not the product of conscious decision" (p. 357). Brett's suggestion is at least

---

[2] This is the first of a number of definitions, a summary of which can be found in the Appendix.

consistent with what I suggested in Chapter 2, that is, that for Aristotle, Ryle and Wittgenstein, the automation in a particular exercise of a habit (virtue, behavioural disposition, learned response) could be understood in terms of a lack of deliberation about that particular exercise. In this chapter I want to clarify just what such automatism consists in.

Defining automatism requires us to avoid two extremes. On the one hand we do not want to underplay the phenomenology of habitual behaviours, that is, the lack of mental involvement on the part of the agent in performing them. So our definition of automatism would do well to explicitly exclude not just conscious decision, as Brett suggests, but all conscious thought processes, such as the contemplation of whether to do it, the weighing up of pros and cons, as well as decisions.[3]

On the other hand, we do not want to overplay the lack of mental involvement. For if, in trying to capture the lack of mental involvement, our definition of automatism pictures the agent as completely external to the behaviour in question, that could make capturing the intuition that agents are responsible for their habitual behaviours (the task of Chapter 4) impossible. Habitual behaviours are, as we have noted earlier, "second nature" for that agent, so she cannot be thought of as completely alienated from them. What I propose then is to give a definition of automatism which I hope avoids these two extremes.

The obvious place to start is with the thought that that we can simply define automatic behaviours as those over which we do not deliberate. This seems to capture an important aspect of our ordinary experience of habits. As we shall see I don't think a lack of deliberation is all there is to automatism, but it gives us part of a definition. The term "deliberation" could cover a multitude of sins, so we need to say what exactly falls under it.

---

[3] This would not rule out the possibility that such behaviours are done "for reasons", since, as we shall see in Chapter 5, nobody thinks that reasons need be consciously entertained.

6. Deliberation

Deliberation, as I shall understand it, is the familiar process of considering, or reflecting upon, the considerations relevant to performing some piece of behaviour, some of which might count in favour of doing it, or doing it in a certain way, others counting against, weighing these factors against one another, and then finally coming to a decision regarding whether to perform that behaviour. Such a process is inevitably incomplete. Temporal and epistemological constraints dictate that we will seldom, if ever, be in a position to deliberate upon every relevant factor. But that does not prevent us from making decisions about what to do, knowing that we might get things wrong. Nevertheless, if some such process, however partial or unsatisfactory, has been undertaken by an agent, either alone or with assistance from others, with the question of whether the ensuing piece of behaviour should be performed as its subject matter, I shall say that the agent has *deliberated* on that behaviour.

This characterisation leaves us with contentious cases, in particular when no other agent is involved in the supposed deliberation. I want to tighten our definition so that it is as clear as possible when we should say that deliberation has not taken place, so that we can know when a given behaviour is rightly classed as automatic. To this end I want to settle, again largely by stipulation, four controversies.

An initial controversy is over just how little of the process described above an agent needs to have performed in order for it to still count as deliberation. Specifically, as it can merely involve reflection, without any processes of weighing or deciding, what can count as "reflection"? Here I want to say that insofar as we can say that the agent has done something distinct from the behaviour, which might have resulted in her not behaving at all, she has reflected. So, to have deliberated, the agent must have done at least two things - reflecting and behaving - and not one - just behaving. If the only grounds for saying that the agent reflected is that she behaved in such and such a way ("in ϕ-ing she saw that she should ϕ"), this is not sufficient to say that she has done two things. There must be independent grounds for saying that the agent was involved in a process that might not have resulted in the behaviour.

Usually the agent herself will be authoritative on this matter if questioned in a suitably non-coercive way.[4]

A second source of controversy concerns just how *conscious* the agent needs to be of a deliberative process in order for it to count as deliberation. Paradigmatically, deliberation is conscious. The agent is aware of the content of that deliberation at the time. Indeed she might perform such deliberation out loud, or with the assistance of others. If asked, she will be able to say whether she is going to perform the piece of behaviour before she does it. And if asked afterwards why she did it, she is likely to be able to recall the contents of the deliberative process and cite what she then found to be decisive considerations as being amongst her reasons for doing what she did ("I did it because I thought such-and-such"). Needless to say, when deliberation is conscious in all these ways, any resulting behaviour is not automatic.

However, we seem to be able to make some sense of the idea of *non-conscious* deliberation. Deliberation might be thought to be non-conscious in two senses. First, in the sense that the agent might seem to be able to report on an apparent process of deliberation, even if she was unaware of engaging in any such process at the time. I call this *sub-conscious* deliberation. I call it "sub-conscious" rather than "unconscious" because such a process could have been carried out consciously (perhaps given more time prior to the behaviour), and can be brought to consciousness given reflection. Here, the basis for saying that she deliberated is her readiness to make such a report, plus the assumption that such a report is a matter of the agent's remembering a prior process which is distinct from the behaviour itself. These are importantly different grounds from those upon which we construct a "*post hoc* rationalization" of a behaviour, after it has been performed.

I shall say more about *post hoc* rationalization in Chapters 5 and 6, but we need to be clear on a number of features of it now in order to distinguish it from sub-

---

[4] Not always. Deliberation can be done overtly, perhaps out loud, which the agent could not plausibly deny, and presumably would not. At other times, the agent having paused before acting, possibly with a "wondering" expression, furrowed brow, and so forth, will be sufficient grounds for attributing deliberation regardless of what she might later assent to.

conscious deliberation. The idea is that we can imaginatively construct a justification for some behaviour that presents the context in which it took place as it might have struck the agent, as cohering with her overall conception of the world, her projects and so on. We can do a *post hoc* rationalization for virtually any behaviour, whether or not the agent was even capable of a deliberation with that content. Indeed we often perform *post hoc* rationalizations for baby and animal behaviour. So this kind of rationalization need not be performed by the agent herself. Anybody with a good understanding of the agent's perspective will be able to do the same. And of course it follows that rational agents will standardly be able to construct such rationalizations on their own behaviour. So we should not conclude from the fact that an agent has this capacity following a particular behaviour, that she must have deliberated.

Given that such *post hoc* rationalization is possible, then, if we are faced with an apparent example of the agent "remembering" a hitherto unconscious deliberation, we need to be sure that it is not just a *post hoc* rationalization in disguise. Fortunately we do not need to settle here the question of what might constitute adequate grounds for thinking that remembering had taken place. All we need to say is that if there are cases of genuine recollection, that may be reason to say that the actions in question are not automatic.[5] Hence I want to allow that sub-conscious deliberation, if such sense can be made of it, is compatible with the idea of that behaviour being automatic. For as we shall see in Chapter 5, somebody might try to argue that reasons are present in habitual behaviour in this sub-conscious way.

There is a second and more extreme sense of non-conscious deliberation that I do not want to include in the definition. For one might want to say that the agent had deliberated even if she was unaware of any such process at the time, and could not recall the content of any such process either. There may be theoretical motivations for saying this, such as the aspiration for a theory of action which unites thought-out and non-thought-out actions.[6] But taken as a pre-theoretical view about what deliberation

---

[5] This may allow us to include "Freudian" accounts the sub-conscious in this class, because for these, there is always the possibility of revealing the true content of a deliberation in psychoanalysis.

[6] See for example Smith (1998)

is like from the point of view of the experiencing subject, it lacks support. Indeed, theories that misrepresent the phenomenology of actions in this way contribute to the prevailing insensitivity to the distinctiveness of those behaviours - habitual behaviours being a paradigm - about which we do not apparently deliberate. The temptation is to model them on those which we do, but that may be an intellectualist mistake. At this stage then, I do not want to prejudice our account with such theoretically motivated notions of deliberation, so this is not a sense of deliberation that I shall recognise here.

A third controversy is that of *when* a process of deliberation can be said to take place. When an agent deliberates about doing something, perhaps most commonly, she does this shortly before doing it, perhaps only a second or two before, or even less. But the time gap between deliberation and resultant behaviour can vary considerably. One might deliberate and come to a decision a long time in advance of the behaviour. For instance, I can deliberate about going round the world when I retire, which is a number of decades away. At the other extreme, deliberations might take place concurrently with the behaviour. For instance, somebody asks me to push their car for them to get it started, and its only when I have braced myself against the rear bumper that I wonder whether this is a good idea, in view of the size of the car, although I decide that it is, and keep pushing. Although the long-term deliberation might be thought not to specify a particular behaviour, but only a kind of action (and a rather vague one at that), and although the concurrent deliberation could at most interrupt the behaviour, because neither are characteristic features of habitual behaviour, I want to allow that all of these kinds can count as deliberations.

The final two controversies concern the content of the deliberation. The first of these is whether we shall say that deliberation concerns a *particular* piece of behaviour, or about a *kind* of behaviour. This is particularly important if we are going to include habitual behaviours in the class of automatic behaviours, because we have acquired some of our habits on purpose, that is, following a conscious decision to do so. At some point in the past, then, an agent decides that she would like to perform a certain kind of behaviour with less effort than she has done up until now, and so she resolves to make it a habit, which she then acquires by practice. The whole point here

is that her prior decision was made in order to *avoid* the need for deliberation about particular behaviours of the kind that she wishes later to perform. What makes her prior deliberation (about forming the habit) not deliberation in the required sense, is that it is deliberation about a kind of behaviour. Deliberation for the purposes of our definition, then, is deliberation about whether to perform a *particular* piece of behaviour.

The fifth and final controversy concerns whether automatic behaviour should exclude deliberation about *how* to perform the behaviour in question. So far the only deliberation I have specified is over the question of whether one should perform the behaviour. Clearly habitual behaviour does not involve that. However, it might be consistent with some uses of the term that an agent might still wonder how precisely to perform that particular behaviour, the particular style of performance, or taking into account novel features of the particular context. In Chapter 2, §3 we saw how this might affect our understanding of the automatism of virtuous actions. For when an agent acts virtuously, say in an exercise of courage, she may need to resolve questions about the details of how to be courageous in this particular case, although the question of whether to be courageous, quite properly, does not occur to her. So whilst we need not maintain that thinking about how to do something always occurs in habitual behaviour, it would seem to be an unnecessary restriction to rule such thinking out of our definition. Thus if our definition of automatic behaviour only rules out deliberation about *whether* to do it, other sorts of deliberation are left as possibilities.

Given the above clarifications, I want to adopt a lack of conscious deliberation, concerning whether to perform a given particular piece of behaviour, as a necessary condition for that behaviour to be automatic. However, I don't think it is also sufficient.

7. *Trying*

Anthony Kenny (1989) writes: "If one has a habit of doing X then it is harder not to do X than if one has not" (p. 85).[7] I think this expresses rather well the idea of how a habitual behaviour can be said to be "second nature" to the agent. But I do not think that saying of automatism that it consists only in a lack of deliberation fully captures this thought. For deliberating may, to many of us, be no "harder" than not deliberating, and when that is so, simply saying that one did not deliberate about performing a given behaviour might not yet capture what is easy about it. I propose then to add another condition to the definition of automatism which more fully captures the lack of difficulty in habitual behaviour, and that is by saying that automatic behaviour does not require the agent to *try*.

This is a more contentious thing to say than it may seem, however, since Brian O'Shaughnessy (1973) and Jennifer Hornsby (1980) have both claimed that all actions involve trying, and Philip Pettit (1993) has even said that exercising habits in particular involves trying. I think it is worth taking the time to see how these sorts of move can be resisted. I tackle the more specific, and thus potentially most damaging, objection first.

I mentioned Pettit's (1993) "ethocentric" account of rule following in Chapter 2, §8. Pettit thinks that rule following can be understood as the exercise of the kinds of habits he calls "extrapolative inclinations" (1993, pp. 86-97). In order for extrapolative inclinations to provide an account of *following* a rule, as opposed to merely acting in accord with one, Pettit thinks that the agent must "try" to exercise them. The idea of trying is in turn understood in terms of the agent's having certain beliefs and desires, thus they are squeezed into the mould of "intentional" actions. Trying, then, for Pettit, is nothing more than the possession of mental states.

There are a number of objections to this move, not least of which is that Pettit offers a contentious account of what it is to try. More seriously, if what I said in

---

[7] The context is Kenny locating habits within the class of dispositions. Unfortunately he does not develop his suggestive remark.

Chapter 2 (§§6-8) about rule following is right, extrapolative inclinations understood in this way cannot provide an account of rule following, since there is a question of how to interpret the mental states, which in turn will require further mental states, and so on *ad infinitum*. In the terms of Chapter 2, §6, Pettit has chosen the position that McDowell calls Scylla.

Even if Pettit's view fails as an account of rule following, he may still be right about habits always involving trying. But I do not think that this claim does justice to the evident phenomenology of habitual behaviour, which, given reasoning analogous to that made about deliberation above, gives us no grounds for saying of an agent that she not only behaves in a certain way, but that she tries to as well. But here we run into the more general objection, which states that on the contrary, all actions are in fact accompanied by trying.

There are two writers who defend such a claim. O'Shaughnessy (1973) proposes that "trying is an essential constituent of intentional action as such" (p. 53). Similarly, Hornsby (1980) claims that "Every action is an event of trying or attempting to act" (p. 33). If either of these writers is right, it would undoubtedly mark significant progress towards a unified analysis of actions. But for our purposes, defining automatic actions as those which (amongst other things) the agent does not try to do, would result in the immediate exclusion of habitual behaviours from the class of actions. To avoid this outcome, which would present a significant problem for the argument of Chapter 4, I need either to distinguish my sense of trying from that used by O'Shaughnessy and Hornsby, or to find a way of responding to their arguments. I shall begin with the latter option.

Both writers recognise that the *prima facie* evidence counts against them. O'Shaughnessy notes the "oddity" of saying "'He tried to walk across the road', of a normal able-bodied man in a setting of rural peace" (1973, p. 53), though he goes on to assert that such sentences "must be perfectly intelligible, and capable of truth and falsity, in humdrum circumstances of this kind." (1973, p. 53).

Similarly Hornsby writes,

> It seems that it is only appropriate for a speaker to say that an agent tried to ϕ, if, for some reason or other, the agent did not - or it was thought that he did not - straightforwardly and easily ϕ. (1980, p. 34)

Nevertheless, she notices that trying seems only appropriate when various "doubt or denial" ("D-or-D") conditions are fulfilled, an idea she takes from Grice (1961). And following Grice, she admits that "if this condition is not fulfilled, the utterance of a statement [that says he tried] may well be extremely misleading in its implication" (1980, p. 34). The fact that such a statement would be "misleading" does not, according to Grice or Hornsby, entail that it is false. Indeed Hornsby, like O'Shaughnessy, thinks it is true.

First of all I want to agree with both O'Shaughnessy and Hornsby, that it is "odd" or "misleading" to say that an agent "tries" in "humdrum" circumstances. And it seems to me that we can explain this oddity by appealing to phenomena such as habitual behaviours. When an agent behaves habitually, the circumstances are utterly familiar to her, and she naturally "finds herself" engaged in exercising the habit. There is no question of failure, either for us observers, or for the agent herself. As a result, we will not find it at all natural to describe such behaviours as requiring the agent to "try". That is why I think citing the absence of trying is a good way of capturing the automatism of habitual behaviours.

Where I need to disagree with both O'Shaughnessy or Hornsby, however, is in thinking that in such cases, it is still true to say that the agent tries. The way that both writers try to persuade us that this is right is by means of thought experiments which place doubt about the success of an action in the mind of an observer, though the action in fact proceeds successfully.[8] Hornsby's example is of an onlooker who has been misinformed about the difficulty his friend will have in moving a boulder. As it

---

[8] O'Shaughnessy writes: "though we speak of trying only when success is in doubt, that doubt could dwell in someone *other than* agent or speaker; so it seems that we misunderstand this linguistic rule if we suppose it unconditionally to forbid mention of trying when one is in no doubt. It follows that sentences attributing trying to an agent must be perfectly intelligible, and capable of truth and falsity, in humdrum circumstances of this kind." (1973, p. 53).

turns out there is no difficulty and the onlooker says "I was right about one thing at least. I knew that my friend would try to move it" (1980, p. 35).

Now I don't say that these examples should persuade us.[9] But even if they are found persuasive for one-off behaviours, I don't think we should accept that they work when the behaviour is habitual. The reason is that insofar as there is doubt about the success of a behaviour, from whatever point of view, that would compromise the claim, from the same point of view, that the circumstances in question are normal, in the sense outlined in §4 above. This is a point about the logic of normal circumstances. Hence, by definition (R), the behaviour in question would be disqualified from being habitual on the sole grounds that it is not repeated, never mind how the rest of the definition of habitual behaviour turns out. In other words, one cannot consistently claim both that a behaviour is repeated according to (R), and that there is some doubt about whether it will happen. The doubt is always an indication that something abnormal is afoot.

We can see this kind of move at work even more clearly if we consider Hornsby's own example of an allegedly habitual behaviour that she thinks involves trying. She asks us to,

> Consider Brown who has a daily routine of waking up, getting dressed ..., which he carries out habitually, without reflection or ado. ... One day someone is provided with a reason to think that Brown ate something on the previous night that causes specific motor disturbances, and to think that as a result Brown will not be able to tie his knot this morning. (Hornsby 1980, p. 37)

Brown, as it turns out, is unaffected, and as before, Hornsby concludes that the observer "always knew that he would *try* to knot his tie" (1980, p. 37). Again Hornsby clearly trades on abnormal circumstances - the suspect food - which, insofar as they

---

[9] Both writers rely on a questionable distinction between how (presumably competent) language users use the term "trying" on the one hand, and on the other, either the "linguistic rule" which governs its use (O'Shaughnessy, see n. 8 above) or the conditions which make sentences containing that term true (Hornsby, e.g. 1980, p. 38). This sort of distinction would be rejected by Wittgenstein for one (see Chapter 2, §§6-8).

obtain (or are thought to obtain), diminish the plausibility of the claim that the behaviour is habitual (or is thought to be habitual). The fact that an observer simply got things wrong, should not affect what it is right to say. What we need is an example of a habitual behaviour which, although the circumstances are normal in every way, and are acknowledged to be so by all parties, still involves trying. And I doubt we can find any such example.[10]

In fact, my disagreement with O'Shaughnessy and Hornsby may not be as great as it seems. For both writers later admit that when actions are "successful", the agent does not do two things - try and successfully act - but one. Thus they think that a successful act is identical with the agent's trying to do it.[11] So what we have here may in fact be a disagreement about terminology after all, since they are prepared to call a successful act "trying" whilst I am not. My reply is that insofar as they insist on this appropriation of the ordinary use of the term "trying", applied to habitual behaviours this revision lacks justification, and for present purposes it would prevent us from making the distinction we want. If this means that we shall have to look elsewhere for a unified account of action, so be it.

Thus I want to understand trying in what I think is a natural way, so that it does not "misleadingly" apply to all behaviours, but rather informs us about characteristics of particular sorts of behaviours. Trying, then, is whatever extra effort is required to perform a behaviour given that the agent has expectations about its possible failure. I want to say that insofar as habitual behaviours are automatic, they do not involve such trying.

In sum, I want to define automatic behaviours as follows:

---

[10] In Chapter 8 I shall suggest how this sort of response to Gricean moves might fruitfully be developed.

[11] See O'Shaughnessy's "Postscript" (1973, pp. 73-4); Hornsby writes "if ever we try to ϕ and succeed in ϕ-ing, then our trying is our succeeding" (1980, p. 39). I am aware that Hornsby also thinks that "all actions occur inside the body" (1980, p. 14), which would definitely be harmful to the present thesis. However, the view is, I think convincingly, rejected by Jonathan Lowe. See the extended exchange between them in Lowe (1981, 1983, 1984) and Hornsby (1982, 1983).

(A) A behaviour ϕ is *automatic* if and only if the agent performing ϕ engages in no process of conscious deliberation about whether to ϕ, either before ϕ is performed, or during its performance; *and* the agent does not try to ϕ.

Whilst this definition characterises automatic behaviour in purely negative terms, I take it that it nevertheless captures something of the phenomenology of habitual behaviour: it is a natural and simple kind of activity. Importantly, because the characterisation of automatic behaviour is negative it does not outstrip the phenomenology. If other philosophers wish to analyse habitual behaviour into component parts, they will have to go beyond the phenomenology, and the onus of justifying such an analysis will be with them.

The idea of automatism also captures an interesting feature of habitual behaviour: that we *find ourselves* engaged in it. Everybody has had that experience of being engaged in some familiar routine, say brushing one's teeth in the morning, or putting the kettle on (early morning routines seem replete with examples), and then reflected on the fact that one never made any sort of decision or effort to do that thing. One just finds oneself doing it. This may seem alienating, like finding somebody else doing these things. One might think: "Here is a person (or 'body') going through some routine, upon which I am a little more than a spectator". But I would want to say that the experience of finding oneself engaged in a habitual behaviour is a proper part of healthy, non-alienated, agency. The routines are familiar to one, since one has oneself repeated them many times before. One has been at least vaguely conscious of what one is doing before "coming to", though the consciousness is inarticulate, and little more than a sense of being awake. And importantly, one becomes aware of the behaviour, as it were "from the inside", a perspective from which one cannot view others. Thus I think that any sense of alienation in finding oneself behaving habitually is either a sign of mental illness, or an illusion. The occurrence of the phenomenon is, I think, rather to be viewed as a good indicator that a repeated behaviour is automatic.

It may also be worth noting that the requirement for automatism places limits on the sorts of repeated behaviours that could become habitual. For some behaviours will never become automatic, no matter how many times we repeat them. These

might include complex routines such as playing a game of football; or activities that are intellectually demanding, such as playing a game of chess or writing a philosophy paper. Whilst initiating such activities may become automatic, and various elements of them may also be like this, carrying them through from start to finish will never become automatic for creatures at our point in evolution, at least. Trying or deliberating will be an inevitable part of such activities. It is for that reason that habitual behaviours will tend to be relatively simple routines which can be mastered in the way suggested.

8. Learning and Losing Habits

We can now say what habitual behaviours are, namely, all and only behaviours which are both repeated according to (R) and automatic according to (A). This definition puts us in a position to say what it is for a habit to be *learned*, and what it is for a habit to be *lost*. This will not affect our definition of habitual behaviour, but that we can account for these things is one test for the adequacy of that definition. Let us consider learning first.

Any kind of learning involves a transition of an agent, or knower, from one state to another. From a state of ignorance to a state of knowledge. This is no different for learning habits. For as long as there was a time when the agent could *not* perform this kind of behaviour automatically, and that there is now a time that she can do it automatically, we have identified a transition from one state of the agent to another. That it what it is to learn a habit. From simply being able to do something, to doing it naturally, as a characteristic part of who one is. This is the acquisition of practical knowledge. In Ryle's terminology, one becomes *disposed* to behave in a new way.

The transition marks a change in the significance of the normal circumstances. When we say of a behaviour that it is merely repeated, the normal circumstances are those in which the repetition, as far as we know, just happens to take place. The fact that the behaviour usually takes place when the normal circumstances arise may be purely accidental, and the only predictive inferences that can be drawn from one's occurrence to the other's would be based purely on statistics. But once the agent can

be said to be disposed to behave in this way, the connection between the normal circumstances and the behaviour is seen as non-accidental in a new way. Since the behaviour in question is only automatic in these sorts of circumstances, it can now be regarded as a *response to* those circumstances, which has more than a statistical significance. The occurrence of the circumstances plus the disposition of the agent, could now be said to *explain*, in a new sense, *why* the behaviour in question happened. The sort of explanation we have here would be comparable to explaining a leaf falling by saying that the tree it came from is deciduous and winter has arrived; or by explaining the cat's pouncing by saying that it likes playing with wool, and it's spotted a ball of the stuff.[12] The transition in question, then, is potentially of considerable importance.

The transition from not being able to perform a behaviour automatically, to being able to, typically comes about in one of three ways, which I call *intentional*, *unintentional*, and *non-intentional* respectively. Intentional learning comes about when an agent deliberately chooses to acquire a particular habit. So I might decide that it would be a good idea if I got up early every day to get some work done before breakfast. I train myself to do so by setting my alarm for 7 a.m. and after a few weeks it starts to get rather easier. I get used to it, and it soon becomes harder not to do it even at weekends. I may even find myself getting up at that time even though my alarm has stopped working. Since I now do it automatically, we can say that I have got into the habit of getting up at that time. We can train ourselves to acquire many other useful habits in this way.

Habits can also be acquired unintentionally. Simply repeating a kind of behaviour on numerous occasions can be enough for this to happen, though one does not intend it to. So acquiring the habit of watching television in the evenings is something I never intended to happen, but the repetition alone makes me do it automatically now (a fact that I find rather annoying in view of the poor quality of the programmes on many evenings).

---

[12] I will develop this idea a little in Chapter 6, §4, where I suggest that this sort of explanation is causal.

The third way that habits can be acquired is not merely unintentional, but is better described as non-intentional. These habits are acquired during upbringing, and can pre-date our linguistic and deliberative capacities altogether. So there is no question of the individual having made or failed to make any sort of decision. Rather, the natural reactions of the infant become gradually refined as she matures, through the natural mechanisms of imitation and repetition. This process is part of the acquisition of second nature that we saw McDowell describing in Chapter 1.

This account of learning is consistent with the ideas proposed by Aristotle, Ryle and Wittgenstein discussed in Chapter 2. These writers all think that the repetition is part of the learning process. And whilst their respective accounts of the particular modes of learning might differ, that need not harm the claim that they are each talking about behaviours which are both repeated and have become automatic. This further justifies counting these three writers' notions as habits in our sense.

An account of losing habits, in contrast, like the idea of not knowing something any more (or forgetting it), goes the other way around. If there was a time when I had some habit, that is, I used to repeat the behaviour and could do it automatically, but now I either have to deliberate or try in order to do it, if I can do it at all, this would count as my having lost the habit. Notoriously losing habits is much more difficult than gaining them, though there are more and less effective strategies. Simply being away from the normal circumstances for a given habit for a long period is one way of losing it. Performing alternative actions in those same circumstances may be more effective. Experience also shows that losing a habit need not be permanent. Indeed getting a habit back - *re*learning it - is likely to take far less in the way of practice than it did the first time.

Is being learned a fourth feature of habits? One might thinks so since it would be a way of excluding repeated automatic behaviours that are not learned, such as the heartbeat, blinking and the digestion, from our definition of habit. But we shall see that such behaviours can be excluded anyway given our third criterion, responsibility, which is to be explicated in Chapter 4. I cannot see any other reason to include learning as an additional criterion.

9. *Definition*

We can now state our definition of habitual behaviour as follows.

(HB) A behaviour ϕ is *habitual* if and only if it is repeated according to (R) and automatic according to (A).

This definition captures the first two features of habits, and will form the basis of our definition of habitual action, the topic of the next chapter.

4

ACTIONS, RESPONSIBILITY AND CONTROL

1. Actions or Compulsions?

In Chapter 2 I argued that Aristotle, Ryle and Wittgenstein all opposed intellectualism by employing the idea of a behaviour which is repeated, automatic and responsible. In Chapter 3 I developed the idea of a habitual behaviour by describing in detail how we should understand the first two of these features. In this chapter I want to complete the definition of a habit by spelling out the sense in which we can be said to be personally responsible for some of our habitual behaviours. This will support my claim that we should call such behaviours habitual *actions*.

According to definition (HB), habitual behaviours are not distinguished from other acquired behaviours which are also repeated and automatic. Addictions such as smoking, various kinds of neuroses, compulsive behaviours like binge eating, expressions of phobias such as vertigo, and nervous tics are all of this kind. Because these behaviours, which I group under the heading *compulsions*, are most obviously thought of as things that happen to agents - we are victims of them - rather than things that agents do, it seems reasonable to say that they should not be counted as actions. With only the two features so far discussed, then, we have no reason to think that habitual behaviours as such should count as actions, any more than compulsions do.

But in this chapter I shall argue that there is a class of habitual behaviours which are not compulsions, and deserve to be thought of as genuine expressions of agency. That requires us to decide upon a suitable criterion of action-hood, which allows us to distinguish repeated automatic behaviours which are compulsions from those which are not. As we shall see, settling on such a criterion is not straightforward.

In §2 I explain why I shall not adopt a standard criterion for actions, according to which they are intentional under some description. In §3 I defend my proposed criterion for actions, which is that they are behaviours for which we are personally responsible. In §4 I propose an intuitive test for personal responsibility which derives from a suggestion by Strawson, which I call the *reactive attitudes test.* I apply this test to show that we seem to be responsible for some habitual behaviours, and not others. In §§5-8 I develop a notion of control that explains the results of this test, and thus how we can be said to have personal responsibility for many habitual behaviours. This will be the basis of my completed definition of habitual actions.

2. *Intentional Actions*

When philosophers enquire into what an action is they generally begin by noticing the intuitive distinction between things that agents do on the one hand - actions, and things that merely happen to agents - mere behaviours, on the other.[1] Contemporary writers then tend to make a second move, which dates back at least as far as Anscombe's (1957) *Intention,* which is to characterise the agent's doings as being "intentional" under some description. The thought is that a piece of behaviour can be described in many ways, but only when it is described in certain ways does a description capture it as something the agent intended to do (or "meant" to do, or did "on purpose"). So, to use Anscombe's (1957, pp. 37-41) example, a man might be replenishing the water supply and poisoning the inhabitants, but it is only the first of these that says what he intends. As long as we can settle on at least one intentional description of a behaviour, we can say that it is an action. The task then is to offer an analysis of intentional actions thus construed. But I think that for our purposes this is a bad strategy, and I shall now explain why.

Analyses of intentional action vary greatly, but can be roughly divided into those which emphasise the role of the agent's will, or some variant of it, and those that emphasise the role of rationality, understood as the agent's responsiveness to reasons.

---

[1] This characterisation pervades introductory texts on action. See for instance Dretske (1988, p. 1), Moya (1990, p. 2).

I take my cue here from a recent survey paper on intentional action by Al Mele (1992b), who organises his discussion around two "proto-analyses" of action:

A1. S intentionally A-ed if and only if S A-ed in the way that S intended to A.

A2. S intentionally A-ed if and only if S A-ed for a reason. (Mele 1992b, p. 200)

Those that adopt A1 and thereby emphasise the will, whose variants include "volition", "trying" and "intending", include Jonathan Lowe (1996), Carl Ginet (1990), Brian O' Shaughnessy (1980) and Jennifer Hornsby (1980). Those that adopt A2, and thereby emphasise rationality, include Anscombe (1957), Donald Davidson (1980), Alvin Goldman (1970), Al Mele (1992a), Jonathan Dancy (1993, 2000) and Rowland Stout (1996). This classification is rough, and some writers offer analyses which might be said to fall into both groups (Hornsby is perhaps one of them). But that will not matter unduly. For whilst one or both of these ways of analysing action could potentially provide the grounds for distinguishing habitual actions from compulsions, I don't want to adopt either.

The reason is that I think that both kinds of analysis embody prejudices against the very idea of a habitual action. This may be due to a tendency to assume, following influential writers like Ryle (1949) and Winch (1958), that habitual behaviours are mere mechanical responses to stimuli which shouldn't really count as actions anyway. In Chapter 2 we saw some reasons why this view is impoverished.

Nevertheless, some writers appear reluctant to say that habitual behaviours are definitely not actions. Davidson (1963), for instance, seems to want to make room for the possibility of habitual actions, since he seems to allow that a habit can be (at least part of) a reason.[2] For considerations I shall give in Chapter 5, however, I think Davidson is wrong in this latter regard. Another approach is taken by Alvin Goldman (1970). Rather than trying to incorporate habitual behaviours into his analysis of intentional action, Goldman is candid about its inadequacy in this respect. He writes that "It must be conceded ... that the analysis of action I have given is not

---

[2] Davidson's characterisation of a "pro attitude", one component of a "primary reason", includes "social conventions" and "permanent character traits" (1963, p. 4).

ideally suited for dealing with habit, nor has it been designed with habitual behaviour in mind" (1970, p. 91). Goldman's own theory is reasons-based, but I think the same anti-habit prejudice infects will-based analyses too.

This pervasive prejudice leads to intellectualist distortions in both kinds of analysis. Will-based analyses are vulnerable to two sorts of distortion. Such analyses may posit "internal" items which are vulnerable to the sorts of difficulties incurred by the Cartesian intellectualism that Ryle attacked (Chapter 2, §4). Or they may posit explanatory items for which there may be no phenomenological grounds. We saw how this worked in the case of "trying" in the previous chapter (§7). We are asked to accept that every intentional action is accompanied by some push from the will.

Those analyses that emphasise rationality tend also to make intellectualist assumptions about what it means for an action to be rational, namely, that it is done for reasons, an idea which is understood in a particular way. This is a position which I shall criticise at length in Chapters 5 and 6. But, to anticipate, I think those that hold reasons-based analyses still conceive of intentional actions as things that agents, as Stout puts it, "*make happen*" (1996, p. 9). And again, we look for an injection from the intellect, this time in the form of reasons, to mark out intentional actions from other phenomena.

Of course, any analyses that accommodate both will and rationality, are vulnerable to both sorts of intellectualism.

It is because of these sorts of difficulty that I do not want to adopt either kind of analysis to settle the question of how habitual behaviours can be said to be actions.

What I think lies at the heart of the problem is not the particular analyses of intentional action, but their starting point. The very idea of an intentional action is understood as a behaviour in which the agent has some quite direct intellectual involvement, which has a content sufficiently determinate and articulated for us to settle on some definite descriptions of the behaviour under which it is purported to be intentional. But in habitual behaviour, which is by definition automatic, we have no reason to think that any determinate or articulated characterisation of what the agent is doing must or will be available to her. Yet in their determination to find a unified

theory of action, many contemporary analytic philosophers of action are prepared to gloss over such details.

Of course, there may be ways of understanding the idea of an intention which require less intellectual involvement from the agent. But my suspicion is that using the idea of intentional action to decide which habitual behaviours are actions will be a bad place to begin. For none of them are clearly intentional anyway. I want rather to make a fresh start by going back to the original intuition about actions.

3. *Other Characterisations of Actions*

Fred Dretske (1988) is one of the few philosophers of action who does not begin with the idea of intentional action. Instead he begins with a much more "general notion", for which he prefers the term "behaviour", and which applies to "animals, plants, and even machines in very much the same way as it applies to people" (p. 3). Significantly for the present project, Dretske's notion "applies to people ... when there are no purposes or intentions" (p. 3). Thus for Dretske, the following examples all qualify as "behaviour":

> People shiver when they get cold. That is something they do. They also perspire when they get hot, grind their teeth when they are asleep, cough, vomit, weep, salivate, blush, tremble, hiccup, inhale, exhale, choke, fumble, stammer, fall asleep, dream, wake up, and a great many other things that are in no way voluntary, deliberate, or intentional. (1988, pp. 3-4)

Although there is no explicit mention of exercising habits in Dretske's list, I don't think they would be out of place. My reason for thinking this is that Dretske's list seems sensitive to the way we standardly describe the behaviours on his list. We saw how this could work in the discussion of Wittgenstein in Chapter 2 §8. The suggestion there was that one way of determining whether a given happening is thought to be an action is to look at whether is it described by means of a subject ("I", "he", "Maureen", etc.). The behaviours on Dretske's list are all standardly attributed to a subject in this way, and exercising a habit is something that is described in the same way ("*I* exercise my habit of nail-biting"). If this is right, Dretske might have a way of capturing the sense in which habitual behaviours are things people do.

But there are at least two reasons why Dretske's scheme will not work. First, many of the behaviours in his list are the sorts of things that I earlier called "compulsions", in the sense that we are victims of them. Usually, when we cough, weep or stammer, we are doing something which we cannot help, and indeed something we would rather not do if we had the choice. Notwithstanding the way we describe them, they are not expressions of agency. So it looks as though Dretske has drawn the line in the wrong place, at least for our purposes.[3] This also marks the unreliability of the linguistic approach.

The second reason I don't think we should go with Dretske is that the analysis he offers of the distinction between "behaviour" and other sorts of thing that can happen to creatures, is problematic. He cashes out that distinction in terms of whether the change is caused "internally" or "externally" (1988, p. 3). And for Dretske an "internal cause" is understood literally, as that which originates "inside or underneath the skin, fur, fins, feathers, or whatever" (p. 3, n. 3).[4] Now whilst nobody would deny that *something* goes on underneath our skins whenever a behaviour in Dretske's sense takes place (though we can expect it to be quite different for each sort of behaviour), it is highly contentious to claim that reference to such happenings could tell us anything about what it *means* for us to behave in such a way. After all, we can learn what actions are without learning anything about what goes on under the skin. The conceptual investigation into actions is the prior task, and what we are concerned with here. Dretske's approach then, is not as much help as it first appeared.

In the absence of other non-intentional characterisations of actions in the mainstream literature, we are pushed back onto our own resources.

I want to return to the intuition that actions are what agents do, as opposed to what happens to them. I propose to fill out that intuition in the way we did with our

---

[3] Karlsson (2002, pp. 62-3) has recently made a similar objection to Dretske's approach, and advances an Aristotelian alternative, in a paper to which I would like in future to give a fuller treatment.

[4] I think there may be an alternative way of understanding the "internal"/"external" contrast which does allow habitual behaviour to have an "internal" source of change. See §8 below.

working definition of habits in Chapter 1, noting that, whatever else "things agents do" may be, they are behaviours for which agents are personally responsible. Putting things this way gives us an approach to actions which emphasises the contrast between actions and phenomena which are merely causal. This is not to say that actions are not causal phenomena, they are just not only causal. In Chapter 2 §3 I suggested one way in which we can understand this distinction, and that is by saying that whereas things for which we are personally responsible warrant attributions of praise or blame, things that bear only causal responsibility, do not. I want to develop this idea as a better criterion for actions.

4. Reactive Attitudes

Sir Peter Strawson (1962) connects the fact that agents are morally responsible for what they do with the fact that we hold various "reactive attitudes" such as gratitude, resentment and forgiveness, towards them.[5] Having such attitudes towards one another seems to be an inescapable part of being a human being with interpersonal relationships. Now unlike Strawson, our concern is not with the compatibility of free-will with determinism, nor is it with specifically moral responsibility, but we can employ the same kind of conceptual connection that he uses. We can say that we know that an agent is personally responsible for a given piece of behaviour if, as a result of various bits of information which concern that behaviour, we could harbour, and perhaps express, reactive attitudes towards that agent. So if, given the right surroundings, I could resent an agent for her arm going up, perhaps because I knew she wanted to raise the bid, that shows that her arm going up is an action; her action. That I could not resent her, in any surroundings, for her arm moving - after all, somebody else nudged it - shows that it is amongst the mere behaviours. The same thing could, I take it, be said of the coughing, weeping and stammering. I call this test *the reactive attitudes test.*

---

[5] According to Jay Wallace (1994, p. 8, n. 10), who discusses Strawson, the connection between responsibility and such attitudes goes back to Joseph Butler's (1726) sermon "Upon Resentment".

I am going to use the test not as the ultimate arbiter of what is a responsible behaviour and what is not, but merely as a way of bringing out our intuitions about whether some bits of behaviour seem responsible or not. That should be enough to motivate a further enquiry into what underlies these intuitions.

Let us apply the reactive attitudes test to habitual behaviours. For instance, because I live in Britain I am in the habit of driving on the left. I do it repeatedly and automatically. Exercising this habit is something that I could be thanked for, given that it gets me and my passengers from A to B in safety. If driving on the left were something that merely happened to me, such an attitude would make no sense. Similarly when I indulge my long-standing habit of nail-biting I regularly attract the attitude of impatience from my partner. Unfortunately I can't pretend I am completely blameless for this minor self-mutilation. My suggestion, then, is that at least some habitual behaviours will pass the reactive attitudes test for personal responsibility, and this should raise our suspicions that they are actions.

Compare our attitudes towards the behaviours I call compulsive. We might find somebody's nervous tic annoying, and we might sympathise with him for having this disposition. But such behaviour would not tend to attract our resentment or admiration. Similarly, somebody who is often overcome with an urge to eat a whole trifle late at night deserves our help and compassion, but not our resentment (providing of course one didn't have an eye for it oneself).

Other compulsions are more contentious. Consider smoking. We may loathe the practice, and resent anybody who does it. Similarly for compulsive tidying. Does this mean that compulsions should also be counted as responsible? I do not think so. However, at this point we need to dig deeper into these examples to see what supports these attitudes. This will allow us to make a stipulation which allows us to say that we should not have reactive attitudes to the victims of compulsions.

Return to the smoking example. Consider how our reactive attitudes might change if we learn that a particular smoker has done all they can to give up, tried the patches, the gum, hypnosis, and so forth, and still cannot give it up. When informed of circumstances like this our reactive attitudes tend to subside. What this brings out is something important about what governs our reactive attitudes. If the agent

literally "can't do anything about" behaving in some way, we tend not to hold them responsible for it.

The sorts of things that we can be said to "do about" our habitual behaviours can vary considerably. Something might be done at two levels: at the level of the particular behaviour, when the agent might decide to do otherwise, which may not be possible in cases of genuine compulsion; or at the level of the habit or compulsion itself, when the agent might adopt strategies to eradicate the tendency to indulge in such behaviour.[6] But whilst the success of such strategies, which might include a change of scene, counselling or hypnosis, is unlikely to be solely in the hands of the agent, she is likely to be in a position to put such strategies in train (in modern Western society at least). To that very minimal extent she may be said to attract the reactive attitudes. In §8 I shall be in a position to explain these two levels of reactive attitudes for habits.

Deciding on which attitudes make sense is a delicate matter, and requires close attention to the details of the particular case. But what we are looking for in each case seems reasonably clear. Insofar as we know that somebody can't do anything about a habitual behaviour, manifestations of that behaviour do not attract the reactive attitudes, and we can surmise that they are not personally responsible for these manifestations. The suggestion is that these are the compulsions. In contrast, insofar as the agent is in a position to do something about an instance of habitual behaviour, and does not, they do attract the reactive attitudes, and hence seem to be behaviours for which the agent is personally responsible. They are the candidates for being habitual actions.

Thus the idea of being able to do something about a behaviour gives us a way of distinguishing compulsions from other habitual behaviours. We could say that if a habitual behaviour is such that the agent cannot do anything about its exercises, it should be classed as a compulsion. Otherwise we should class it as an action. This would give us a basis for our intuitive distinction between these things.

---

[6] We saw how losing a habit can be understood in Chapter 3, §8.

Even given such a distinction, there will be behaviours which are close to the borderline between compulsions and habitual actions. Sometimes we will not be able to tell if a kind of behaviour is a habitual action or a compulsion until on some occasion we try to stop doing it. Addictions commonly work like this. For other behaviours, we might find that we can resist on some occasions, whilst on others we cannot (my nail-biting *might* be such an example, though in some moods I won't admit it). There is no need to be any more stipulative here. For the principle is clear in particular cases. If the agent genuinely cannot do anything about the particular habitual behaviour, it is to be classed as a compulsion, and not an action.

What the reactive attitudes test suggests is that the idea of "being able to do something about" a behaviour is at the heart of our intuitions about personal responsibility. In the next two sections I want to support the idea that we have personal responsibility for non-compulsive habitual behaviours by developing more precisely what "being able to do something about it" amounts to. To this end I shall develop a specific notion of control which captures this currently rough thought.

5. *Control of Habitual Behaviours*

I want to develop the idea of a kind of control that we can be said to have over non-compulsive habitual behaviours which I call *rational intervention control.* I claim that this kind of control is sufficient for personal responsibility. It follows that if actions are those behaviours for which agents are personally responsible, non-compulsive habitual behaviours are actions.

I begin with some suggestions about a kind of control that we might have over habitual behaviours which I find in Aristotle and Ryle. It has two aspects, involving both intervention and rationality. In this section and the next I develop the more generic idea of intervention control and tackle a possibly fatal objection to the claim that this kind of control grounds responsibility made by Fischer and Ravizza (1998). In §7 I explain the significance of saying that intervention control is "rational".

In Chapter 2 I suggested that virtuous action should be thought of as both automatic and responsible. In §3 of that chapter I said that virtuous actions count as responsible ("virtue in the strict sense", as opposed to "natural virtue") because

Aristotle holds that virtue "is not merely the state in accordance with right reason, but the state that implies the *presence* of right reason" (*NE* 6.13, 1144b26-7). What is important here is that the "presence of right reason" allows us to say that a virtuous action is responsible without having to give up on the idea that it is automatic. For to say that the faculty of reason is "present" is not to say that it reasoning actually takes place. The suggestion then is that reason has some controlling function which accounts for virtuous actions being responsible, but that control need not be directly exercised. Reason controls whilst being, in some sense, "in the background". But it is not obvious how something that is merely "in the background" could be said to control anything.

We get a clue from Ryle in his account of intelligent capacities (see Chapter 2, §5). Ryle's idea is that when an agent exercises an intelligent capacity she "minds" what she is doing, by exercising "care, vigilance, or criticism" (1949, p. 42). Now whilst this might be said to bear some resemblance to Aristotle's idea of the presence of right reason, reason being present in the form of the agent's care, vigilance and criticism, I think Ryle is adding two insights. First, Ryle notices something special about a variety of intelligent human behaviour. Namely, manifest in that behaviour is a kind of watchfulness over, or attentiveness to, it. This idea of watchfulness is the idea of something in the background, as it were, paying attention, albeit entirely passively, rather as a shepherd can watch over his sheep.

But secondly, Ryle writes of the "exercise" of care, vigilance or criticism, and this implies that some kind of intervention would be made were the relevant circumstances to arise. In the same way, the shepherd would rescue a sheep were it to stray. There is an active aspect to these notions too. The exercise of an intelligent capacity, then, can be said to be controlled, because the rational faculty is not just in the background, but is attentive, and ready to intervene in the behaviour, as and when necessary. This, in outline, is what I call rational intervention control. And I argue in the next section that we have this kind of control over our non-compulsive habitual behaviours.

6. *Intervention Control*

Harry Frankfurt (1978) notices the significance of intervention:

> A driver whose automobile is coasting downhill in virtue of gravitational forces alone may be entirely satisfied with its speed and direction, and so he may never intervene to adjust its movement in any way. This would not show that that the movement of the automobile did not occur under his guidance. What counts is that he was prepared to intervene if necessary, and that he was in a position to do so more or less effectively. Similarly, the causal mechanisms which stand ready to affect the course of a bodily movement may never have occasion to do so; for no negative feedback of the sort that would trigger their compensatory activity may occur. The behaviour is purposive not because it results from causes of a certain kind, but because it would be affected by certain causes if the accomplishment of its course were to be jeopardised. (Frankfurt 1978, p. 75)[7]

Frankfurt's passage nicely illustrates the intuition that the *possibility* of intervention, as opposed to actual intervention, is all that is needed for control, and hence responsibility, for a piece of behaviour. This is the case for the behaviour of the car as well as the bodily behaviour. According to Frankfurt, it can be correct to say that the driver drives the car down the hill "purposefully", and this entails that it is he that does it, even though he may not have initiated the coasting, and did not actually intervene in it. Similarly, it might also be correct to say that an agent's behaviour is "purposive", she does it, regardless of how her bodily movements might have been caused, and irrespective of the fact that she may not have actually intervened in those movements.

Now in the light of the considerations canvassed earlier, we might baulk at Frankfurt's suggestion that the behaviour in question is "purposive". For this may carry with it the suggestion that what we have here is a kind of intentional action, and one might then argue, as Mele (1997, pp. 8-11) does, that it can be incorporated under standard analyses. But we need not read so much into the term as this. One thing I

---

[7] We can remain agnostic about Frankfurt's overtly causal interpretation of what it means to intervene in a process.

think Frankfurt is getting at by saying that the behaviour is "purposive" is that an agent can be behind a behaviour without intervening in it at all, and that is all we need.

The idea of control in play here seems to fit in rather well with the idea of a habitual behaviour. Because habitual behaviour is always automatic, there is no question that it has been initiated by an act of thinking (deliberation or trying). However, if something unexpected happens, or if the agent so chooses, she will become more aware of what she is doing, and at this point may or may not intervene. If she intervenes her behaviour is no longer automatic, and so ceases to be properly habitual. However, when none of these contingencies arise, i.e. the circumstances are normal, and she does not intervene, the behaviour remains habitual. The control over her behaviour consists in her being able to intervene should it be necessary. And if my understanding of Frankfurt's intuition is right, this is enough for her to be personally responsible for it.

But there is dissent. In their recent book *Responsibility and Control* (1998), John Fischer and Mark Ravizza agree that some kind of control is necessary for responsibility.[8] They write:

> It seems to be a basic presupposition embedded in the way we think about these matters that an agent must in some sense control his behaviour in order to be morally responsible [for it]. (pp. 13-14)[9]

However, Fischer and Ravizza argue that the kind of control in question could not be the intervention control Frankfurt describes in the passage above. If they are right, this may be a threat to the idea that intervention control is sufficient for personal responsibility. If we look at their argument, we see that there is such a threat.

---

[8] In the remainder of this section page numbers will be to this work unless otherwise stated.

[9] Not much turns on Fischer and Ravizza's use of the term "morally". They employ it primarily to contrast this kind of responsibility with causal responsibility (pp. 1-2, and n. 1). For our purposes, then, we can take their notion of "moral" responsibility to be equivalent to our idea of personal responsibility.

Fischer and Ravizza are impressed by counter examples to what is called "the principle of alternate possibilities", which states that an agent is only morally responsible for a piece of behaviour if she could have done otherwise. The principle originates from an earlier paper by Frankfurt (1969). In that paper he famously argues against the principle on the grounds that there may be no alternative (for Frankfurt, "alternate") possibilities "genuinely open" to an agent, yet she can still be held morally responsible for what she does.[10] Frankfurt invents a number of (increasingly bizarre) scenarios in which various outside agencies would have intervened had the agent tried to do otherwise, and made things turn out in the same way anyway. These scenarios involve coercion, brain manipulation, mad scientists, and the like. In such cases, argues Frankfurt, no alternatives are genuinely open to the agent. However, Frankfurt's intuitions are that the agent is nevertheless responsible for what she does. If this is right, the principle of alternate possibilities is shown to be false, and we must look somewhere other than to counterfactual scenarios to ground responsibility.[11]

It is this that motivates Fischer and Ravizza to introduce a distinction between *guidance control* and *regulative control* by means of the following example:

> Let us suppose that Sally is driving her car. It is functioning well, and Sally wishes to make a right turn. As a result of her intention to turn right, she signals, turns the steering wheel, and carefully guides the car to the right. ... She controls the car, and also has a certain sort of control *over* the car's movements. Insofar as Sally actually guides the car in a certain way, we shall say that she has "guidance control". Further, insofar as Sally also has the power to guide the car in a different way, we shall say that she has "regulative control." (pp. 30-1)

---

[10] "Genuinely open" is Fischer and Ravizza's phrase (1998, p. 30).

[11] There may be a serious tension between this result and the claim of Frankfurt's later (1978) paper referred to above, since in that paper he seems to be relying on alternate possibilities (the possibility of the driver intervening in the movement of the coasting car) in his account of "purposive" action. The coherence of his overall position may thus rely on some sort of distinction between two sorts of control, like that made by Fischer and Ravizza, discussed and criticised below.

Let us suppose for the sake of argument, that regulative control is equivalent to our intervention control.

Fischer and Ravizza claim that the two kinds of control "can at least in principle pull apart" (p. 32). They show this by extending the example so that Sally's car has dual controls, and an instructor at the other controls, "who is quite happy to allow Sally to steer the car to the right, but that if Sally had shown any inclination to cause the car to go in some other direction, the instructor would have intervened and caused the car to go to the right (just as it actually goes)" (p. 32). This is structurally equivalent to one of Frankfurt's cases. Thus, as Sally has guidance control over the car, but not regulative control - the instructor has that - and since, following Frankfurt, Sally seems still to be responsible for the car's movements, they reason that "the sort of control necessarily associated with moral responsibility for action is *guidance control*" (p. 33).[12] They conclude that "Guidance control, and not regulative control, is the control that is associated with moral responsibility" (p. 34), and they thereby avoid any reliance on the principle of alternate possibilities. If this argument is sound it amounts to a refutation of my claim that intervention control can be sufficient for responsibility.

But there are a number of ways in which we can resist this conclusion. First consider the instructor case, which is Fischer and Ravizza's only example of the two kinds of control coming apart. We could refuse to accept the analogy between the instructor case and cases of habitual behaviour. In the instructor case there are two agents; in habitual behaviour there is only one. And it is not clear that a single agent could ever have guidance control without also having regulative control. What is telling is how Fischer and Ravizza characterise guidance control, that is, in terms of Sally "carefully" (p. 30) guiding the car to the right. A natural understanding of this is that Sally is attentive to feedback (both visual and through the steering wheel), and would in the light of such feedback, make corrections if necessary. In other words guidance control entails some regulative control, so alternate possibilities are still in the picture.

---

[12] On the use of "moral" see n. 9 above.

This objection may of course reveal that I have misunderstood what Fischer and Ravizza mean by guidance control - unfortunately they define the idea solely by means of the single example. But it does not look as if there will be an easy fix to this criticism. For if guidance control involves no regulative control at all, the claim that guidance control alone underwrites responsibility looks far less plausible.[13]

But even if we accept the analogy with a two agent scenario, in the light of intuitions like Frankfurt's, we can deny Fischer and Ravizza's intuition that Sally, and not the instructor, is responsible for the car's going to the right. True, it is not exactly obvious what we would say in such an unusual situation, but it would certainly be odd - or symptomatic of being in the grip of a theory - to deny flatly that the instructor has *any* responsibility for what happens to the car. The instructor will, after all, earn some of our reactive attitudes for the car's movements going so smoothly. It is these same kinds of attitudes that will become transferred to Sally herself once she learns to drive, and these attitudes that explain why we are more impressed with her cornering when she does it on her own.

Perhaps there are replies to these objections. If there are, there is a nettle still to be grasped. For if we are going to rely on the idea of regulative control, or some other notion of control which requires the idea of alternate possibilities, to explain our responsibility for non-compulsive habitual behaviours as I suggest, we shall need to show that this notion of control is not subject to Frankfurt-type counter examples. How is this to be done?

I have two answers. First, we can maintain that none of the Frankfurt-type cases represent the *kind* of intervention that we need to invoke to secure responsibility for habitual behaviour.[14] The kind of intervention that I think is necessary is the intervention *of the agent*. So, for instance, the possibility that *I could make a decision* not to bite my nails now, not to go the usual way to work today, and so forth, is all that is needed for intervention control. These are the counterfactual scenarios that do the

---

[13] This objection could also be made to Frankfurt's overall position, see n. 11 above.

[14] This argument has similarities with Davidson's (1973, pp. 74-5) reply to Frankfurt, though it does not rely on Davidson's idea of "attitudinal conditions".

work in grounding responsibility. In contrast, the kind of intervention that I do not think is necessary to ground responsibility is the intervention of outside agencies, mad scientists, aliens, omnipotent gods and driving instructors, which fill the literature on alternate possibilities. It seems to me that if any of these are genuine possibilities, then it would be futile to try to say that they are what ground the agent's responsibility. For it would seem far more natural to say that it is precisely *because* these are possibilities of *alien* intervention that the agent has *less* control over her behaviour than she might have thought she had. It is the aliens that have the control; not her.

The second answer is structurally similar to the argument against the presence of "trying" in Chapter 3, §7. The idea is that none of the behaviours in Frankfurt-type scenarios could count as habitual behaviours, since in every case, the circumstances in which the behaviours are supposed to take place are abnormal. This is what the definition of habitual behaviour (HB) implies, and the move also enjoys intuitive plausibility. With outside agencies wired up to an agent, poised and ready to intervene, severe pressure is put on the supposition that she is involved in anything repeated, never mind habitual. In contrast, it would be quite usual for the agent herself to be in a position to intervene should that be appropriate or desirable. Thus Frankfurt-type cases are not counter examples to the idea that we have intervention control over habitual behaviours, for they presuppose that the behaviours in question are non-habitual. Our willingness to be impressed by thought experiments of this kind may reveal how deep our habit blindness goes.[15]

In sum, I do not think that Fischer and Ravizza have shown that the only kind of control relevant to responsibility is guidance control. And I do not think that we should be as impressed as they are with the Frankfurt-type cases that motivate their claim, both in general, and in the particular context of habitual behaviours. Since we have seen no good counter argument, it still seems entirely plausible that intervention control is sufficient for personal responsibility. In the next section I shall develop the

---

[15] Questions are raised here about the use of thought experiments in the philosophy of mind and action, suggesting a further avenue for future research.

notion of intervention control by spelling out the sense in which intervention is rational.

7. The Rationality of Intervention Control

I want to say that we can understand the particular kind of intervention relevant to the control of non-coercive habitual behaviours by saying that it must be *rational* intervention. Though I shall not be addressing the topic of the rationality of habitual behaviour properly until Chapters 5 and 6, it is necessary to broach the topic of rationality here in order to argue that intervention control is sufficient for personal responsibility. For the purposes of the current argument we can understand rational intervention as intervention "for reasons", an idea of rational action which is well embedded in the literature.[16] We can for now understand that phrase simply as indicating actions that are sensible things to do from the point of view of the agent.

As we saw in the previous section there are lots of ways of intervening in the progress of a piece of behaviour, and not just any kind of intervention will show personal responsibility. Possible interventions include being restrained or pushed by things in one's environment, being restrained or coerced by somebody else, being overcome by tiredness, being anaesthetised, having one's brain manipulated, and so on. None of these are rational interventions, however, because they are not done for reasons. This relation to an agent's reasons gives us a clear sense in which an intervention may or may not be authored by the agent. When an agent intervenes by acting for reasons she is the author of that intervention. When she does not, as in the above interventions, she is not. The idea is that insofar as a piece of behaviour can be intervened upon by an agent in this way, that agent is author of the behaviour, whether or not she actually intervenes. I take this to be consistent with Frankfurt's intuitions about the coasting of the car.

---

[16] As we shall see in Chapter 5, actions for reasons are thought by many to exhaust the class of rational actions. This is something I challenge in Chapter 6. I argue there that rational action need not be understood as actions for reasons, for many habitual actions are rational in another sense. This leaves open the possibility that rational intervention may consist in exercising other habits. I take this up in Chapter 8.

We can now give a definition of rational intervention control.

(RIC) An agent has *rational intervention control* over a behaviour if and only if she has both a capacity to notice when an intervention on that behaviour is rationally appropriate, and a capacity to intervene appropriately on such occasions.

I want to spell out the idea of rational intervention in a little more detail by distinguishing between two sorts of intervention that may take place when exercising a habit. I call them *minor interventions* and *major interventions*. Minor interventions are adjustments of the behaviour, whilst continuing to exercise that habit. Major interventions involve abandoning that habitual behaviour altogether, so it is no longer an exercise of that habit.

Consider my habitual behaviour of walking to work along a certain route every day. Reasons for minor interventions include the fact that it is raining, which should prompt me to take an umbrella; a crowd of people standing in my usual path, which means I should walk around them; and there being large volumes of traffic at a road crossing, which just means I have to wait for longer than usual. Some of these sorts of interventions may be done quite subconsciously, by the body alone. Others could be described as rational, in the sense that I could be said to have such reasons to intervene. However such interventions are understood, they do not compromise the status of my behaviour as an exercise of *this particular habit*. For the circumstances within which these interventions are made fall within the characterisation of normal circumstances specified in Chapter 3, §4. Any deliberation involving such reasons would not typically be deliberation about *whether* to exercise the habit or not, but how to adjust things appropriately in this particular case. Hence it is in virtue of the possibility of minor interventions that habitual behaviours can be said to be monitored, and corrected by, our deliberative faculty, without that faculty necessarily being directly involved.

Examples of reasons for major interventions include the fact that due to floods all work is cancelled; my breaking my leg; some life-threatening situation at home; that I wanted a change; and that I wanted to lose the habit. If I came to have any of these reasons, acting on them would be to intervene in such a way that I would not

exercise my habit at all today.[17] The sorts of circumstances in play here are relevant to the habitual behaviour in prospect, since they would all be reasons for me not to perform it. So an attempt to persist in performing the behaviour, perhaps as a result of not having noticed the strangeness of the circumstances, or having noticed this but not really thought properly about the implications, I would be doing something irrational, since I would be failing to respond appropriately to reasons. Thus we can see that intervention control works at a second level, that is, in monitoring significant abnormalities in the circumstances, and opting out in appropriate ways. A failure to do so would render our otherwise automatic attempts to behave habitually irrational.[18]

Thus we can see that the two capacities involved in rational intervention control can both be present at two levels corresponding to minor and major interventions. Minor interventions occur when the agent adjusts what she is doing to the particular, though normal, circumstances, and continues to carry out her habitual behaviour; and major interventions, when the circumstances are abnormal, and the agent opts out of the habitual behaviour altogether. Thus we can see that failures in either capacity, that is, to fail to notice when an intervention is required, or to fail to intervene appropriately, in a piece of habitual behaviour, is a failure in rationality. Were I to go out without my umbrella when it's raining, "blindly" walk into the crowd of people, try to cross the line of moving traffic, ignore the radio, or even my own wishes, at best I could be accused of being absent minded; at worst stupid or rude. These kinds of failures would be culpable irrationality - I would be responsible for these mistakes. It would be no justification for me to say that "after all, it was a habit of mine!", but an excuse, and a rather feeble one at that. I would be guilty of failing to pay proper care and attention to what I am doing. We call this kind of failure *negligence*.[19]

---

[17] All are ways in which I might end up losing the habit, as the last reason makes explicit.

[18] I shall say more about the rationality of major interventions in Chapter 6, §6.

[19] Note that if I genuinely was not able to stop myself from ignoring these reasons, I would still arguably be manifesting irrationality, though in a different way. As we saw in §4, the inability would indicate that

I think that rational intervention control is sufficient for personal responsibility because it is the agent herself who needs to be attentive enough to spot any reasons not to do it; and it is the agent herself who would have to act otherwise on pain of the charge of negligence.[20]

I think we can find support for this claim in some literature on intentional actions. For some accounts of intentional actions draw on the idea of interventions to solve problems about where exactly, in spatio-temporal terms, an action begins and where it ends.[21] For instance, Stout (1996, pp. 31-2) takes up Adam Morton's (1975, p. 14) suggestion that "Intentional action is action that is guided by information to which it is responsive". The idea is that only whilst the guidance continues can the behaviour be deemed an action.

However, both writers propose views according to which actions also aim at some outcome, and are thereby intentional. And I don't think we can in general say that sort of thing about habitual behaviours. When we behave habitually we just do what is usual for us in this context. What is more, we seem to be able to detect what is unusual in our surroundings, and correct for it, without having any conception of what we were doing in the first place. We revert to what we were doing, whatever that was. In this way we can still, as Stout puts it, "make a good shot at it" (1996, p. 31). Thus we can have the idea of being guided by information, in the form of rational intervention control, without thereby conceding that habitual behaviours are intentional. Adopting Morton's scheme with this modification, there need be no problem with determining when an agent is acting: it will be for the duration of the agent's having rational intervention control over her behaviour.

This characterisation of control has an interesting, and perhaps surprising, upshot, which will become significant later. For nowhere in the above definition have

---

we are dealing with a case of compulsion, not habit. We can see compulsion as a particularly unfortunate kind of irrationality, which would not be culpable (or at least not in the same way).

[20] I shall say more about the irrationality of inattentiveness in Chapter 6, §5.

[21] There is a helpful discussion of this issue in Dretske (1988, pp. 11-22). Stout (1996, pp. 29-32) offers a brief, but helpful synopsis of the positions.

we needed to specify what kind of behaviour is actually being controlled, or "under control". This means that agents could be said to have rational intervention control over occurrences whether or not they involve the agent's own body. Frankfurt's coasting car example is not merely an analogy for the kind of control in question, but a genuine case of it. The driver has rational intervention control over the movement of his car down the hill, even though he may have his hands off the steering wheel, and his feet off the pedals. Similarly, a footballer can be said to have rational intervention control over the movement of a ball which he "guards" as he runs with it, though he need not make contact with it. And a child-minder has rational intervention control over his children even though he might be sitting in the same room reading a book, or even just listening in another room on a baby alarm. As long as the agent is attentive and can intervene if necessary (which may require technological support) they can be said to be in control of any given occurrence.

This is not to deny that behaviours which involve the agent's own body could be said to enjoy a special status. Indeed, in view of our characterisation of behaviour (Chapter 3, §1), habitual behaviours always do involve the agent's body. The point is that the involvement of the agent's own body in the behaviour is not essential to her being personally responsible for it. What follows is that an agent can have rational intervention control over a piece of behaviour whether or not that behaviour is itself rational in the sense that it is done for her reasons. This will become important in the next two chapters.

8. Control of Habits

In §4 I suggested that the reactive attitudes, when applied to habitual behaviours, are appropriate at "two levels": at the level of the particular *exercise* of the habit; and at the level of the habit itself. Given that we have rational intervention control over particular habitual behaviours, we can now add something about the sort of control - and hence responsibility - we have over habits themselves. For we can now be said to control our habits in two ways: we have control over which new ones we learn, and, to some extent, over which ones we lose.

In Chapter 3, §8, I described how we can acquire habits intentionally, that is, by deciding to practise the kind of behaviour we would like to become automatic in contexts of the appropriate sort. But whilst the prospect of automatism may motivate us to intentionally acquire habits, "programming" ourselves in this way has its dangers. For although we have rational intervention control over our non-compulsive habitual behaviours, so one is never completely at the mercy of the "program" (providing of course that the habit does not become a compulsion), this requires continued vigilance. Better then to acquire good habits than bad. But gaining habits, good or bad, is nevertheless one way in which we can intentionally adjust our own future behaviour, and hence such behaviour is something for which we are *doubly* responsible.

Because we have rational intervention control over non-compulsive habitual behaviours, we can also, to some extent, control which habits we lose. Providing an agent knows that habits can be lost through retraining, and has reason to lose a particular habit, it is the fact that she has rational intervention control over her habit's exercises that makes a certain sort of habit loss possible, that is, when an agent has no choice but to be presented with the normal circumstances. Such an agent could always decide to intervene in a particular habitual behaviour, and do something else, or nothing at all. Doing this repeatedly in normal circumstances for that habit will help. But actually losing a habit - for it to cease to be automatic - will often take considerable perseverance, and, depending on our age (apparently due to the reduced "plasticity" of the brain), it may never happen.

Thus it is because we enjoy rational intervention control over particular habitual behaviours that we have a certain amount of control over which habits we have. One implication is that in the mature and self-reflective agent, there ought to be a good deal of harmony between the habits one has, and the habits it would, on reflection, be good to have. It ought to be no surprise that for such agents, many exercises of habits turn out to meet precisely the same standards of correctness as those actions which flow from careful deliberation. There is a kind of harmony between habit and reason. But as we shall see in Chapter 5, to blur the distinction is a mistake.

For the purposes of this chapter, we now have not one, but two, justifications for saying that non-compulsive habitual behaviours are *the agent's* behaviours, and are thus marked out from compulsions. First, we can say that non-compulsive habitual behaviours are the agent's because it is the agent herself who has rational intervention control over them in particular cases. Second, we can say that non-compulsive habitual behaviours are the agent's because it is the individual agent herself who has cultivated (in the case of intentionally acquired habits), or refrained from losing (in the case of unintentionally or non-intentionally acquired habits), the habits of which they are exercises.

There are at least the following two implications. First, both justifications draw on a conception of an agent as a being extended in time, with a rich history and a repertoire of capacities, which are not all manifest at any one moment. Accounts of personal identity may be affected. Second, this more complex picture of personal responsibility can help to explain some of the limits that exist in our willingness to hold, and revise, our reactive attitudes to one another, as we negotiate the fine line between compulsion and action.

9. Habitual Actions

In this chapter I have developed the idea of a kind of control, rational intervention control, which we have over particular habitual behaviours. I have defended the claim that an agent's having this kind of control over a behaviour is sufficient for them to be personally responsible for it. Not only that, but we also have some degree of control, and hence responsibility for, which habits we have. I conclude that we are now justified in calling habitual behaviours habitual *actions*.

The complete definition of the notion of a habitual action can now be stated:

(HA) A behaviour ϕ is a *habitual action* if and only if it is repeated according to (R), automatic according to (A), and the agent has rational intervention control over ϕ.

The corresponding definition of a habit is as follows:

(H) An agent has a *habit of ϕ-ing in C* if and only if the agent performed habitual action ϕ on most of her recent encounters with C.

I anticipate (but will not show) that these two definitions capture a considerable portion of those human behaviours, and kinds of behaviour, that we commonly call "habitual actions" or "habits".

My starting point was to try to capture something generic in the anti-intellectualist ideas put forward by Aristotle, Ryle and Wittgenstein. Whether or not I have completely succeeded in that endeavour I am not sure. It would require further exegesis and argument in each case, armed with our new definitions. But regardless of whether I have precisely captured a common conceptual core, the true test for the notion of a habitual action will be whether it can be employed to cure contemporary intellectualisms. But before we can go into that issue, we need to make good a crucial claim about such actions. Namely, that habitual actions themselves, and not their interventions, represent a distinctive class of *rational* actions. That is the job of the next two chapters.

5

ACTIONS FOR REASONS

1. Rational Actions

In the previous chapter we arrived at the definition of habitual actions by abandoning standard analyses of action, which, I suggested, may have been prejudiced against habits. But abandoning the standard analyses comes at a cost. For those analyses tend to license us to say that actions for which we are personally responsible are also *rational.* Indeed many analyses simply assume that action-hood and rationality come and go together. If we are going to admit habitual actions, or any subset of them, into the category of rational actions, then, we have more work to do in showing they belong there.

One might at this point want to say that we have already come across a sense in which habitual actions can be said to be rational. This sense arises from the kind of control that we have over them, described in the previous chapter (§7). One might say that because we have rational intervention control over habitual actions, that in itself provides a sense in which they are rational. But in fact, all that follows from our having this kind of control over our habitual actions is that the counterfactual scenarios are rational, that is, when the agent intervenes for reasons. It does not follow that the actual scenarios of habitual action, in which no such intervention occurs, are rational. What we still need to do then, is to identify a sense, or senses, in which actual habitual actions are rational.

The obvious approach is to adopt a widely held assumption about what it means for an action to be rational, namely, that all rational actions are actions for reasons. We met the idea of actions for reasons in the last chapter (§7), albeit briefly. I said there that actions for reasons can be thought of as sensible things to do from the point of view of the agent. As we shall see there are different ways in which such a locution can be cashed out, but the central idea is uncontroversial. This assumption

about rational actions can be traced at least as far back as Anscombe's (1957) *Intention*.[1] And perhaps because the view has such a pedigree, contemporary philosophers of action, with few exceptions, consider it a platitude.

Nevertheless I consider the above assumption to be a substantive claim, and a contentious one at that. Hence I give it a name. I shall call the view that all rational actions are actions for reasons *the reasons theory of rational action* (or just the *reasons theory*). In this chapter, however, I shall assume, with the majority, that the reasons theory is correct. Hence for the moment the question of whether habitual actions are rational just is the question of whether habitual actions are actions for reasons. A major factor in my thinking that the reasons theory is contentious is that, as I shall show, the answer to this question comes out negative.

In answering this question we must keep in mind the importance of the descriptions of actions in determining their status, whether it be rational or habitual. We noted in Chapter 4, §2 that actions are only intentional under certain descriptions. The same thing goes for an action's being rational.[2] Under the description "brushing my teeth", my action may be done for my reasons, whereas under the description "wearing the brush out", it probably will not. Similarly for habituality. We saw in Chapter 3, §3 that behaviours are only repeated under certain descriptions, but not others. And because habitual actions are a kind of repeated behaviour, the same thing applies to them. Described as "brushing my teeth", my action may be habitual. But described as "occupying my mouth", it is doubtful that it is, since there are no characteristic circumstances in which I do *this*.

---

[1] Anscombe writes: "What distinguishes actions which are intentional from actions which are not? The answer I shall suggest is that they are the actions to which a certain sense of the question 'Why?' is given application; the sense is of course that in which the answer, if positive, gives a reason for acting" (1957, p. 9). Whilst Anscombe has certainly been influential in bringing about the orthodoxy, it is notable that she does not say that intentional action *must* be actions for reasons. Unlike many who are influenced by her, she leaves room for other analyses.

[2] Indeed, for some, e.g. Davidson (1980, p. 61) an action's being rational and it's being intentional comes to the same thing.

Of course the question of whether such alternative descriptions genuinely pick out one and the same action, is highly contentious. To say that they do may require some underlying ontology *a la* Davidson. For reasons given in Chapter 1, this is something I think we should avoid. If we instead assume that no identity can be established, the question to be addressed must be whether habitual actions, *under their description as habitual,* are rational, *under their description as actions for reasons.* This then is how I shall proceed.

But still matters are hardly simple. For when we come to look at the literature on actions for reasons, we find there is little in the way of agreement about what this idea amounts to. The lack of consensus is not just a problem for the question at hand. It also means that the counterfactual scenarios upon which the idea of intervention control relies may require further clarification. Nevertheless, as for any disagreement, there does seem to be common ground, and I shall begin in §2 by canvassing popular views on the topic, and giving a characterisation neutral enough to be acceptable to all holders of the reasons theory. I shall then divide going accounts of actions for reasons into two groups, according to what kind of thing they take a "reason" to be. I call these views respectively *reasons internalism* and *reasons externalism.* In each case I consider whether habitual actions could be actions for reasons on this construal of a reason. In §§3-5 I describe reasons internalism and say why I think that holders of that view have difficulty in accommodating habitual actions. In §6 I describe two well-known problems with reasons internalism. This not only motivates the opposing position, reasons externalism, but also explains why I reject reasons internalism as an account of the counterfactual scenarios. But whilst reasons externalism might avoid internalist difficulties, it remains, I argue in §§7-9, difficult to accommodate habitual actions as done for reasons in this sense. I shall conclude, in §10, that whilst reasons externalism may represent a preferable account of reasons, neither view of reasons can accommodate habitual actions as actions for reasons.

2. *The Reasons Theory*

The idea of "a reason for which an agent acts" is technically defined. Such reasons must, as Davidson first said, *rationalize* the action:

> A reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action - some feature, consequence, or aspect of the action the agent wanted, desired, prized, held dear, thought dutiful, beneficial, obligatory, or agreeable. (Davidson 1963, p. 3)

The term *rationalization* has been adopted in the literature ever since to express the relation between reason and action.[3] The notion of a reason expressed here is well embedded in the literature, and we don't need to agree with very much of what Davidson says to accept its intuitive appeal. Specifically, and with due respect to other reasons theorists, we need not accept that reasons can always be restated in terms of a pro attitude plus a means-end belief, nor that this pair is causally related to the action. For current purposes we can be neutral on such matters.

Slightly different locutions have been employed, which I take to be equivalent. Here, for instance is McDowell:

> A full specification of a reason must make clear how the reason was capable of motivation; it must contain enough to reveal the favourable light in which the agent saw his projected action. (McDowell 1978, p. 79)

Here we have the idea that the reason is capable of "motivation", in virtue of its revealing "the favourable light" in which the agent saw his action. Both these terms enjoy currency today. Here, for instance, is Dancy in a recent book:

> normally there will be, for each action, the reasons in the light of which the agent did that action, which we can think of as what persuaded him to do it. When we think in terms of reasons in this way, we think of them as *motivating*. (Dancy 2000, p. 1)[4]

---

[3] Robert Audi (1985) has suggested that in our everyday methods of making sense of actions, the rationalization of an action is often *contrasted* with its proper explanation. I shall follow philosophical convention in making nothing of such a contrast (if it exists), and regard rationalization as a species of explanation.

[4] The context here is Dancy introducing the distinction between "motivating reasons" and "normative reasons", a distinction that Michael Smith (1994) in particular has made popular in recent years. There are a number of factors which lead me to think that the distinction is unhelpful, not least of which is Dancy's own denial that we are here talking about two sorts of reason, "There are not. There are just two

Reasons, then, can be thought of as items which not only explain an action, but show it to be *justified from the agent's point of view.* We can gather from the above characterisations that the idea of justification is meant to be weak, which is why it is sometimes glossed as "sensible". It certainly falls short of *moral* justification. But the agent nevertheless has a sense of why she should do it.

This characterisation marks reasons out from at least two other sorts of things that we might naturally call "reasons". There might, for instance, be a number of considerations which could justify an action, but as it happens, the agent is not aware of them. So the fact that I would meet an old friend was a good reason for my going for a walk, though this fact played no part in my decision to do so. We can follow Dancy (2000, p. 1) in calling these "good reasons". There may be many good reasons which are not the reasons *for which* an agent acts, the latter being within the agent's purview when she acts.

A second sort of item which we would ordinarily call a "reason", but which does not rationalize in the sense required, are those factors which explain an action, but do not justify it. So we might say that the reason that I fell asleep was that I had taken a sleeping pill. Such reasons do not justify what I do, though they may help us to understand it. Again following Dancy (2000, pp. 5-6) we can distinguish these non-justifying reasons from those for which agents act by calling them "reasons why".

The reasons theory of rational action would hardly be compelling at all, however, if we did not accept one important qualification concerning the phenomenology of actions, whether or not they are habitual. More often than not, just why we should perform some action hardly ever goes through our minds either shortly before, or during, its performance. And this need not be because it is something we have done repeatedly. So, for instance, we might at some earlier time, have made a plan about what to do, and the present action may be a part of implementing that plan. Any number of considerations could have been before our

---

questions that we use the single notion of a reason to answer" (2000, p. 2). Dancy also seems to undermine the case for this taxonomy in his (1995b) paper in which he argues that "there is really no such thing as the theory of motivation". See also Norman (2001) for a rejection of the distinction.

minds at the planning stage, but we are unlikely to be rethinking these thoughts when we come to execute the plan.

We have of course come across unthinking actions before, in Chapter 3, §§6-7, when we defined the sense in which habitual actions are automatic. It would be convenient if we could say that non-habitual unthought actions were automatic in the same sense. But I don't think we can. There is a much stronger case for saying that the agent tries when she acts non-habitually. Certainly, my argument against the presence of "trying" does not go through if the action in question is not repeated, since the idea of normal circumstances used there does not apply. And as we shall see, I suspect that there may be a sense in which reasons theorists would like to say that some sort of deliberation does in fact accompany non-habitual actions, albeit of a sub-conscious variety. So I think it would be a mistake to assume that non-habitual actions are automatic. But in terms of understanding the reasons theory, we need to bear in mind that for many actions, whether habitual or not, there are no phenomenological grounds for ascribing reasons to the agent.

This is not to say that we are never aware of what justifies a non-habitual action when we perform it. Typically this will happen when we are not sure about whether to do something, or where the consequences of an action are particularly significant or important to us, or both. In such cases, providing there is time, we will quite properly deliberate. But even given that deliberation may only be partial, actions preceded by such a process are relatively few and far between. Thus a fair reading of reasons theorists is that they do not deem it necessary for reasons to be within the agent's consciousness at the time of acting. If this were a requirement it would turn out that hardly any of our actions were rational, never mind the habitual ones. As an account of rational agency, this would be something of a paradox.[5]

The reasons theorist will nevertheless want us to believe that for all rational actions, the reasons are in some sense "present" to the agent when she acts. They cannot be completely external to the agent's subjectivity, or the reasons theorist would be unable to say that this action was intentional under a description at all, and this in

---

[5] I return to this paradox in Chapter 6.

turn would compromise its status as an action. But neither need they be present to the agent in the way that they are when we deliberate about what to do. Nevertheless, according to the reasons theory, in rational actions, they are always there. One might speculate that their attribution will be somehow linked to the agent's willingness to self-ascribe these reasons if pressed on the matter after acting. But even if this were maintained, their ontological status is left wide open.

I am going to assume that all of the reasons theorists I am going to discuss would agree on the above characterisation of reasons, and the rationalizing relation. Given this characterisation, the reasons theory has a number of things going for it. The reasons theory gives an account of the *normativity* of rational actions, that is, the sense in which we *should* do them, since the actions are not just explained but justified by the reasons. The reasons theory gives an account of the sense in which rational actions are *authored* by the agent, since reasons must be within the agent's point of view. And whilst the issue of the causes of actions is hotly debated, the reasons theory at least has the potential to deliver an account of the *etiology* of rational actions. For the reasons could plausibly form a part of a story of what brought the actions about. I take it that these factors go some way to explaining the theory's popularity. The implication is that if habitual actions fail to be rational according to this theory, as I think they do, any alternative may find it hard to live up to these standards. I shall follow up this challenge in Chapter 6.

At the risk of oversimplifying, I shall now divide philosophers into two groups according to their views about reasons: *reasons internalists* and *reasons externalists* (abbreviated as *internalists* and *externalists*). Very roughly, reasons internalists believe that reasons are psychological states of the agent. Reasons externalists deny this and claim that reasons are states, or possible states, of the world, which are in the purview of the agent. I want now to canvass each view and ask in each case whether habitual actions can be said to be done for reasons so conceived.

3. *Reasons Internalism*

The view that reasons are psychological states of the agent can be traced back to Hume. Recall that in his *Treatise*, Hume divides human psychology into "reason" and

"passion", and claims that reason is "the slave of the passions" (Hume 1740, p. 415). Since reason in some sense "guides" our actions, every action requires a contribution from both sides: a "desire" for something, and a "belief" that the action in question will satisfy that desire. This, at least, is the received view.

More recently, since Davidson's (1963) seminal paper, internalism has been thought attractive because it opens the way for the idea that reasons might be related to the actions which they explain causally. For a thesis such as this to be plausible, reasons internalism will seem mandatory, because only a state which is realised in some physical way in the agent's body, could be said to cause a movement of that same body.[6] Such a thesis places a constraint on how we understand the notion of a "psychological state", that is, as a state that supervenes on the physical state of the agent's body. But whilst reasons internalism might support a causal analysis of action, it does not entail it.[7]

Alongside Davidson, present day reasons internalists include Alvin Goldman (1970), Al Mele (1992a), Michael Smith (1994) and Simon Blackburn (1998). Whilst their accounts differ in their details, all believe that the reasons which rationalize an action consist of the characteristic pair of psychological states: a desire plus a means-end belief. Whilst other kinds and configurations of psychological states are theoretically possible, this is by far the most popular version of the view, and it is accordingly how I shall understand internalism.[8]

The terms "desire" and "belief" are broadly construed by internalists. "Desire" covers a range of attitudes, including wants and emotions, and is for that

---

[6] At least, that is, given popular assumptions about causal explanation. In contrast, Stout (1996) argues that there is an account of causation which allows a causal explanation of action to be compatible with the rejection of the internal conception of reasons.

[7] Smith reserves this possibility by arguing that reason explanation could be understood "teleologically" (1994, pp. 101-4), though he admits (p. 114) that as a matter of fact, he also holds a causal thesis. (See also Smith 1998, p. 18).

[8] Other possible views could say that just beliefs or just desires rationalize. "Intentions" might also figure in the internalist's picture. I anticipate that all such variations will be subject to the forthcoming arguments.

reason sometimes glossed, following Davidson, as "pro attitude". The following passage shows the diversity included under this latter notion:

> 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. (1963, p. 4)

Now I think there is a genuine question about whether some of these items deserve to be termed "psychological", but that is something I shall be discussing shortly. "Beliefs" may also include genuine states of knowledge, as well as states which fall short of it.

Reasons internalism is not the implausible view that our folk rationalizations always mention a belief and a desire. As the example of the explanation of my going for a walk "for health", illustrates, we do not need to mention such states to be understood. However, no matter what we might say in such explanations, according to internalists, until we have mentioned such states, we have not displayed the real reasons for which the agent did what she did. To display them, our everyday formulations can, and should, be recast in terms of the internalist's preferred psychological states. In the above example, we might for instance say that I went for a walk because I *wanted* to be healthy and I *believed* that by going for this walk I would help to bring that about. Here we have the standard pairing of belief and desire which internalists think must be the reason for every rational action.

There seem to be two possible ways in which a habitual action could count as an action for a reason in the internalist's sense. First, the habit itself might count as a psychological state of the agent. Second, the habit is not itself considered to be a psychological state, but it generates, and thus explains the existence of psychological states, which in turn rationalize the action. I shall consider these possibilities in turn.

4. *Are Habits Psychological States?*

An answer to this question will turn on how we understand the idea of a psychological state. Unfortunately, deciding on a correct conception of such things is not something we can hope to do here, so we shall have to settle for some fairly rough and ready characterisations of some possible views. I want to consider three conceptions of psychological states upon which an internalist might draw in order to argue that habits are states of that kind. These conceptions are respectively that psychological states are attitudes with propositional content; that they are dispositions; and that they exhibit a first/third personal epistemic asymmetry. I shall argue that in all three cases, either the conception of a psychological state is impoverished for the internalist's purposes, or there are problems in subsuming habits under that class of state. Thus habits are not psychological states.

I turn first to the conception of psychological states as attitudes with propositional content. If the reasons internalist believes that only propositions can provide proper justifications, she will say that psychological states are required to have propositional contents of some sort. The obvious thing is to say that the psychological states in question are the sorts of states which are commonly called "propositional attitudes". Beliefs and desires can both be of this form since one can believe *that* such and such is the case, or desire *that* so and so be the case.

But even if we accept that reasons are psychological states of this kind, there is a difficulty in thinking of habits in this way, since habits do not seem to have a propositional form at all. One has a habit of ϕ-ing in C, which is not the habit *that* anything. Habits just don't seem to be the sorts of things that can be captured propositionally.

Somebody might object by suggesting that there are ways for the internalist to capture the idea of a habit propositionally, and indeed this can be done either in terms of a belief or a desire. The suggestion might run as follows: my habit of ϕ-ing in C can be expressed either as my belief that when in C I should ϕ; or as my desire that when in C I ϕ.

Now these attributions seem on the face of it to be little more than theoretical posits, and artificial ones at that. They seem to be ascribable only as a consequence of knowing an agent has a habit, and hardly the sorts of states which the agent will self-ascribe. But even if we accept the legitimacy of such posits, there is a more serious problem with the suggestion that these states capture the idea of a habit. For it would be compatible with an agent having a habit of ϕ-ing in C that they have the opposing beliefs and desires. So I might actually believe that when in C I should not ϕ; my problem is that my habit points me in the opposite direction. Similarly, I might desire that when in C I should refrain from ϕ-ing, it's just that my attempts at losing the habit have yet to succeed. So, assuming that the suggested beliefs and desires are not supposed to be completely independent of the agent's other beliefs and desires, somebody who wanted to lose a habit would be guilty of a kind of incoherence. It seems then that if we understand them as genuine beliefs and desires, they do not capture the idea of a habit. Habits can operate independently of the agent's propositional attitudes. And this is even true of those attitudes which concern the very behaviours involved in a habit. So unless the posited attitudes are conceived as entirely independent of the agent's actual beliefs and desires, which would thereby fail to capture anything of the agent's point of view, they must be rejected as an account of reasons. Habits then are not propositional attitudes.

Denying propositional content to habits may seem to be a radical step. Indeed those gripped by the reasons theory may resist this move simply on the grounds that if habits do not have propositional contents, they certainly could have no role in rationalizing an action. Others may resist this move because of the implications for the individuation of habits. After all, according to the current suggestion, psychological states can be individuated by means of their propositional contents. Denying that habits are of this kind will render this method unavailable. But I think we can admit that there may be such challenges, without conceding that giving an explanation in terms of a habit is just to deny an action's rationality.

The above considerations lead to a second way in which the internalist might try to say that habits are psychological states, namely, to say that psychological states are a kind of *disposition* to act in certain ways. If I have a desire to go rowing, you will

expect me to take the opportunity to do so when it arises, or else you will suspect that I didn't really have that desire after all. Similar things can be said of beliefs. If I believe that Caracas is the capital of Venezuela, you will rightly expect me to say certain things when I describe South American countries, and to give certain answers to questions on that topic. The idea that psychological states are dispositions and nothing more, is often, and to my mind unfairly, attributed to Ryle.[9] Nevertheless, adopting this view of psychology would provide a rationale for thinking of some of the more dubious items in Davidson's set of "pro attitudes" as psychological. Here I am thinking of what he refers to as "social conventions" and "permanent character traits" (1963, p. 4). And more to the point, this account of psychology would also be conducive to the internalist counting habits as reasons. For since habits can be regarded as a sort of disposition, it follows that they, on this account, are psychological states.

But whilst the dispositional view might give the internalist a quick victory, I think the internalist requires more of a psychological state than this view can offer. In particular, the internalist needs to capture the sense in which reasons are from the agent's point of view. And dispositions do not seem to capture that thought at all. I am disposed to blink when something passes close to my eyes, but my disposition does not make this sort of behaviour seem attractive to me. The disposition might be an explanation of what I did, but it is not a reason in the required sense.

That is not to say that dispositions are an unimportant aspect of psychological states. But having a disposition cannot be all there is to being in a psychological state of the relevant sort. Something more is needed which captures the sense in which a psychological state expresses the agent's perspective on things.

---

[9] This attribution is usually accompanied by the charge, meant in a derogatory way, that Ryle is a "behaviourist", see e.g. Fodor (1987, p. 292). Compare Ryle: "I am not, for example, denying that there occur mental processes. Doing long division is a mental process and so is making a joke" (1949, p. 23). So don't think Ryle would endorse the "and nothing more".

One way of understanding this idea is in terms of an *asymmetry* between the way I ascribe such states to myself, and the way I ascribe them to others.[10] This asymmetry may be understood in a number of ways. It may for instance, be due to the states in question having certain first-personal "feel" which is not available to anybody else. This would be particularly evident for some desires like hunger, or longings, though less so for others, such as the desire to watch television, and arguably absent for beliefs. Perhaps more plausibly, the asymmetry may consist in the fact that unlike others, only I can know my own psychological states non-inferentially, that is, without making any inferences from my verbal and non-verbal behaviour. Or the asymmetry may be understood in terms of a certain sort of authority I have over which states I am in, that nobody else has. Any of these construals of asymmetry can be said to capture the agent's point of view. A purely dispositional account of psychological states simply misses that aspect out.

It is on these grounds that I would dispute Davidson's inclusion of "permanent character traits" and "social conventions" amongst his pro attitudes. For even if an agent were to have one of these "attitudes", that does not entail that he "saw, or thought he saw" (1963, p. 3) anything attractive about manifesting one of these dispositions. The reason is that one can have these "attitudes" without having any sort of distinctive perspective on them. This is something which is also true of habits. Indeed, on a plausible interpretation of Davidson, the two "attitudes" in question are kinds of habit. One can have a habit (or character trait, or social convention) without any idea *that* one has it, or of what, if anything, can be said *in favour* of having it. In such cases, the "attitudes" in question fail to rationalize the action on Davidson's own terms. So I think we have to reject their inclusion in the legitimate set of pro attitudes.

The above problem with the dispositional account naturally suggests a second way in which the internalist could understand psychological states, which picks up on the asymmetry of such states. Dispositions can still constitute part of the account, but it will have to be supplemented with some criterion of asymmetry. The question is, is

---

[10] See for instance Lowe (2000, p. 68).

there any understanding of asymmetry, which both allows standard beliefs and desires to be psychological, and does not rule out habits? I do not think so.

In support of this view I shall canvass the three interpretations of asymmetry sketched above. First, using phenomenological feel is unpromising. Since such feels are missing from all of the beliefs and many of the desires typically involved in reason attribution, it will not do as a criterion of the psychological for the internalist. It may be a sufficient condition for a state to be psychological, but it is not necessary.

Second, can asymmetry be understood in terms of whether an inference is required? The idea here is that I only know other people's psychological states by making inferences from their verbal and non-verbal behaviour, whereas I can self-ascribe such states to myself immediately, that is, without any such inference. Now there may be some habits, including those which we have had from infancy, which we can self-ascribe in this way. I don't, for instance, need to infer that I eat three meals a day, and that I speak English. But an inference is likely to be required in the self-ascription of other habits. In particular, this will be the case if I have not previously reflected upon whether I have a particular habit, and nobody has brought it to my attention.

There is also a question of how satisfactory an asymmetry in kinds of inference is as a criterion of the psychological. I seem, for instance, to know certain intuitively non-psychological facts about myself non-inferentially, whereas others may not, such as what my name is and how tall I am. Like my self-ascriptions of well-known habits, I can check these things empirically, but I still seem to know them non-inferentially. So it looks doubtful that this way of spelling out asymmetry will provide a necessary condition of the psychological.

The third possibility is whether an agent's having a certain sort of authority over her states is a plausible criterion of their being psychological. The idea, which is known as *first-person authority*, is that I cannot be mistaken about my own psychological states in the way in which I can be mistaken about the states of others.[11]

---

[11] See for instance Glock & Preston (1995), who attribute this view to Wittgenstein.

Now I don't want to take issue with this way of spelling out asymmetry as a plausible criterion of the psychological, and indeed one that could be adopted by the internalist. The problem is, I don't think we enjoy this sort of authority over our habits.

Here we have to be a bit careful. For because of the way things are defined, it might be suggested that there is a sense in which we have a kind of first-person authority over our habits. Specifically, I have authority over whether any particular candidate habitual action is automatic. Nobody else is authoritative over whether I tried or deliberated over a particular action, or range of actions - that may be just a matter for me. Even if I visibly paused beforehand, I might have been deliberating about something other than the action itself. However, whilst I would admit that I have authority over whether or not I tried or deliberated, I do not think that this amounts to authority over whether I have a particular habit or not. This latter fact will depend upon there being similarities across a range of my own behaviours, and whether there are such similarities is something over which I am definitely not authoritative.

Indeed, I would venture that if anybody has authority over whether my behaviour is repeated, it will lie with people who know me well, rather than with me. This can in some cases be because I do not have a very good view of some of my behaviours, in particular facial expressions, and bodily postures. But most obviously, I do not have authority because certain other people, in particular those who know me very well, and with whom I cohabit, are confronted with my ways continually. There is a sense in which such people need to know my habits, so that they can get used to the tolerable ones, and bring the unbearable ones to my attention! So whilst I cannot be intelligibly informed about whether I did something automatically, I can be informed about whether doing that thing is a habit of mine; and that is why learning of such things can come as a surprise to me. This is not to say that I may not, as a matter of fact, know at least some of my habits better than anybody else. But that is simply a contingent matter of my having observed the behaviour more often than anybody else, and not a logical feature of habit attribution.

I have not come across an account of this sort of authority in the literature, but I call it *second-person authority*, and I shall say that habits enjoy second-person

authority. The upshot is that if we characterise psychological states in terms of an asymmetry of authority in favour of the first-person, habits are not psychological states. If we seek an authority about such states, it would be better to ask somebody else.[12]

So we have not found a way of spelling out asymmetry in a way that will be both conducive to the reasons internalist, and allow her to count habits as reasons. And neither of the other two interpretations of psychological states allowed this either. There may be other ways of understanding the internalist's idea of a psychological state which I have missed, and which are not subject to the sorts of difficulties described above. I do not know what they are. In the absence of such an understanding, we must conclude that habits are not psychological states, and so are not reasons in the internalist's sense.

5. *Do Habits Explain Psychological States?*

The other possibility for the internalist is not to say that habits themselves are amongst the reasons for a given habitual action, but rather that habits explain the reasons, themselves understood as standard psychological states, which in turn rationalize that action. Goldman seems to have something of this kind in mind in the following passage:

> I frequently call a certain number in Yonkers, 914 YO 5-3438. Each time I do this, I begin with an occurrent want to dial the whole number. As I proceed, a smaller want is formed to dial each successive digit. On one occasion, however, I want to call a different number in Yonkers, 914 YO 8-7183. Having dialled 9-1-4-Y-O, habit results in my forming an occurrent want to dial 5 instead of 8. (Goldman 1970, p. 90)

The suggestion seems to be that habit operates at one remove from the real reasons associated with habitual actions, which include "occurrent wants". These wants are, in turn, be conceived as components of reasons according to Goldman's internalist picture of these things, so the dialling of each digit is done for reason, even though the

---

[12] Presumably a similar thing could be said about Davidson's "character traits" and "social conventions". And this further strengthens the case against their inclusion as internalist reasons.

dialling of the entire wrong number is not, since the agent in question at no point wanted that. So here it looks as though the habit is operating to generate reasons, although the habit itself does not confer rationality on the resulting actions.

The question is, could habitual actions in general be incorporated into the internalist's scheme in this way? I don't think so.

For a start I think there is room to deny that any of the posited occurrent wants exist at all. Of course, the context of Goldman positing these things is defending his own internalist, and causal, picture of intentional actions, and he would like that picture to apply to as many actions as possible. But I think that he is being over-ambitious in this case, since the occurrent wants enjoy no pre-theoretical support whatever. Indeed, if the agent in Goldman's example did not want to dial the entire number, then surely it follows that he did not want to dial each digit of that number either, notwithstanding his habit. I suggest Goldman's positing of such items says more about his commitment to a particular theory of action, than it does about the way habits work, which, as we saw in the previous section, can be contrary to our beliefs and desires.

What I think gets to the heart of the matter, however, is this. On the scheme suggested, it is the belief and desire which are rationalizing the actions on their own merits, and regardless of how the agent acquired them. So the actions in question are not actually being assessed as rational under their description as habitual. In describing the justifying states as "occurrent wants", the idea of their having been generated by a habit drops out of the picture. The actions in question are instead being assessed under a description which is independent of the habit. So the scheme cannot show how specifically habitual actions are rational, which is what we wanted. It seems we must conclude that the reasons internalist cannot show how habitual actions are actions for reasons in this way either. And I do not know how else they could go about it.

If we combine the fact that habitual actions cannot be accommodated by reasons internalism, with the popularity of that view, we get part of an explanation of why habitual actions are thought of as, at best, only marginal cases of rational actions. And this may in turn go some way to explaining why so little philosophical attention

has been paid to this class of actions. But there are difficulties with reasons internalism, and if we can show it must be rejected, this may give habitual actions a second chance at being actions for reasons.

6. *Problems with Internalism*

Reasons internalism has been criticised in recent years. Critics include McDowell (1978, 1995b), Quinn (1993), Stout (1996), Collins (1997), Stoutland (1998) and Dancy (2000). Here I want to sketch what I take to be two of the most important criticisms of internalism, without going into details or considering objections, which I leave to others. My purpose is just to illustrate the kinds of problems that internalism faces, regardless of its success or failure at accommodating habitual actions. Not only will that give us reason to consider habitual actions in the light of the externalist alternative, the job of §§7-9, but it will also support my view that internalism cannot provide an adequate account of the counterfactual scenarios necessary for the control of habitual actions.

I shall look at two ways in which reasons internalism is criticised, both of which, I take it, would be endorsed by the externalists named above. Internalism is criticised first, on the grounds that desires cannot rationalize, and second, on the grounds that beliefs cannot rationalize. I describe the criticisms in that order.

The need for desires (or pro attitudes - I shall speak indifferently of desires for now) in rationalizations has been criticised perhaps most persuasively by Warren Quinn (1993, pp. 228-55). The idea under attack is that desires are "non-cognitive" states of the agent. The argument can be put quite simply. Agents can have desires for all sorts of things, some of them quite bizarre. Examples in the literature include Anscombe's man who has a desire to drink a saucer of mud (1957), and Davidson's "yen ... to drink a can of paint" (1963, p. 4). Quinn gives the example of somebody who just loves to switch radios on, though he doesn't do it to listen to the music, for information, or for any other end (1993, p. 236). Quinn's central point is that an agent's simply having a desire for something does not tell us what is good about it. And if it does not say that, it cannot do the rationalizing job. Thus internalism does not do proper justice to what I called in §2 the normativity of rational actions. True, if

we know that somebody likes performing some kind of action, we would normally expect there to be something good about it. But just being informed that they like it does not yet tell us that there is anything good about it, or what that something is.

Of course, if desires were thought to be cognitive items, that is, states that somehow latched onto things that were in themselves good, that is, desirable, as opposed to merely desired, such difficulties might be lessened. But since internalists tend to follow Hume in the strict division between reason and passion, and since denying this would seem to entail that the world is populated by values as well as facts, the metaphysical price may seem too high. Non-cognitive states face Quinn's difficulty, and there does not seem to be an easy way out.

This difficulty does not just affect desires, but it affects habits such as Davidson's character traits and social conventions, too, and bolsters the earlier argument against accepting such things as being reasons. Explaining what I do by saying that it is my habit does nothing to say what is good about it. It might be a terrible habit (nail-biting), just as a character trait might be sinister (avarice), and a social convention could be extremely harmful (slave-ownership). To be fair, Davidson himself might be able to evade this problem because of the way he characterises rationalization - which must reveal something that "the agent wanted, desired, prized, [or] held dear" (1963, p. 3), all of which are arguably non-cognitive attitudes - but if Quinn's criticism is right, this does not make proper sense of the action from anybody's point of view, never mind the agent's. This is why, I would suggest, Davidson should not allow that a yen to drink a can of paint could ever rationalize it.

I shall say more about the significance of the distinction between good and bad habits in the next chapter. For now we can conclude that the fact that an agent has a particular desire, conceived as non-cognitive state of the agent, cannot rationalize what she does. Of course, if desires are not conceived as non-cognitive, things may look more hopeful for the internalist. I shall say more about such states below.

What about the rationalizing role of beliefs? Arthur Collins (1997) gives an example of a rationalization which is attributed to Joe, who usually drives across the bridge to work. Today, however, the bridge is closed, and Joe decides to take the ferry. Thus, the explanation is as follows:

(O) Joe is taking the ferry because the bridge is closed. (Collins 1997, p. 109)

This explanation is stated in terms of the objective circumstances, namely, that the bridge is closed. According to internalism (what Collins terms "the standard outlook"), this requires a "psychological restatement" as:

(P) Joe is taking the ferry because *he believes the bridge is closed.*[13]

Collins, however, argues that "Appeals to objective circumstances need not be deleted in favour of psychological realities. As a matter of fact ... claims about objective circumstances cannot be deleted without dropping the explanation altogether" (p. 109). Collins suggests that the mistake of the standard outlook is to think of beliefs as "stored representations" upon which we "report" when we self-ascribe beliefs. He states the problem with this view as follows:

> the explanation proffered by 'I am doing it because I believe that *p*' absolutely depends on the fact that these words do express the speaker's commitment to the truth of *p*. The explanation makes the claim about the objective circumstance and cannot delete that claim. (Collins 1997, p. 118)

So we cannot coherently say of an agent that she believes that *p* and at the same time hold that she is indifferent about whether *p* is the case. Expressing a belief, then cannot simply be the agent's making a reference to her own internal resources. This does not of course mean that psychological restatements like (P) are in error, but it does mean that we cannot think of the fact that it is expressed in terms of belief as crucial to (P)'s rationalizing Joe's action. To have the requisite normative import, then, rationalization requires the right kind of content, and that content cannot be understood independently from the agent's commitment to the way the world, containing bridges and such like, is arranged. It is not the fact that an agent *believes* that *p* which rationalizes what she does, but simply the fact that *p*. The internalist's

[13] I change Collins' exact phrasing (1997, p. 109) to the third person, I take it without loss.

insistence on the restatement in terms of belief, then, has things the wrong way round.[14]

My presentation of these moves is doubtless too schematic to persuade the internalist.[15] Indeed in Chapter 7 I return to some internalist defences made by Smith and Blackburn, that have been put forward in the context of a debate in moral philosophy, though I shall argue there that ultimately these defences fail. But the above difficulties should be enough to show why, from here on, I shall take it that it is reasons internalism which requires defence.

In the light of the two anti-internalist arguments described above, the position I call reasons externalism would seem to follow quite naturally. Here I go beyond what can strictly be attributed to Quinn or Collins. If, in response to Quinn's criticism, we drop the claim that desires are non-cognitive, it seems that a "cognitive" conception of these states might be able to do the rationalizing job. On such a conception desires, like beliefs, depict the agent's perspective on some feature, or possible feature, of the world. Here we need the idea of a *possible* feature of the world because desires are typically for states of affairs which do not yet obtain, and require action to bring them about. But this need not lead us to reject the label "cognitive".

It may be thought that such a state should not be termed "cognitive" because states concerning merely possible states of affairs are not items of knowledge. However, calling desires "cognitive" captures two things about them. First, many desires (those that are not mere fantasy) depend a good deal on what is actually the case, and what the agent could realistically attain, so in this sense it is very misleading to brand them non-cognitive. Second, if desires are conceived non-cognitively, there is only one way in which such a state could "make sense" of an action, namely, that the action *satisfies* it (where "making sense" falls short of rationalization). Calling

---

[14] See also Dancy (2000, Chs. 5 and 6), who draws on Collins, and advances a similar line. See also Stout (1996, Ch. 1).

[15] For instance, Michael Smith (1998) has recently claimed that "while an explanation in terms of a fact presupposes the availability of a Humean belief/desire explanation, the reverse is not true" (p. 20). Smith does not however seem to be aware of the arguments presented here.

desires "cognitive" brings out the sense in which they are *for* something more than their own satisfaction: they are for something *in the world*, and not "in" the agent.

This blurring of the distinction between belief and desire on grounds of their cognitive features alone, then, places in doubt the standard internalist picture which requires both kinds of state. Properly understood, one or other of them could be sufficient. But if we then also accept Collins' criticism of internalism about beliefs, and generalise it for the cognitive conception of desire, we see that the idea that psychological states are the motivating reasons can be dropped altogether. If Collins is right, it is the agent's acting in the light of the way the world is, or (granted the above extension) might be, that justifies, and so rationalizes, her action, and not her *thoughts* to that effect.[16] This is the position I call reasons externalism.

For now, then, I shall assume that externalism provides the best going account of actions for reasons, and hence the best account of the counterfactual scenarios which are necessary for rational intervention control. Having rational intervention control over a habitual action, then, cannot be adequately conceived as merely attending and responding to one's psychological states, as the internalist could have it. Rather, it is better thought of as the agent's attending and responding to the way the world is, or might be, irrespective of her own beliefs and desires.[17]

7. *External Reasons*

Before turning to habitual actions specifically, I need first to further clarify how reasons are conceived by reasons externalists, *external reasons*, as I shall call them. Like all reasons theorists, externalists need to distinguish between those reasons for

---

[16] The idea of "the way the world is" is expressed in various ways by different writers. These include "external situations" (Stoutland 1998, p. 43) and "external fact" (Stout 1996, p. 35). I take these to be compatible with Collins' suggestion. I also think that the rather technical construal of an "external reason" whose coherence is denied by Williams (1981) and defended by McDowell (1995b) is compatible. Williams' uses of "internal" and "external" reasons should not be confused with mine.

[17] I shall suggest in Chapter 8 that this externalist conception of the counterfactual scenarios may itself be nuanced in the light of a conception of rational action corrected to accommodate habitual actions.

which the agent acts, and those for which she could have acted, but in fact didn't. They still need to make sense of the idea that the reasons which rationalize are "from the agent's point of view". Stout explains the requirement thus:

> If the agent did not have some immediate mental access to the reasons for their behaviour, then the involvement of their agency in this behaviour would be brought into question. The immediate reasons for an agent's action must be *their* reasons. (Stout 1996, p. 38)

This requirement for the agent to be aware of her reasons is appreciated by other externalists. We saw in §2 that McDowell speaks of a "favourable light being shed" by the reasons on the action (1978, p. 79). Presumably the light shed is one from which the agent can benefit. He writes of the reasons being "something of which, on each of the relevant occasions, [the agent] is aware" (1979, p. 51). We have already seen Dancy writing of reasons "persuading" the agent to act (2000, p. 1). Presumably such persuasion cannot be done without the agent having some awareness of the items in question. And here is Dancy again in a revealing passage:

> [The explanation in question] specifies the features *in the light of which* the agent acted. It is required for this sort of explanation that those features be present to the agent's consciousness - indeed, that they be somehow conceived as favouring the action; so there must always be a way of making room for this fact, in some relation to the explanation that runs from features as reasons to action as response. (2000, p. 129)

The way externalists distinguish the agent's reasons from other worldly facts, then, is by saying that the agent must be "aware" of those reasons. This claim can be separated into two aspects, one conceptual, the other temporal. The conceptual aspect is what Dancy expresses above as that the features "be somehow conceived as favouring the action". That is to say, the agent must know what the relevant facts are, not just under any description, but under a description which justifies that action. Note also that such justifications need not be justifications for everybody, but must be justifications for that agent. So the fact that there is a stamp fair in town is not a justification for me to go there, since I have no interest in stamps, though it will be for a philatelist.

The temporal aspect is that the agent must be aware of the justifications in this way *at the time that she acts*. Now since actions have a duration, this is not very specific. But if externalists are going to make sense of aborted actions, that for reasons outside the agent's control fail to be completed, the awareness will be required at least at the moment that the action begins, if not for its entire duration. Certainly, the awareness is not something that may or may not occur after the action is complete, as might be the case were the agent questioned about her reasons afterwards, and only at that point constructed a justification (what I called in Chapter 3, §6 a "*post hoc* rationalization"). Rather, for externalists, the agent's awareness of the reasons is supposed to be an essential part of the action itself as it takes place, and hence ineliminable from its proper rationalizing explanation.

The need for both of these aspects of externalist rationalization is particularly vivid in Stout's (1996) teleological theory of action. He proposes that reasons externalism is compatible with a causal interpretation of rationalization, in which the reasons specify the goal and direction of a causal process. He writes:

> My claim is that actions are the immediate results of causal processes which are sensitive to actual (external) means-end considerations [i.e. reasons]. (Stout 1996, p. 3)

It seems clear that such causal processes could not operate if the reasons which govern them, in a form apt for justification, were not present to the agent for their duration.

Given this clarification, we can now turn to the question of whether habitual actions are actions for external reasons. There are two ways in which this might be claimed. The first is to claim that habits themselves are external reasons. The second is to claim that the circumstances to which a habitual action is a response constitute external reasons. I consider these in turn.

8. Are Habits External Reasons?

The idea here is that an agent's habit is part of the world in the sense required by the externalist, and hence the habit is her external reason for its exercise. The basic idea may seem obvious enough: the habitually acting agent acts in the light of her past behaviour.

I think this is what Philip Pettit (1993) has in mind for his notion of an "extrapolative inclination", which is central to his account of rule following (pp. 76-106). Pettit thinks that for a rule to be something the agent can try to follow (a requirement I would question), it must be "independently identifiable". He suggests that "examples of [the rule's] application" (p. 86) are identifiable in this way. "Looking at the instances, [the rule following agent] will be able to have the rule as a salient object of belief; [the agent] will be able to target or address the rule" (p. 87).[18] From Pettit's analogy with a chimp who learns the "triangularity" rule, we can gather that for him, these instances are historical: "The chimp forms a disposition in response to the actual examples presented" (p. 89). But I don't think habits can be external reasons in this historical sense.

There are at least the following two problems. The first is that I do not think that there are any general grounds for attributing a conception of the habit to the habitually acting agent at the time that she acts. The second is that facts about the agent's past behaviour do not by themselves justify continuing to behave in that way. I consider these problems in turn, before suggesting what might be a deeper difficulty with the proposal.

First, I want to consider two sorts of grounds for attributing a conception of the habit to the agent. We know from our definition of habitual actions that we cannot seek the grounds for attributing to the agent a conception of a habit by appealing to the agent's phenomenology. Should the agent be consciously aware at the time of acting of any facts about her past behaviour, that would just indicate that the action was not in fact habitual. So the habit cannot be present to the agent in the sense the externalist requires.

This is not to rule out the possibility that one can conceive of oneself as having a particular habit, and act in the light of that fact. I may, for instance, find myself being unsure about what to do in some familiar situation, but then decide that on balance it would be better to do what I've always done in this kind of position, or do

---

[18] Pettit's use of "belief" here belies his commitment to internalism, but his claim about what kind of content is relevant can be read externally.

what comes most naturally. However, if the foregoing accurately describes my action, it is not a genuinely habitual action by our definition, for such pausing and thinking disqualifies the action from being automatic. If we let the habit do its usual work, this will be an intellectualised misdescription of the phenomenology.

A second source of grounds would be in what the agent says about her reasons after acting habitually. Were she asked why she acted as she did, the replies "that's what I've done before" or "that's what I always do", or even "that's my habit", might be thought grounds for saying that the habit is her external reason. But even if this were the only justification that the agent were prepared to give, we still lack sufficient grounds to conclude that the fact cited was present to the agent *when* she acted.

Here we should recall the distinction from Chapter 3, §6, between *post hoc* rationalization and genuine recollection of hitherto "subconscious reasons". What sorts of grounds may be given for thinking we have a case of the latter? If the agent is questioned after the action, the two may not be easy for the agent, or anybody else, to distinguish. Grounds for the recollection might be the agent's protesting that she "remembers" what she thought at the time; or that she is extremely reluctant to endorse all but one rationalization of what she did. But here we run into difficulties about how much store we should put on what agents are inclined to say in such situations. We shall need to take into account such things as their sincerity, their motivations (are we offering them incentives, or punishments, for saying one thing rather than another?), and how theory-laden their claims may be. We may in fact need to bear in mind the possibility that our agent has done some philosophy of action. So I can't see that any of these grounds are going to settle beyond doubt the question of whether we have a genuine case of recollection of hitherto sub-conscious deliberation.

What is more, there is reason to think that in the case of many habits, the chances of an agent offering either of the above sorts of self-justification will be zero. For if we reflect on the fact that many habits are naturally acquired either unintentionally or non-intentionally, and that, as I described in §4 above, habits enjoy second-person authority, then supposing that the agent to always have the requisite self-knowledge for such justifications will appear hopelessly optimistic. Without a

conception of her own habit, no particular stretch of her past behaviour will be salient to her. Indeed we could rather expect that there will be occasions when a demand for self-justification is met by no response at all from the agent. That is not to say that the habitual actions in question are not perfectly sensible things to do. But it is to say that if the availability of an after-the-action self-justification is the sole ground for ascribing a conception of the habit to the agent at the time of her action, it will be in general unavailable. And even when it is not, it looks to be inadequate.

I confess that I don't know what other pre-theoretical grounds we might use for ascribing the requisite conception of the habit to the habitually acting agent.

The second problem for the proposal that the habit's history is her external reason, is that even supposing that content could be correctly ascribed to the agent at the time of acting, we should not assume that it will be a justification. This is a problem is one we came across above (§6) in Quinn's argument against the idea that desires can rationalize. For habits the problem is that it does not follow from the fact that I have done something before that I should keep doing it. That I have bitten my nails for years is no justification for me to continue doing so. Sometimes indeed the fact that I have done something repeatedly gives a justification not to do it. The fact that I am unduly critical when I mark student essays is a justification for not doing it again. The fact that I keep doing it, as opposed to having done it just once or never before, makes my getting it right next time all the more urgent. So a behaviour's past history does not justify its continued performance.

Somebody may suggest a remedy whereby we distinguish between good habits and bad habits. If we have such a distinction we might say that the good habits are the ones whose history does provide a reason to continue with them, whereas the bad ones, like nail-biting, and being over-critical in one's marking, do not. I think this may be an important distinction, and one which I shall discuss further in Chapter 6. But it is no solution to the current problem, because the evaluation of a particular tract of one's history as good or bad is carried out independent of its habituality. It amounts to the concession that habituality by itself provides no justification. It is an explanation only, not a justification.

I think the above two problems are sufficient for us to reject the idea that an agent's past history could constitute her external reasons.

Before moving on I want to raise a further, and perhaps deeper, difficulty with the idea of habits being external reasons, which requires a slightly different understanding of that idea than the one with which we have been working so far. Up until now we have been assuming that for a habit to be an external reason just means that the agent's past history of exercising that habit is the reason. But that doesn't get things quite right. For it seems that unlike other external reasons, habits cannot really be separated from the actions they are supposed to explain, for the present action itself is an exercise of the habit too.

Compare habits with other external reasons. Paradigm external reasons, such as bridges being closed, other than being reasons for an agent, are otherwise independent from the action. But habits are not independent from the actions they are purported to rationalize. This is because every habitual action is *part of* the habit which is supposed to rationalize it. So the action itself is part of the putative reason. The relation between habitual action and habit is one of part to whole. And whatever the supposed relation between reasons and actions has been thought to be by philosophers engaged in these debates, it has never, as far as I know, been conceived of in this way. So a habit is not something an agent acts "in the light of", but something she acts *with*.[19] It is the wrong sort of thing to be a reason. This leads me to think that the externalist suggestion at hand may be more confused than either of the first two criticisms acknowledge.

9. Are the Circumstances External Reasons?

The other possibility for the external reasons theorist is to suggest that the reasons for a habitual action are the circumstances within which that action takes place. This may look more promising since the circumstances are not part of the agent, but part of the world, and since the circumstances are what the agent is in some sense, responding to,

[19] I am open to suggestions about better ways of expressing this idea.

the agent cannot be wholly ignorant of them. The question is, can these circumstances be external reasons?

There are two difficulties. First, there is the difficulty of whether the habitually acting agent can be said to conceive of the circumstances when she acts, in any particular way at all. For instance, in habitually brushing her teeth, an agent may said to be responding to the fact that it is early in the morning, that she has just walked into the bathroom, or that she craves the taste of peppermint in the mornings. Even if we stay within the range of what the agent might plausibly assent to, there seems no upper limit on the number of possible descriptions of circumstances to which she could be said to be responding.

However, as we have seen, the external reasons theorist requires that the agent is aware of the circumstances under some determinate description at the time of acting. Without such a description, the circumstances cannot provide justifications, for "these circumstances" alone does not justify anything. But just as we saw in the previous section when considering how a conception of the agent's history might provide reasons, without phenomenological grounds, such a description is hard to find. Analogous difficulties will also arise if we seek grounds for attributing a description in what the agent later says, since the interpretation of a self-ascription will be indeterminate between recollection and *post hoc* rationalization. So it is unclear how the externalist could go about settling on a description of the circumstances that is unequivocally available to the agent at the time she acts habitually.

This difficulty should not be overstated. For we don't want to deny that there may be some, perhaps "low-level", awareness, on pain of admitting that the action was not really a response to the circumstances at all, and perhaps even just an accident. But how are we to understand this sort of "awareness"?

One suggestion is by comparison with the sort of awareness we can be said to have when we are waking up after being asleep, that is, a sort of absent-minded daze. In such a state, we are not characteristically aware of anything in particular, for we are not typically focusing our attention on anything specific. However, it would also seem wrong to describe such a state as one of being *un*aware of the world around us. Whilst it is undoubtedly possible for us to bring the content of this awareness under

various descriptions given a sort of "coming to" - it is not in principle out of reach of our concepts - it would be a mistake to think of this content as coming to us with any determinate conceptual shape. I think our awareness of the world in habitual actions could plausibly be of this sort, not only because we can exercise some habits when actually half-asleep, but also, and more often, because we exercise habits when thinking about something else entirely. In these latter cases, we are not "awake" to our automatic exercises of habits.

If this is indeed the sort of worldly awareness that we have in habitual action, what counts decisively against the external reasons theorist is that the content of this awareness is not of the required shape to provide justifications at all. There is nothing good or bad about the world as such which could make our acting seem worth doing. And this is the second problem for the external reasons theorist.

So now the dilemma for the external reasons theorist is this. Either they attribute to the agent awareness of the circumstances with content determinate enough to justify. But we lack pre-theoretic grounds for fixing on such a content. Or they admit that there is no specific content to the awareness of the circumstances. But they now have difficulty saying how such content justifies. Either way, the prospects for saying that the circumstances are external reasons do not look good.

This is not to say that there are no good reasons for acting habitually, and nor is it to deny that these reasons may be quite correctly said to be aspects of the circumstances in which such actions are performed. We may be able to locate in those aspects of the world our goals, and ways to reach those goals, which could justify what we do. On the other hand, we may not be able to locate such justifications. It would seem very optimistic to suppose we always could, for we seem able to acquire the most pointless habits. Examples would be hair-twiddling and mannerisms, of which even to say they are done "for the sake of it" seems strained. But when we do locate such justificatory aspects in the circumstances in which an agent habitually acts, whilst they are *the agent's* justifications in the sense that they justify what that agent does, they do not seem to be reasons *for which* she acts, in the externalist's sense of that term. Indeed, it seems likely that there are no such reasons.

What we have uncovered here then is a residual intellectualism, perhaps a hangover from internalism, in the externalist's picture of things. For the externalist too requires the agent to have an intellectual involvement in reasons when she acts. We look for something that "persuades" the agent to act (Dancy), or something which gives a "means-end justification" of what she is doing (Stout), or something "in the light of which" she acts (McDowell). But in habitual actions nothing of the sort seems to take place. We don't need persuading to act habitually; we do it naturally. We don't need to have a justification at hand when we do what we habitually do, though we may be able to put one together afterwards. And if there are reasons which speak in favour of a habitual action, we act in the *dark* of them. I think that this intellectualised understanding of an external reason explains why the externalist is unable to say that habitual actions are actions for reasons.

If this is right, it should be possible to try to solve the problem by offering a more radically externalist characterisation of actions for reasons, one which requires no intellectual involvement whatever from the agent at the time of acting. This is an avenue that I shall explore in Chapter 6. For now, if only for the sake of keeping our terminology consistent with what real externalists say, I think we should accept the conclusion of this section, which is that a reasons externalist should deny that habitual actions are actions for reasons. The best explanation for a habitual action on their account may then be that the agent acts out of habit. That such an explanation falls short of a rationalization for the externalist as well as for the internalist, need not be thought to create undue difficulties. Reasons theorists of both kinds should concede that I have identified a class of actions to which their theories do not apply; there are still plenty of non-habitual actions left over to which they do.

10. Habitual Actions are not Actions for Reasons

If we construe rational actions in the standard way, as actions for reasons, it seems that habitual actions are not rational. Whilst externalism is more promising than internalism, neither accommodates habitual actions. On the assumption that these two accounts exhaust the available accounts of actions for reasons, it follows from the reasons theory that habitual actions are not actions for reasons.

I do not think that this conclusion is controversial. Of the small number of writers who have considered the question of whether habitual behaviours are actions for reasons, most have argued that they are not. We have already seen Ryle's view of habits, thinking them as "mere" and "dispensing" with intelligence (1949, p. 42), the manifestations of which being contrasted with "motivated" behaviour (pp. 106-9). Along similar lines, N. S. Sutherland (1959) writes that "if someone did something purely from force of habit, it excludes the possibility of a motive explanation so that actions performed purely from force of habit must be excluded from the class of actions for which there is a motive explanation" (p. 146). And R. S. Peters (1958) points out that "to ask for a person's motive rules out the suggestion that he might be acting out of habit" (p. 152). The implication is the same as that supported by the argument of this chapter, namely, that habitual actions should, on account of their lack of rational motivation, be lumped together with other non-rational behaviours upon which the idea of a reason fails to get a grip. Of course it would be wrong to assume that the above writers were working with the same idea of habit as the one I have developed (indeed we saw in Chapter 2 that Ryle is not). But we may conclude with them that habitual actions are not rational actions at all.

This would be a mistake. For whilst we can accept the premise that habitual actions are not actions for reasons, we need not share the conclusion. For this chapter has revealed not that habitual actions are not rational actions, but that they are not members of *one class* of rational actions that we call actions for reasons. We have been assuming the truth of the reasons theory all along. And we may be able to maintain that habitual actions are rational actions if we reject it.

The reasons theory rests on an intellectualist bias in the philosophy of action itself. That bias is to think of all rational actions on the model of deliberated actions. We may flatter ourselves by thinking that most of our actions, if they are not actually deliberated, at least appear to be deliberated, and this is what makes them rational. As if we always get a (subconscious?) glimpse of the favourable light in which the world presents itself to us, and act in that light. Hence the fixation with actions for reasons. But now it should seem unsurprising that habitual actions which, since they are by definition automatic, are unlikely to fit into this deliberative mould. When we

act habitually we do not do two things - see the light, then act - but just one: we act. No matter how well thought out some habitual actions *appear* to be, we should not confuse appearance with reality. The insistence on classing all rational actions as actions for reasons is confused. It is just a modern form of intellectualism.

In Chapter 6 I shall show how we can hold that habitual actions are rational whilst avoiding the error of intellectualism.

6

RATIONAL HABITUAL ACTIONS

1. Problems with the Reasons Theory

In the previous chapter we saw that the reasons theory of rational action creates a number of difficulties with thinking of habitual actions as rational. In the face of these difficulties it may seem best to conclude that habitual actions and rational actions are mutually exclusive categories. On this view, the idea of a rational habitual action is an oxymoron. In this chapter I want to argue that this is a mistake.

Considerable support for the idea that habitual actions are rational comes from the anti-intellectualist projects in the history of philosophy, described in Chapter 2. If these interpretations of Aristotle, Ryle and Wittgenstein are right, their credibility would be weakened if it turned out that the common solution to intellectualism was a kind of behaviour which could be exhibited by non-rational animals. The exercise of virtue would then be not the expression of a rational human nature, but something closer to a blind response to stimulus. Ryle's dispositions would be similarly compromised in their proposed role as constituents of the mental - presumably the very seat of rationality - if manifestations of them turned out to be non-rational. And if following a rule was a non-rational activity, we would steer into what McDowell dubs Charybdis: "the picture of a basic level at which there are no norms" (1984, p. 242). Many philosophers think these projects are successful, yet they all strongly suggest that habitual actions must be in some sense rational.

There is a second and more general problem with denying that habitual actions are rational. In the previous chapter (§2) we noted that if the reasons theory is to be credible as an account of rational actions, it had better not turn out that most of the time we do not act rationally. But given the available interpretations of a reason which we have now canvassed, we might now seem pressed to draw that unpalatable conclusion. For it seems reasonable to say that most of the time, that is, at least when

we are awake, and reasonably healthy, we rational agents act habitually. Most of us do not in fact deliberate about what we will do very much of the time. Thus most of the time we conform, without trying, to stable patterns of action characteristic of habits. But if this is right, and if habitual actions are not rational, it follows that most of the time, rational agents do not act rationally. This paradoxical conclusion violates our intuitive understanding of what a rational agent is. Though a minority of the time, rational agents may not act rationally, an individual could not do this most of the time and still claim to be a rational agent. That most of the time we act rationally, then, provides a constraint on acceptable theories of rational action, which I shall call the *conceptual constraint*. If habitual actions are as prevalent as they seem to be, and our theory of rational actions forces us to deny that they are rational, we should to that extent be suspicious about that theory. And that implicates the reasons theory.

There are two options: modify the reasons theory, or reject it. Whilst modification may seem less radical, and therefore initially preferable, there are a number of disadvantages with it.

First, nothing I have said in criticism of the reasons theory prevents us from maintaining that actions for reasons, as they are currently defined, represent an important category of rational actions. Indeed we have seen how useful the idea has been for unseating reasons internalism. Keeping the idea of actions for reasons in play allows us to show that externalism at least purports to be doing the same sort of job as internalism. And there may be other debates that I haven't touched upon, which rely on the idea of actions for reasons. All this may suggest that to mark off actions for which the agent has an awareness of their justification at the time of action, is to identify a natural boundary in the class of human behaviours. Modifying the reasons theory would compromise this possibility.

There is a second disadvantage of modification which concerns the motivation for such a move. Insofar as modification is part of a quest for a unified theory of rational actions, which specifies the essential nature of such actions, it may be poorly motivated. I shall say more about this difficulty later in the chapter. But for these reasons I think we should resist modifying.

Of course, the reasons theory is so entrenched in philosophical thinking that abandoning it seems a radical step, perhaps too radical if our aim is to accommodate the apparently rather marginal category of habitual actions. But I think that the step is worth taking, not just because it allows us to avoid the above difficulties, but also because abandoning the reasons theory casts doubt on the very assumption that habitual actions are only marginal members of the class of rational actions. The idea that habits are marginal may depend rather too much on the assumption that the reasons theory is true. If we reject it, we might discover that habitual actions are not marginal at all, but are candidates for the role of paradigms of rational action.

I begin in §2 by identifying a hidden assumption of the reasons theory, the rejection of which points us towards possible alternative accounts of rational action. In §3 I introduce a promising alternative account, which I shall call *the permissive conception of rational action.* In §4 and §5 I fill out some of the details of that conception. In §6 and §7 I describe how the permissive conception accommodates both habitual actions and actions for reasons respectively. This puts me in a position to argue in §8, that the permissive conception allows us to meet the conceptual constraint.

In §9 I consider another sense in which it might be said that actions in general, and habitual actions in particular, are rational, and that is in the sense that they are *reasonable* things to do. I suggest that this sense of rationality can be given an externalist interpretation without intellectualism.

In §10 I summarise my findings. I will by then have explicated two senses in which an action's being habitual is compatible with its rationality. The idea of rational habitual action will thus be defended.

2. *An Assumption of the Reasons Theory*

In his book *Morality: Its Nature and Justification* Bernard Gert (1998) is critical of accounts of rational action which understand them to be "based on reasons" (p. 60). Gert does not go into the ways in which the expression "based on reasons" could be understood, so I cannot say whether he would agree with my arguments in Chapter 5.

Nevertheless, the explanation Gert gives for the failure of such accounts is relevant here. He writes:

> All attempts to provide a positive definition of rational action are inadequate because rational actions share no positive feature or even any significant similarity other than simply not being irrational. Defining a rational action simply as an action that is not irrational does not impose a fictitious and misleading uniformity on all rational actions. (1998, p. 61)

Unfortunately Gert does not say very much in defence of this interesting and fertile claim. In this section I want to describe what I find in Gert's complaint about "positive definitions" of rational action. In the next section I shall take up his positive suggestion. First, then, to Gert's complaint that attempts to provide "a positive definition of rational action are inadequate".

Gert attacks an assumption about rational actions that is contentious. The assumption is that every rational action is rational in virtue of sharing some "positive feature" with all other rational actions. Two clarifications are in order. First, I take it that the idea of a common "positive feature" refers to properties that all instances of a kind *have*, and that excludes properties that they *lack*. So, for instance, hardness is a positive feature common to all stones, but whilst all stones lack the property of being animate, a lack of animation is not a positive feature of all stones in the sense intended. Bearing in mind this qualification I shall hereon abbreviate the expression to "feature". This will become significant shortly.

Second, I take it that this supposed common feature of rational actions cannot merely be that they are just that: rational actions. For that, it seems reasonable to think, would yield no informative theory at all. Rather, the sought common feature is thought to be something else, something which *makes* all of those actions rational. Something which, in some sense, underlies all rational actions and licenses our calling them this. In other words, something *essential* to rational actions. Call this common feature the *rationality-maker*. Once we have assumed that there are such things, the philosopher's task is to locate such a rationality-maker.

In the case of the reasons theory, the posited rationality-maker is one or more of the agent's reasons. And this may seem as promising a candidate as any, if we are

optimistic about spelling out the idea of "the agent's reasons". But as we saw in the previous chapter, according to existing accounts of reasons, habitual actions do not seem to have a rationality-maker in this sense. As I read Gert, he suggests that we should put pressure on the assumption that makes this quest seem mandatory.

Wittgenstein (1953, §§66-77) pointed out that there are many kind concepts which work perfectly well despite the fact that the items to which such a concept applies may share nothing save being of that kind.[1] To use the hackneyed example, if we want to characterise the idea of a "game", a search for a feature shared by all games, and essential to them, will be in vain. Here we find the importance of the idea that a "feature" excludes lacks, since we might be able to think of something that all games lack, for instance, a spouse. But the lack of a spouse is not a feature of games in the required sense. There is no game-maker. This absence does not, however, inhibit our having a perfectly good sense of which things are games and which things are not. There will of course be marginal cases whose game-hood could go either way, and over which we could debate. But the existence of such cases does nothing to undermine the fact that there is large agreement on many clear cases of games and non-games alike. Monopoly is a game; the number 2 is not a game. When asked what licenses our calling some candidate "a game", we are not obliged to give the same answer every time. Rather we can point to a number of features which the candidate might be said to share with other items whose game-hood is undisputed, at least for the purposes of this exercise. In so doing we are pointing to what Wittgenstein famously calls "family resemblances" (§67).

If this is right for games, it cannot be assumed that for some given kind concept *C*, there is a *C*-maker which items normally classed as *C* share. It will have to be argued not only that there is some common element which accurately tracks our normal use of the term "*C*", but what is more, that the presence of this element alone licenses such uses.

Now terms with well established definitions may permit such arguments to go through. These might include technical terms such as the scientist's notion of water.

---

[1] For an account of Wittgenstein's suggestion see Bambrough (1961), to which I am indebted.

For the scientist, having the molecular constitution H_2O is the water-maker, though it may not be for the rest of us. There are other non-technical examples, such as unmarried man-ness being the bachelor-maker. In both sorts of case it is obvious how to continue the sentence "to be a [or some] *C* is to be ...". But when definitions are vague or contentious, as they tend to be in philosophy, completing such sentences will be impossible.

Prior to a theory of rational action, the notion has only a vague and contentious definition. Indeed, unless the definition were contentious, there would seem little point in seeking a theory in the first place! That is to say, before we have a theory in mind it is not obvious how we might continue the sentence "to be a rational action is to be ...". Because the definition is contentious, it is also contentious to assume that there is a rationality-maker. For all we know, "rational action" might be another "game". So it need not be a lack of ingenuity that has prevented us from reaching consensus on the definition of a reason. Rather, a presupposition of the quest, that there must be a rationality-maker, ensures that it will be in vain. This is a deeper reason for resisting a modification of the reasons theory.

The reasons theorist cannot deny that the definition of rational action is contentious by claiming that it is being employed technically. Employing a technical notion of rational action depends on *deciding* to do so, and supplying a rationality-maker. If this is not to be question-begging, the decision requires justification which shows how the technical notion refines our pre-theoretical idea of rational action. And the requirement to supply an independent argument for the essentialist assumption that there is a rationality-maker remains.

Even without such an argument, however, failing to make the essentialist assumption will seem to generate a serious problem. For giving up on the idea of a rationality-maker would seem to be tantamount to abandoning the theoretical enterprise of accounting for rational action. After all, how else can we give a unified account of the phenomenon of rational action, than by seeking a feature that all rational actions share?

The answer is not to abandon the idea of a single, unifying theory, for all else being equal, unity would be preferable to plurality. Rather, the answer is to weaken

our conception of what it means for a theory to be "unified". If we do this, we shall see that to abandon rationality-makers is not to abandon a single theory altogether. For there are at least two ways of offering an informative account of rational action, which can be said to unify in senses weaker than sharing a common feature. I shall describe the first of these in the remainder of this section, and argue that it is unsatisfactory because the sense of unification is in fact too weak. In the next section I shall describe the second, more satisfactory, way, proposed by Gert, which unifies in a stronger sense, without returning to essentialism.

The first way of keeping the possibility of a theory of rational action alive is to take a *disjunctive* approach. That is, to say that all and only rational actions are *either* such-and-such *or* so-and-so (*or* ...). The disjunction can be as long as we need it to be. This allows us to drop the assumption that there is a single rationality-maker, and instead suppose that there are many of them. In the present case we might say that all and only rational actions are *either* actions for reasons *or* habitual actions of a certain sort. Call this *the disjunctive theory of rational action*. But whilst this approach meets the conceptual constraint mentioned in §1, it does so only by unifying the phenomenon in a very weak sense, a sense which threatens the claim that habitual actions are genuinely rational.

In the disjunctive theory what we have is not an account which captures the way in which the multiplicity of things that we call rational actions are all the same kinds of thing. But they are only the same kinds of thing by dint of the disjunctive theory itself. This is tantamount to admitting that we do not have two ways in which an action can be rational, but a straightforward ambiguity in the term "rational". Compare a disjunctive theory of the concept "bank": all and only banks are *either* at the edges of rivers *or* are financial institutions which retain and lend money. The sense of unification is very weak since the disjuncts merely reflect two different meanings of the same word, which are otherwise completely independent. The theory does not capture any deep connection between riversides and financial institutions because there is none to be captured.

And a disjunctive theory of rational action does not articulate any genuine unity between actions for reasons and habitual actions, but is silent on the matter. It is

then open to an opponent of the view that habitual actions are rational to say that the disjunctive theory merely trades on an ambiguity in the notion of rationality, and reveals no more unity between actions for reasons and habitual actions than exists between financial institutions and riversides. The objector can deny that the "rationality" which is being claimed for habitual actions is rationality proper, but is at best of a second-class kind, and one which the serious philosopher of rational action can justifiably regard as marginal, if not a complete change of topic.

I think that this is enough to for us to put disjunctive theories to one side. And of course further motivation for us to do so comes from the availability of an alternative approach, which I think can be regarded as unifying in a stronger sense, without requiring rationality-makers at all. It is to such an approach that we now turn.

3. *The Permissive Conception of Rational Action*

I want to develop Gert's suggestion about how we should define the idea of a rational action. The result will be what I call *the permissive conception of rational action.* Like the disjunctive theory considered above, the permissive conception does not require a single rationality-maker, but unlike the disjunctive theory, it unifies the phenomenon in a sufficiently strong sense. This conception will put me in a position to argue, in later sections, that some habitual actions are rational, along with actions for reasons, and thereby to meet the conceptual constraint.

In the previous section we saw that Gert is, like me, suspicious of the idea that a rational action is one that is, as he puts it, "based on reasons". In Gert's opinion this is "misleading" because,

> on any plausible account of "based on reasons," these definitions exclude from the category of rational actions all those rationally allowed actions that are done without a reason, or simply because one feels like acting in that way. Although all rationally required actions can be described as based on reasons, rationally allowed actions are also rational, and many of them cannot be correctly described as based on reasons. (1998, p. 60).

Here we have a suggestion about why the idea of a rationality-maker might be a bad way to characterise rational actions, namely that this assumes that all rational actions are rationally *required.* If we make this assumption it will be tempting to think that there must be something which underwrites that requirement for every rational action. This is the role played by reasons (whether internal or external) in the reasons theory. But according to Gert the assumption is questionable. For, he suggests, some actions may merely be rationally *allowed,* and it is clear from the above passage that Gert thinks that they too can be thought of as rational. If he is right about this, we shall be less inclined to look for something that underwrites the rationality of all rational actions because if an action is merely allowed, we will not be tempted to posit something which "makes it" allowed. Rather what we will check for the absence of criteria for the action's being *dis*allowed. So Gert suggests that we define a rational action "simply as an action that is not irrational" (1998, p. 61). This is what I call the permissive conception of rational action. Formally,

(PC) An action is rational if and only if it is not irrational.

Rather than supposing that rational actions are those actions which possess a certain feature, we instead suppose that rational actions are those actions which *lack* some feature, or set of features, which, were they present, would *disqualify* the actions from being rational. Instead of a rationality-maker we have one or more *rationality-disqualifiers,* specifically, ways in which an action can be said to be irrational.

I think (PC) unifies in a sufficiently strong sense since on the one hand it specifies something shared by all rational actions, namely that they are not irrational, so unlike the disjunctive account, it captures a genuine commonality. But on the other hand, the permissive account does not fall foul of the Wittgensteinian critique since what is shared is not a "feature" in the sense defined, but rather a lack of a particular sort of feature.

The permissive conception requires two ideas which standard accounts of rational actions do not require. First, we need to make sense of the idea of an action independently of whether it is rational. Accounts of action which tie rationality and

action-hood closely together will have difficulty making sense of this idea.[2] Second, we need to know what it means for an action to be irrational. After all, to use J. L. Austin's somewhat outdated though still useful phrase, according to the permissive conception of rational action, irrational actions "wear the trousers". I shall say more about these two ideas in the next two sections. Taking them both for granted for a moment, however, there is more to say in general terms about the permissive conception.

The permissive proposal may seem idiosyncratic, but it is not without pedigree. Bas Van Fraassen (1989) has made a similar proposal about rationality in the context of theory change in the philosophy of science. He thinks that if we subscribe to the idea of rationality Bertrand Russell assumes to be right, it is impossible to show how it can be rational to believe a theory which goes beyond any available evidence, and lacks support from previous opinions. Van Fraassen thinks we should instead reject that idea of rationality and adopt another. He contrasts these two conceptions of rationality by analogy:

> The difference is analogous to that between (or so Justice Oliver Wendell Holmes wrote) the Prussian and the English concept of law. In the former, everything is forbidden which is not explicitly permitted, and in the latter, everything permitted that is not explicitly forbidden. When Russell is still preoccupied with reasons and justification, he heeds the call of what we might analogously call the Prussian concept of rationality: what is rational to believe is exactly what one is rationally compelled to believe. I would opt instead for the dual: what it is rational to believe includes anything that one is not rationally compelled to disbelieve. ... *Rationality is only bridled irrationality.* (1989, pp. 171-2)

It is by using the "English" concept of rationality (which he later parenthetically calls the "the permissive concept of rationality", which is where I get the name) that Van Fraassen can show how, when we invent new theories which lack support, "we are not only prone but rational to embrace them" (1989, p. 172). He thinks that it is too demanding to ask for "reasons and justification" for our beliefs to count as rational,

---

[2] Examples would include the reasons-based accounts of intentional action referred to in Chapter 4, §2.

since the quintessentially rational enterprise of scientific theory belief would thereby be disqualified. Thinking of rationality instead as a permissive notion, which only disqualifies irrational beliefs, captures this thought and hence better reflects our ordinary understanding of the idea of rational belief. This is a view which Van Fraassen still defends today.[3] If it is right as a view about what it is rational to *believe*, why should the same not follow about what it is rational to *do*?

Van Fraassen's remarks also lend support to the thesis at hand since they suggest that rationality is not the only permissive concept. This should reduce any suspicion that the introduction of such a concept is the result of gerrymandering. For Van Fraassen is surely right to say that the ordinary English concept of legality is permissive in the way he describes. We can define English legality by saying that an action is legal if and only if it is not illegal, where what is illegal is explicitly stated in the statute books. So for example, if I am not sure whether building model aeroplanes is legal, the lack of mention of that activity in the statute books ought to satisfy me that it is. There is no other criterion of legality, and in particular, there is no positive characterisation of it. Legality then, appears to be a permissive concept with which we are all familiar, and there are others. Candidates that have occurred to me include "healthy" (not ill); "normal" (not abnormal); and "adequate" (not inadequate). I think permissive concepts are commonplace, and that just adds to the plausibility of the view that rational action might be one of them.

The permissive conception of rational action will be controversial at least in philosophical circles, since it gets rid of the standardly used three way distinction between rational actions, *non*-rational actions and *ir*rational actions. With the permissive conception there is no distinction between non-rational and irrational actions, so we only have a two way distinction on the rational-irrational axis. The category of actions which were, on the standard scheme, both non-rational and not irrational, are now admitted into the class of the rational (and the hope is that these will include some habitual actions).

---

[3] He employs the idea in a recent (unpublished) lecture.

Someone might object that a move which denies us a potentially important distinction, should be avoided. In response I would say two things. First, the standard scheme is not obviously the best way to divide things up. For according to that scheme the term "non-rational" is used not only to pick out those actions which are neither rational nor irrational, but also to pick out behaviours which are simply not assessable as rational or irrational in the first place, since they are not actions at all. All that the permissive conception does away with is the first of these categories, and leaves the second intact. This is a distinction which the standard scheme obscured, so the permissive suggestion does no worse in terms of the number of distinctions we are allowed.[4]

Second, I doubt the three way distinction is important. For philosophers seem to have had little use for the category of actions which are both non-rational and not irrational. Perhaps indeed its main purpose is to provide the conceptual space for awkward cases.[5] But if that is right, the fact that such cases exist should raise our suspicions about the adequacy of the theories which generate them, particularly if they include behaviours as widespread as habitual actions.

The permissive conception is a radical proposal for a number of other reasons. We might previously have thought that we could give an account of rational actions without delving into the nature of irrationality, but this now looks unlikely. Philosophical accounts of irrationality, whatever form or forms it may take, are essential to our understanding of rational action. Other debates are demoted. Specifically, if the permissive conception is right, the current debates about actions for reasons must be repositioned. The debates between those who understand the relation between reason and action to be causal and those who deny it, and between reasons externalists and reasons internalists, can no longer be seen as debates about

---

[4] We could, following Hursthouse (1991) introduce the terminology of "arational actions" to stand for actions which are neither rational nor irrational, but this would be to depart from Hursthouse's particular use of the term for actions performed "in the grip of an emotion" (p. 59).

[5] In a similar vein, Lowe (1978, 1980) identifies that according to standard analyses of intentional actions, there is a category of actions which is neither intentional nor unintentional, which he terms "non-intentional".

the nature of rational actions. Rather they are at best debates about *one way* of avoiding irrationality. And it remains to be shown that acting for reasons is the only, or indeed the most important, way.

There are significant challenges for the permissive conception. In Chapter 5, §2, we saw some of the benefits of the reasons theory, which a competing theory of rational action would do well to match. These were that the reasons theory accounted for the normativity, authorship, and etiology of rational actions. The permissive conception may not seem to measure up very well in these respects. The normativity of rational actions is captured only in the rather weak sense that one should do what is not irrational. Nevertheless I see no reason not to accept that this is a viable sense of normativity (though I shall suggest how we might bolster it in §9).

The authorship of rational actions is not captured at all by the theory of rationality, since there is no requirement that there be reasons belonging to the agent involved in acting rationally. However, this is not a problem since authorship is guaranteed by the account of actions *simpliciter* I shall shortly give.

Etiology may present more of a difficulty. Since the most popular candidate for causes, that is, reasons, may not be present at all in rational actions, we shall have to look elsewhere for an account of what causes them. I have one suggestion in this respect, which may work for habitual actions, since it connects with some remarks I made in Chapter 3, §8, about the explanation of habitual behaviours. I give the merest sketch here. For it seems plausible that in saying that an action is habitual we have thereby specified what causes the action, namely, her having the habit plus her encounter with the normal circumstances. We can borrow some terminology from Fred Dretske (1988, pp. 42-4) at this point, and speak of the habit as a "structuring cause", and the encounter with normal circumstances a "triggering cause". As Dretske points out, there is no need to privilege one or the other as "the" cause. They both have a contribution to make, as might many other factors. So the habit plus the encounter are amongst the causal antecedents of habitual actions. Clearly this

suggestion requires development.[6] But it shows that the permissive conception of rational actions is at least compatible with an account of the causation of those actions, even when reasons are absent.

I now turn to the two notions upon which the permissive conception relies: actions *simpliciter* and irrational actions.

4. Actions Simpliciter

The permissive conception of rational action (PC) requires the idea of an action which is not defined in terms of its rational status. We saw accounts which are defined in that way in Chapter 4, §2, as the reasons-based accounts of intentional actions of Anscombe and her followers. We now have another reason to reject those accounts, since they tie rationality and action-hood too closely together to facilitate the proposed account of rational actions. For if an action just is a behaviour that is done for reasons, it would seem that according to any sensible account of irrationality, it cannot also be irrational, so (PC) would be an empty theory. Fortunately, however, the account of actions developed later in Chapter 4, albeit specifically for the purposes of making sense of habitual actions, does not make the tie too close.

The proposal of Chapter 4 was that habitual behaviours are actions when we are personally responsible for them. And I argued that rational intervention control (RIC) is sufficient for responsibility, and that we have such control over some of our habitual behaviours. A natural extension of this idea is that personal responsibility can be used as a general criterion of action. It need not be restricted to those behaviours which are repeated and automatic, but applies equally to one-off behaviours, and those over which we deliberate.

This does not entail that all personal responsibility is underpinned by rational intervention control. For there may be other sorts of control which are also sufficient for responsibility, such as the control required to initiate a behaviour. But we saw in

---

[6] One thought is that if we conceive of a habit as *part* of the agent, the account sketched here may be compatible with some idea of "agent causation", along the lines of that developed by e.g. Chisholm (1976) , Taylor (1966) and Danto (1973).

Chapter 4 that this latter sort of control does not seem necessary for responsibility. On the contrary, we found that insofar as an agent is in a position to intervene appropriately in a piece of behaviour, she is to that extent responsible for it, *no matter how it is initiated.* Indeed the process which got the behaviour going may be a sound deliberation, it may be an unsound one, or it may be a purely causal process of some sort. Thus, if rational intervention control is sufficient for responsibility, this gives us at least *a* sense in which a behaviour can be said to be an action without thereby conceding that it is done for reasons. This I think provides enough independence between reasons and action-hood for the purposes of the permissive conception of rational action.

The idea of rational intervention control is not of course completely independent of the idea of acting for reasons. For only agents who are *capable* of doing what is rationally appropriate, which for current purposes can be understood as acting for reasons, can be said to have this sort of control over their behaviour. So, for instance, non-rational animals will not be capable of actions. Reasons then are still in the picture, but they are in the background. For reasons are only involved counterfactually, at the level of the agent's *capacities.*

Of course the above remarks hardly constitute an adequate defence of rational intervention control as a sufficient condition for actions in general. But as long as we don't think it is utterly implausible, it's success in making permissive rationality viable might motivate the pursuit of such a defence. The current test is whether this notion of an action can capture the idea of an irrational action, and to answer that, we need a better understanding of what the irrationality in question amounts to.

5. *Irrational Actions*

Irrationality is what Davidson picturesquely calls "a failure in the house of reason" (1982, p. 449). Irrational actions are a particular class of such failures. I want to begin with two general points about the sorts of things failures are before going into details about specific sorts of irrational actions which are relevant to habits.

First, it may seem that the permissive conception is violating a very general truth. Namely, that if we are going to decide what counts as a "failure" we must

begin with a conception of what successful cases are like. In other words, our notion of failure is parasitic on our notion of success. This is obviously true in some cases, as for instance, when one has built something with a specific purpose in mind, and it fails to fulfil that purpose. We know it fails only because we know what success would have been. But as a general rule, this is false. For we can often say when things have failed without having any clear conception of what "succeeding" would be. I can know, for instance, that a persistent alcoholic has failed in some sense as a person, without having any clear conception of what it would mean for a person to be "successful". The most I know for sure about such a person is that they are not a persistent alcoholic, and that is not to say a great deal about them. Here my idea of success is being constructed from, and is hence parasitic on, my ideas about failure. There is good reason for the priorities being this way round. For failures tend to disrupt our lives, and when they do, we notice them, and put them right if we can. Failures are thus the standard fare of classification and explanation. Successes, on the other hand, we can just rely on without comment or intervention, and they thus stay in the background, unnoticed.

This tells us two things about irrational actions when they are conceived as failures in the house of reason. First, we need not expect to begin with an idea of rational action in order to derive an idea of irrational action. This is another advantage of the permissive conception which depends only on an understanding of irrational actions. It reflects the fact that for everything we understand about irrational actions, we thereby know something about rational actions too.

Secondly, since the ways things can fail form a heterogeneous set, the same can be said of irrational actions. We can expect a plurality of ways in which actions can be said to be irrational, each one corresponding to a kind of explanation (or pointing to how an explanation might go) of what the failure consisted in. But there need be no common feature, or systematic connection of any kind between them. Unlike our concept of rational actions, irrational actions need not be unified in any way.

There is a limit on what can count as an irrational action, and that is that they must be actions, in the sense described in the previous section. If it is genuinely the

case that an agent could have done nothing about a supposed irrational action, then it is not an action. So there will also be a class of irrational behaviours, which will include the results of brain seizures, involuntary muscle spasms, and many bodily reflexes. The question of how much control an agent has over a particular piece of irrational behaviour, is, however, a delicate issue, and we can expect it to be varied and contested in particular cases.

There are some very familiar kinds of irrational actions. *Akratic* or weak-willed action, when an agent acts against her better judgement, has received a good deal of discussion.[7] Self-deceiving actions and actions which exhibit wishful-thinking are have also received attention. These phenomena are all philosophically interesting because they generate paradoxes, and paradoxes are the philosopher's favourite food. If we look at what the agent *says*, she seems to believe or desire one thing, but when we look at what she *does*, she seems to believe or desire precisely the opposite. The irrationality in these actions, then, is understood in terms of a problem with the coherence of the agent's psychological make-up.

But whilst I have no objection to including the above categories of irrational actions in our definition of permissive rationality, and I would refer the reader to the literature on the topics for a fuller understanding of these ideas, they will not suffice. For they look to the agent's reasoning capacities to understand the particular mistakes exhibited in actions, and hence exhibit an intellectualist bias. The problem is that since we are supposing habitual actions to be free from direct intellectual involvement, the above features of irrationality cannot be used to show how habitual actions might be irrational. In habitual action there may be no contrast between what an agent says and what she does, since she may have nothing to say which is relevant to what she does. And that does not mean her action is not irrational.

But not all forms of irrational action are of the above kinds. Davidson suggests that it might be "irrational, given the dangers, discomforts, and meagre rewards to be expected on success, for any person to attempt to climb Mt. Everest without oxygen (or even with it)" (1982, p. 450). The irrationality exhibited here cannot be understood

---

[7] See Walker (1989) for a survey of the issues surrounding weakness of the will.

in terms of the agent's incoherence, but rather in terms of the kind of activity in question. When it comes to Everest assaults, Davidson thinks the negative considerations always outweigh the positive ones, so it is always irrational. The agent who does it anyway evidently does not properly appreciate the relevant facts and their respective weights, and is for that reason irrational.

Now this sort of irrationality could be given an intellectualist interpretation. We could say that an Everest summiteer is irrational because she made a mistake in her *reasoning* prior to the attempt. In other words, she knew all the relevant facts, but just hadn't, for some reason, worked out that some facts should weigh more heavily, either positively or negatively, than others. Thus it was her thought processes that were at fault.

But we need not foist that sort of interpretation on the idea. For we can say instead that her failure was to *see* (or "appreciate") which considerations were relevant, and thereby important, to her planned assault. This is a failure in perception, and not in deliberation. For if the failure happens at this perceptual stage, the deliberative processes don't get the chance to go to work on the relevant subject matter. This perceptual form of irrationality is in this way prior to deliberative irrationality. And this is one sort of irrationality which can be exhibited in habitual actions.

I now want to list three ways in which this "perceptual" irrationality can manifest itself in habitual actions.

(i) Inattentiveness

When an agent fails to notice some salient feature, or features, of a situation which are not only relevant to some prospective action ϕ, but would make it the case that she should not ϕ, I shall say that she is *inattentive*. This might sometimes be called "absent-mindedness", but it is characteristically a perceptual failure. Inattentiveness is a failure in the watchfulness required for rational intervention control. That doesn't mean that behaviours over which one is inattentive are thereby disqualified from being actions, on the grounds of lacking rational intervention control. For one still has such control, even though one is not exercising an aspect of it.

The simplest cases of inattentiveness involve missing something that is either literally "right in front of the agent", or at any rate, very easily available to the agent's perceptual faculties. For example, whilst typing at my computer I fail to notice that the screen, right in front of me, has frozen. This is a reason to stop typing, but my inattentiveness prevents me from responding to that reason.

There are more complex cases of inattentiveness in which the sort of action in question requires the agent to investigate, that is, to engage in supplementary actions in order to establish whether a salient feature is present. For instance, if Robin is driving to work by his usual route, he had still better keep an eye out for cars pulling out of side streets, or pedestrians unexpectedly crossing in front of him. Such vigilance will typically require specific acts of looking to see. If he fails to do these things, his driving is reckless, culpable, and irrational.

The requirement to investigate should not be overplayed. For, on pain of violating the conceptual constraint, it cannot be a failure of rationality to omit to carry out a "complete" investigation into all the considerations relevant to a particular action.[8] Indeed, if we can make sense of the idea of a "complete" investigation, the attempt to carry one out would tend to degenerate into a kind of irrational action of its own, a kind of obsessive behaviour. Nevertheless a case of inattentiveness might involve a failure to investigate any one of a number of important features that might be relevant to a given action.

Other more familiar kinds of irrationality such as weakness of the will, self-deception, and so on, all depend upon the absence of inattentiveness. That is, they presuppose that some feature of a situation has already entered the agent's purview. But there is always the possibility of breakdown prior to any kind of epistemic relation between agent and feature, being established. This kind of irrationality cannot be expressed as a paradox because there is no inconsistency, as it were, "within" the agent, that is, between what she does and what she thinks and feels. From the agent's point of view everything is fine. But inattentiveness deserves to be

---

[8] The suggestion that rationality entails "full information" is made by Brandt (1979), and criticised as too demanding by Gibbard (1990, pp. 18-22).

thought of as a breakdown in our rational capacity because without it our very ability to respond appropriately to reasons to do otherwise is compromised.

(ii) Failure to Deliberate

We saw in §4 that even with permissive rationality the capacity to act for reasons remains a necessary constituent of agency, whether or not we actually exert that capacity in a particular action. An exertion of that capacity involves deliberation, even if it is sometimes deliberation of a very rudimentary kind. What is crucial to agency, then, is that agents can judge when it is appropriate to deliberate about something, and when it is not. This capacity cannot itself be understood as the result of a process of reasoning, since any such process presupposes a judgement to begin reasoning. It is another sort of pre-intellectual capacity that is presupposed by any process of deliberation.

I think it is helpful to think of this capacity in an Aristotelian idiom, that is, as a virtue of rationality. As with other Aristotelian virtues, this virtue comes with a pair of vices which the rational agent avoids. These vices are respectively too little deliberation, and too much.

An agent deliberates too little when she perceives circumstances which are unusual, important, dangerous, or otherwise challenging, she has time to deliberate, but she does not. She instead acts automatically. This might be the exercise of a habit, but might be an automatic action of some other sort, such as an "instinctive" action, which unlike habit, is unlearned. In such circumstances, given that she had time to think and did not, we might say that she acts in a foolhardy way. Of course if the agent left it to instinct, she might get lucky and take the best course of action. If the challenging circumstances were normal for some habit, exercising that habit might increase her chances of doing the best thing. But whether she acts habitually or in some other automatic way, given that the circumstances are of this kind, and she has time to deliberate, she would give herself the best chance of acting well by doing so. A failure to deliberate when deliberation is appropriate constitutes a failure of rationality. As for inattentiveness, this is not a failure *within* reasoning, but is a failure to deploy it.

(iii) Excessive Deliberation

An agent exercises the rational vice of deliberating too much when she perceives circumstances with which she is familiar, and to which she has a natural, perhaps habitual, response, and she has a limited amount of time at her disposal, yet still deliberates before she acts. Those who hold the reasons theory might resist the suggestion that this kind of behaviour is irrational, arguing that if the agent fastidiously considers what reasons there might be, she only increases her chances of getting things right. But I think that on the contrary, deliberation in some circumstances can be a privation of rationality.

Some good examples of when deliberation can be excessive are the skilled actions required for sports. If I am in the middle of a sculling race, possibly the worst thing I can do is to think about which movements to make with my arms, hands, shoulders, back, or legs. Notoriously too much in the way of detailed deliberation on particular movements would cause me to lose my rhythm, and typically, fluff a stroke. This is the difficulty of only being able to think about one or two things at a time. I would do better to think about more general things like relaxation, my position in the race, and so forth, and trust my habits, or even my instincts, to take over the details of my bodily movements. It might of course be helpful to focus my attention, from time to time, on some particular aspect of my rowing technique in order to improve it - and indeed I shall do lots of this in training when there is sufficient time - but such deliberation must be used sparingly when there is limited time (as in a race) and is anyhow only possible if I trust my many other habits to look after the rest of my stroke. When there is limited time, excessive deliberation on familiar techniques, then, is irrational in the sense that it is a poor means to an end.

This rational vice of excessive deliberation on automatic actions comes up again and again. For instance, it would be a rational vice to think about how I am going to pronounce every word when I am speaking. Most likely, I would lose my sense of what I was saying. Of course such deliberation could be helpful were I in some abnormal circumstances, such as being coached in my pronunciation by a speech therapist. But in normal circumstances, such a procedure would be irrational. Another example comes from ethics. If I see somebody in urgent need of my help, say

they are in a life-threatening situation, and they need me to call an ambulance, pausing to deliberate might well be a matter of life and death. Thinking about whether I should really help them, or about how to dial the numbers on my telephone, would both cause unnecessary delays. Circumstances, then, to a large extent, determine not only whether it is rational to deliberate, but also what it is rational to deliberate about. Some self-awareness is also required, that is a kind of confidence in what one can do without deliberating. Nevertheless, deliberating too much is a failure in the rational capacity to judge what is appropriate in the circumstances.

That concludes this brief survey of what might be included as the irrational actions upon which (PC) depends. I make no claims about the survey's adequacy, since I am sure there are many other ways in which rationality can break down. Indeed I shall suggest another way of acting irrationally in the next section once I have introduced the idea of an "irrational habit". However, this short list gives us some grasp of forms of irrational actions which are classified in a way that is not entirely within the agent's point of view. In this way they contrast with the intellectualist classifications of irrational actions with which we are more familiar.

This then completes the outline of the idea of the permissive conception of rational action. The key test for the permissive conception, however, and one which it will have to pass if opponents are to be persuaded to accept all this, is whether it can accommodate enough habitual actions to satisfy the conceptual constraint.

6. *Accommodating Habitual Actions*

I think that according to permissive rationality, some habitual actions count as rational whilst others will not. Take the example of Killian who has the habit of walking the same way to work. One completely ordinary morning, with nothing whatever remarkable about it, Killian exercises this habit, and unsurprisingly gets to work as usual. I think that permissive rationality allows us to count such actions as rational, since Killian's action is not irrational. But I think we can develop this thought into a more general criterion for when exactly habitual actions are rational. I propose that the specifically *rational* habitual actions are those which are exercises of

rational habits which take place in *non-defeating circumstances.* Here I am introducing two new terms.

First, the idea of a rational habit. We noted in Chapter 5, §8, that some habits seem to be good, others bad. Having the habit of washing up after mealtimes is a good one to have, whereas leaving the dishes dirty for days is a bad one. Doing regular exercise is a good habit, whereas watching several soap operas every day is a bad one. So we have an intuitive sense of which habits are good and bad, prior to speculating on what might make such habits good or bad (something which I do not intend to do, though it might be an interesting avenue to pursue).

The idea of a rational habit, as opposed to a good one or a bad one, is easier to define, since we are already working with a notion of the rational, as it pertains to actions. A natural extension of the permissive conception of rational action, would be a permissive conception of rational habit. The proposal would be:

(RH) A habit is rational if and only if it is not irrational.

Analogous to the conception of rational action we have been discussing, the idea of a rational habit is parasitic on the idea of an irrational habit. This way of defining things prevents us from looking for a rationality-maker for habits themselves, and avoiding the criticisms of essentialism discussed earlier. Irrational habits, in turn, can be defined as follows:

(IH) A habit is irrational if and only if its exercises are irrational.

This means we can include the bad habits mentioned above in the class of irrational habits. So there are good reasons not to have the habit of leaving the dishes dirty, for instance, that the food hardens on, it looks untidy, and it means there is nothing clean for others to use. Somebody failing to notice these reasons could be guilty of either of the irrationality of inattentiveness, or, if not, the irrationality of failing to deliberate when it is appropriate (they will not have deliberated too much since the actions in question are assumed to be habitual). We could give a similar account of the irrationality of the habit of watching soap operas, though using different reasons.

Could it ever be rational to exercise an irrational habit? One might think so. There might be, for example, a large charity donation being offered for a person to

watch soap operas for weeks on end. So the person who has the otherwise irrational habit is instead doing something not irrational, but even rather admirable.

But I think there are problems with this claim. For a start, in such cases we need to be sure that the habit is still the same one given this new piece of information. In particular we need to know that the normal circumstances for the habit do not include the possibility of the donation. This would admittedly depend on the truth of various counterfactuals, so is difficult to verify. But even if the habit was judged to be the same one, there is good reason for thinking that its exercise is still a result of inattentiveness, and/or a lack of deliberation on the part of the agent. For given that the habit is irrational in the first place, there is *always* reason to say that the agent exercising the habit is inattentive or failed to deliberate appropriately, since its exercises are by definition irrational. So in exercising a normally irrational habit, the agent is always failing in rationality in this respect. Of course following a deliberation, the agent might come to see that this is a habit worth having, but that transition could give us reason to think that the normal circumstances have changed, and this would in turn lead us to think that the habit is now a rational one. So it is hard to see how exercising an irrational habit could ever be rational.

If the above is right, it gives us a fourth way in which an action could be said to be irrational, to add to the three listed in the previous section, namely, if it is an exercise of an irrational habit. So if we know that the habit being exercised is irrational, we need not enquire further about the circumstances, or the ends being served. We know that the action is irrational under its description as an exercise of a habit of that kind.

Rational habits form a large and heterogeneous group, and include good habits like doing the washing up and taking regular exercise, as well as Killian's walk to work, which has nothing much to be said in its favour. We can also include going sculling every other day, shaking hands when meeting people, using the word "dog" to refer to dogs, and getting up at the same time every morning. The innumerable routines that form the background to our everyday lives, some of them idiosyncratic, others shared, are rational in this sense.

I see no reason also not to include the three varieties of habit considered in Chapter 2. Aristotelian moral virtues such as courage, temperance and patience, are rational habits, though of course more can be said in their favour, since their possession is taken to constitute a flourishing human life, *eudaimonia.* Ryle's behavioural dispositions and Wittgenstein's learned responses would also fall into this category of habits.

The fact that a given habitual action is the exercise of a rational habit, is, however, insufficient to guarantee that the action itself is rational. For a rational habit can be exercised in circumstances perfectly normal for that habit, but it could still turn out that doing so was irrational. For it could be that there are what I call "defeating circumstances". These obtain when there are reasons *not* to exercise the habit as one normally would. In the terms used in Chapter 4, §7, defeating circumstances are reasons for major interventions.

To clarify. From our definition of repetition back in Chapter 3, we know that all habits have normal circumstances. They are part of the identity of a particular habit. But an agent could still be in normal circumstances, yet there could be other facts that obtain, which fall outside our understanding of normal circumstances, which mean that to exercise the habit on this occasion would be irrational. In the example of Killian's walk to work, defeating circumstances would include unexpected eventualities such as there being flooding or noisy road-works which would disrupt his habitual action, or Killian having made a decision to go a different way today. These would all be reasons for him not exercise the habit on this occasion. If such defeating circumstances obtained, and Killian still did what he usually does habitually, and at least begins his journey, it would seem right to say that insofar as he does that, he is exercising the same habit that he usually does. This is usually a rational thing to do, but today it is not, and today he is inattentive to, or fails to respond appropriately to, the defeating circumstances. In such circumstances we can say that Killian acted irrationally, even though he exercised a rational habit.

As long as defeating circumstances do not obtain, the agent exercising a rational habit cannot be charged with inattention or lack of appropriate deliberation, since there is nothing to be attentive to, or deliberate appropriately on. Hence, we can

conclude that whenever a rational habit is exercised in non-defeating circumstances, by the permissive conception of rational action, that action is rational.

7. Accommodating Actions for Reasons

If the permissive conception could accommodate only exercises of rational habits, that may already be enough to meet the conceptual constraint since the bulk of our actions seem to be habitual. However, if it can also accommodate actions for reasons, it would meet this constraint more fully. And the better it meets the constraint, the more powerful and plausible the permissive conception is.

It may seem obvious that the permissive conception will allow us to count actions for reasons as rational. After all, what better way of avoiding irrationality is there than to reflect on the reasons that present themselves to one? However, to think that just by invoking her rational capacity, an agent can guarantee that an action is rational, is to credit the individual with more power than she has. We all make mistakes in our reasoning some of the time. What is more, individuals are not immune to making mistakes in their assessments of their own mistakes. The regress could go on indefinitely. Of course that is not to deny that we get reasoning right a good deal of the time. Indeed, the very possibility of something counting as "mistake" in reasoning presupposes a background of successful reasoning. The problem is that whether a particular instance of reasoning is successful is not completely in the hands of the individual agent who is engaged in that reasoning. Luck is also sometimes involved. Whether an agent reasons successfully prior to acting is an objective matter, settled by such things as whether it survives criticism from others.

This is a parallel thought to one used in epistemology to support "externalist" conceptions of propositional knowledge (see for example, McDowell 1995). Whereas an individual can be sure about what she believes, the question of whether that belief amounts to knowledge is out of her hands. Knowledge is, if you will, *successful* belief,

and success is an objective matter, which is to say that it survives communal criticism.[9]

Given all this, the claim that actions for reasons can be accommodated by the permissive conception can be qualified as follows. Insofar as "acting for reasons" is thought of as what Ryle (1949, pp. 143-7) calls an "achievement" notion, picking out only successful deliberations leading to the right actions, then it immediately follows that the action meets no standard of irrationality, for success just is the avoidance of this, as well as other, failures. This seems to be the understanding of actions for reasons Davidson has, as his analysis applies only "*when* the reason explains the action" (1963, p. 3, my italics). It follows that all actions for reasons in this sense will be rational according to the permissive conception.

But insofar as "acting for reasons" is thought of as what Ryle calls a "task" notion, that is, the agent's attempting to secure success in action by deliberation, then all sorts of things could go wrong. When something does go wrong, it will be due to some sort of irrationality on the part of the agent, and accordingly the permissive conception will deny the action rational status. What may seem surprising is that the action would be irrational even though there may have been deliberation on the part of the agent, and from the agent's point of view, at the time of action, that deliberation could be flawless. We can expect these sorts of occasions to be relatively rare, since one will learn from ones' mistakes, but the possibility remains.

I am not saying that any philosophers actually conceive of actions for reasons as a task notion, as opposed to an achievement notion. But what is crucial here is that if we are to include actions for reasons as rational under the permissive conception,

---

[9] In the cited paper McDowell points out that it is not mandatory to conclude from this that all knowledge amounts to belief plus some extra ingredient which makes it successful, an "internalising" move which encourages scepticism. Rather, some beliefs just are successful, and hence do not fall short of knowledge. The parallel with actions for reasons is that we should not think that all actions for reasons amount to the agent's "trying" to act for reasons plus some extra success ingredient. Rather, many attempts at acting for reasons just are successes. We can say this without the sorts of revisions proposed by O'Shaughnessy and Hornsby, discussed and rejected in Chapter 3, §7.

we must understand this idea as indicating a success in reasoning and action from an objective viewpoint.

8. Meeting the Conceptual Constraint

If the permissive conception of rational action is right, that may seem to make available a quick way of meeting the conceptual constraint. As a matter of empirical fact, it could be said, agents do not act irrationally very much of the time, so it follows from the permissive conception that whatever we do for the remainder of the time, must be rational. But whilst I have some sympathy with this move, I think it may owe too much to the assumed truth of the conceptual constraint. Hence I think it is intellectually healthy to be sceptical about the supposed scarcity of irrationality. For, it could be that we act irrationally rather more of the time than we would like to think, and particularly when we act habitually. I want therefore to consider some arguments which support the idea that we act rationally more of the time than we do not. This will give us independent reasons for thinking that the conceptual constraint should be met.

I want first to sketch two kinds of argument that may be available to support the idea that we have more rational habits than irrational ones. A first kind of argument appeals to human nature and facts about cultural evolution. We are reflective creatures who can detect irrationalities in what we do. Once we have detected an irrational habit, perhaps by having had it pointed out by somebody else, or by self-observation, we can often modify such habits, or lose them altogether. Of course some habits are difficult for us to change or lose, no matter how irrational we think them. But it is in our nature to do all we can to make sure that future generations do not inherit such habits. Knowing our children naturally copy us, we refrain from exercising them in their presence; if we see them doing such things we persuade them against it, with carrots or sticks, and so forth. Because of this, agents, who have a long history of self-criticism, spanning many generations, tend to have rather few irrational habits. We will of course often notice the irrational ones in others, but that is not because they are common, but because they are so unusual. Arguably, as long as they are relatively harmless, some such idiosyncrasies are to be

encouraged because they are what distinguish us from each other. But in virtue of the contingencies mentioned, unusual they nevertheless are.

A second kind of argument for the claim that we have more rational habits than irrational ones mirrors a style of argument used by Davidson (1984). He argues that in order to interpret an individual of whom we currently have no understanding, we must begin by assuming that she has mainly true beliefs about the world. We make this assumption as part of what Davidson calls the "principle of charity". Without this principle, Davidson thinks, we could not begin to interpret what such an individual is doing. In this way we can say that a condition of agency is the possession of beliefs which are mainly true.

We might say that an assumption similar to that about beliefs is necessary with respect to an individual's habits. In order to interpret an individual, we must assume that she has mainly rational habits, on pain of her being opaque to us. If, for example, she exhibited weakness of the will more often than she did not, if we could interpret her at all, it would be with extreme difficulty. The more of her actions that were weak-willed, the less we would be convinced that she had any rational control over what she did, and our willingness to consider her an agent capable of responsible action would be undermined. Persistent exhibitions of other kinds of irrationality, would, I think, present similar difficulties for an interpreter. And the exhibition of a number of different kinds of irrationality could make things almost impossible: imagine trying to make sense of the unfortunate individual who exhibited either weak-willed or self-deceiving behaviour most of the time.

The suggestion is that the assumption that most of an individual's habits are rational habits would have to be part of a Davidsonian principle of charity. Indeed, I suspect that we may be able to say that all there is to the principle of charity is the assumption that an agent has mainly rational habits, the possession of true beliefs being one such habit.[10] Whether or not this suggestion can be carried through, we can still press the analogy: just as rationality can be said to necessarily involve mainly true

---

[10] Similar things might be said of the rational habit of honouring *modus ponens*. See further n. 11 below and Chapter 8.

beliefs, so agency can she be said to necessarily involve mainly rational habits. Of course the Davidsonian approach is not without difficulties, and the specific claim about true beliefs is not beyond criticism.[11] But it would nonetheless seem to be a promising line of argument to support the claim at hand.

But even if it is true that we have more rational habits than irrational ones, that is not enough to meet the conceptual constraint. Even though we rational agents may have a set of (mostly rational) habits, we may not actually encounter normal circumstances for those habits very often. And even when we do, we may not act habitually in those circumstances all of the time, due to defeating circumstances (though we must for a majority of the time on pain of undermining our claim to have that habit). So it could still be that most of the time we do not exercise rational habits. But I want to put forward some considerations which count against this conclusion.

I shall first suggest that we encounter normal circumstances for some habit or other most of the time. Second, I shall suggest that even when we opt out of habitual actions for defeating reasons, we often act rationally. And third, I shall suggest that even in abnormal circumstances, we tend to act rationally too.

First, then, why should we think that we encounter normal circumstances for some habit or other a good deal of the time? Here I appeal to the way in which human life is conducted, that is, in surroundings that are generally stable. We tend to live in the same kind of environment for protracted periods. In the West, we tend to live in the same town, amongst the same people, who consistently do the same kinds

---

[11] One difficulty with the Davidsonian principle is with specifying exactly which beliefs we should count. We would presumably not have too much difficulty interpreting a scientist from three hundred years ago, although a good many of his scientific beliefs might be false, and those beliefs could outnumber his true ones. Replacing the Davidsonian principle with a principle involving only rational habits (see n. 10) could help here. In holding the beliefs he does the scientist is not irrational. In particular, he is not inattentive to the facts, and nor is he making any error in reasoning. That modern-day facts are not available to him is not a problem with his rationality. He may arguably be termed "irrational" once our thought experiment propels him into a modern-day context, when we can (perhaps unfairly) accuse him of inattentiveness to the present-day facts. So we have no difficulty in interpreting an individual who has

of things. Even nomadic people have a great stability in their immediate surroundings, in the group of people around them, their carried equipment, and stable practices. Our daily needs are also stable: we need to eat, drink, exercise, do a job, support a family, enjoy what leisure we have, and so forth. A result of these two kinds of stability is that in meeting our needs we interact with our home environment in ways that we find that, in a broad sense, works. What is more, it is easier for us to do things in ways that have worked before than to keep re-inventing new ways of doing things. Whether we like it or not, we become, as we say, *accustomed* to our home environment. In other words, we acquire a set of habits tailored to that environment. What seems true, then, is that the more we live in a place, the more habits we acquire for living in that place. Virtually every place becomes associated with some habit or other. Some examples from my own life include: the sitting room is where I relax; the kitchen is where I cook or eat; a jar is what I open to get at its contents; the banister is to be held on to; and the step-daughter is to be welcomed back from school. There are very few contexts around my home in which there is not some habitual way for me to act. The same thing could be said for some contexts around Durham, around the Philosophy Department, and so on. These contexts are normal circumstances for my habits. It seems reasonable to say, then, that when agents are in a home environment, they encounter the normal circumstances for some habit or other most of the time.

Of course, an agent's being in normal circumstances for a habit is not enough to guarantee that they will act rationally. The habit may be irrational, though for considerations given above, we can expect such habits to be in the minority. But even when the habit is rational, an agent's simply being in normal circumstances for that habit is not sufficient for her to act habitually, and hence rationally. Because agents have rational intervention control over their habitual actions, they may, on any particular occasion, decide, for defeating reasons, to do something other than act habitually. She can only do this on a minority of her encounters with normal circumstances, otherwise it will become doubtful whether she has the habit at all. But

---

the rational habit of doing what he can to establish the truth. But it is too strong to say that this habit always results in true beliefs, or even mainly true beliefs, which is what Davidson requires for rationality.

even when she does opt out in this way, she will often still be acting rationally. This is because exerting rational intervention control is rational, and as we are currently understanding things, that involves acting for reasons, and when this is successful, as it usually will be (see §7 above), the agent avoids irrationality. It follows that opting out will generally avoid irrationality, and so according to permissive rationality, opting out will be rational.

Agents encounter plenty of abnormal circumstances too. Somebody I have never spoken to before telephones me; there is a new problem in my research; the bath tap is dripping; we are going out to dinner with friends tonight. Although the more such things occur, the more likely it becomes that my associated response, if it is similar each time, will become habitual, there need be no habits that go with such circumstances at present. Because even the most predictable home environments will be disrupted at some points, we can say that there must be some such abnormalities.

When I am in circumstances for which I have no associated habit I am forced back onto other resources for determining how I shall behave. One possibility here is that I act instinctively, that is, I act automatically but unlike in habitual action, instinctive actions bear no necessary connection to what I have done before. Instinctive action is sometimes all we can do when time is limited. However, as we saw in §5, leaving things to instinct may result in irrational behaviour. If time allows, it is better to temper instinct with deliberation. As time often does allow, we tend to be rather good at avoiding irrationality in this way. So even when we are in abnormal circumstances, we generally act rationally.

In summary, it seems that there is good support for the claim that permissive rationality allows us to meet the conceptual constraint. I have sketched why I think we have more rational habits than irrational ones. I have also suggested first, that we encounter normal circumstances for some habit or other a good deal of the time; second, that even when we opt out of habitual actions, we often act rationally; and third, that even in abnormal circumstances, we will tend to act rationally. These considerations all provide independent support for the intuitive idea that most of the time, rational creatures act rationally. Because the permissive conception of rational

action makes available such support, that, in turn, gives us an argument in favour of that account of rational action.

9. *The Reasonableness of Habitual Actions*

Some may still not be satisfied with the permissive conception of rationality as an account of the rationality of habitual actions. The problem, they may say, is that it just does not do justice to the normativity of rational actions, that is, the sense in which we *should* do them. The reasons theory at least gave us a positive account of that, and surely we can do better than saying that rational action is merely a matter of avoiding irrational action.

I agree. I think we can give more substance to the normativity of rational actions, and without returning to the reasons theory or any other sort of essentialism. This will require a supplement to the permissive conception. What is more, I think that this supplement will give us some ways of saying, positively, that habitual actions in particular, should be done.

The idea here is to pick up on one of the everyday uses of the term "rational" when it is applied to actions and other things. When we apply the term in this way we sometimes mean nothing more than that the action is a "sensible", or, to use a term I prefer, a *reasonable,* thing to do, as opposed to its being actually *reasoned.* Thus, an action's being reasonable need not imply anything about what the agent actually saw in the action at all. That is to say, it is possible to construct what I called in Chapter 4, §6, a *post hoc* rationalization, that is, a justification or *rationale* for the action, which makes explicit what is good about doing it, without implying that the agent had any sense at all of what that good is. The agent might think that there is something completely different that is good about it, may think that what is good about it makes it a bad thing to do, or may have no thoughts about it all. Thus an action can be reasonable without being done for reasons. Indeed agents are not the only creatures capable of reasonable behaviour. A rationale can be given for a much broader class of behaviours than human actions: the dog buries the bone to keep it safe from scavengers; the plant bends towards the sun to get more light.

This sense of rational says something more about an action than that it is not irrational. And whilst an action could not be both irrational and reasonable, for being irrational is a way for an action to fail to be reasonable, an action which is not irrational may or may not also be reasonable. So a non-irrational action might be reasonable, say, because it aimed at some good end. But a non-irrational action may also not be reasonable. Examples could include the one-off actions of idly twiddling my fingers, or moving my tongue inside my mouth. So being reasonable is a third sense in which an action can be said to be rational. As a result we now have three senses of rational action: (i) that which is done for reasons (reasons theory); (ii) that which is not irrational (permissive); and (iii) that which is reasonable.

Exercises of rational habits, I want to claim, are not only non-irrational, as I have already argued, but they are reasonable too. We have already seen one sense in which this might be true, and that is when the habit in question is a good habit. Exercising a good habit will always be reasonable since there is always something to be said for doing it, such as the beneficial outcome, even though on occasions, there may be better reasons not to do it.

But there is another sense in which even exercises of rational habits which are not good can count as reasonable. Such actions are reasonable because of certain considerations which could always feature in a rationale in virtue of the special features of their habituality. First, habitual actions always build on the absence of past failure. This arises from the fact they are repeated. And second, habitual actions are always a more efficient way of performing the same behaviour than the alternatives. That arises from their being automatic. Neither of these rationales depend upon the habitual actions in question having any particular outcome.

That habitual actions can always have such things said in their favour sets them clearly apart from finger-twiddling or tongue-moving in terms of their rationality. There is no past success to build on in the finger-twiddling and tongue-moving cases, so the first kind of rationale is not available. And whilst it is difficult to see what is being achieved by finger-twiddling and tongue-moving (a feeling of satisfaction?), it is unlikely that not doing these things will be an any less efficient way of achieving anything, since there really is no point to them anyway. I don't want to

put too much importance on how we class activities like finger-twiddling and tongue-moving. For it may be possible to argue that they too are reasonable, and intuitions may go either way here. However, to simply say that an action is rational because it lacks irrationality is not going to be all that can be said in favour of its rationality. Also saying what is reasonable about it supplements that story. And for habitual actions just such a supplement is available.

That is not to say that acting habitually is always the best thing to do, or the most reasonable thing to do all things considered. In particular situations there may be defeating considerations which would mean that acting habitually would be disastrous. However, that does not affect the fact that in general, there are some things to be said for acting habitually, even though for some habits (the irrational ones) and in some situations (defeating circumstances) these considerations are outweighed. The reasonableness of habitual actions is now secured.

Questions arise about how we might incorporate this result into our theories. One possibility is that somebody will want to co-opt this new sense of rationality into a modification of the reasons theory and thereby take up the suggestion of Chapter 5, §9. The idea is that we understand "reasons" in a more radically external sense than reasons externalists have thus far given. "Reasons" then are the considerations in virtue of which an action is reasonable, and whilst they can be understood as states or possible states of the world, they are not required to be within the agent's purview at all. They are not *the agent's* reasons in the reasons theorist's usual sense of that term. If what I have said about all habitual actions being reasonable is right, this could mean that we no longer need the permissive conception.

However, whilst I have some sympathy with this proposal, there are reasons to resist it. The most obvious is that in adopting it, we are departing from the terminology accepted by reasons theorists. "Reasons" in this new sense, are now things of which the agent may be totally ignorant, so will be unrecognisable even by reasons externalists. In a recent book Stephen Toulmin (2001) supports keeping such terminology separate: he argues that it is often useful to hold apart our sense of what is rational from our sense of what is reasonable.

More seriously for present purposes, externalist theories of actions for reasons which depend upon reasons being in view, such as those of Dancy (2000) and Stout (1996), would have to be discarded. This would leave us without an account of actions for reasons with which to deal with the counterfactual scenarios which underpin authorship.[12]

A third, I think decisive, reason to resist the suggested modification of the reasons theory is that the anti-essentialist considerations discussed in §2 of this chapter deprive us of any grounds for thinking that reasonableness constitutes one common feature of all rational actions. It seems that all exercises of rational habits do share this feature, and indeed, it may be that successful actions for reasons (on the traditional interpretation) may do also. But for all we know, there may other sorts of action, not considered here, which we would want to count as rational, yet which could not be considered reasonable. So we should resist the essentialist proposal. The idea of rationality may have too many overlapping threads to it to be pinned down in a single formula.

Nevertheless, we have done enough to show that if we allow that one of the threads of which the idea of rational action consists, is that of actions which are reasonable, then saying that an action is reasonable can only add to our confidence that it is rational. As far as the rational status of habitual actions is concerned, that can only help.

10. Conclusion

The failure of the reasons theory to accommodate habitual actions as a kind of rational action yields the paradoxical result that most of the time, rational agents do not act rationally. The reasons theory thus violates the conceptual constraint on accounts of rational action. In this chapter I have identified an unjustified assumption made by reasons theorists, namely, that all rational actions share a common feature which makes them rational, a rationality-maker. Rejecting this assumption allows us to

[12] In Chapter 8 I shall suggest that there may be an alternative to these views.

adopt an alternative account of rational action, the permissive conception, which shows the phenomenon to be unified without positing a rationality-maker.

I then argued that according to the permissive conception of rational action, not only can actions for reasons count as rational, but so can many habitual actions, namely, those that are exercises of rational habits in non-defeating circumstances. I have further argued that if we adopt the permissive conception, this allows us to meet the conceptual constraint. This gives further support to the permissive conception as an account of rational action. The corollary is that if the permissive conception is right, most habitual actions are rational actions.

Finally, I have sketched how habitual actions can further be said to be rational in their own right, in virtue of being reasonable things to do. Insofar as we might want to allow that the rationality of an action involves its being reasonable, this supplements the permissive conception, and strengthens the case for thinking habitual actions are rational.

The idea of a rational habitual action is not oxymoronic, as reasons theorists might think, but rather, it makes good sense. An action's habituality and its rationality are not only compatible with one another, but the fact that an action is habitual can actually support the claim that it is rational.

7

HABITS IN META-ETHICS

1. Humeans and Anti-Humeans

There is a controversy in current meta-ethics concerning the nature of moral motivation. An account of moral motivation aims to explain what it means to say that an agent is "motivated" to act morally. This idea of motivation has two aspects, normative and causal. From a normative perspective, an account of moral motivation will explain the sense in which an agent should act morally. From a causal perspective, such an account will say what brings such an action about.

David Hume (1740) unwittingly set the terms for the contemporary debate. Michael Smith (1987, 1994) and Simon Blackburn (1984a, 1998) are leaders in this movement. Both writers call their accounts of moral motivation "Humean", and their accounts deserve this epithet in at least two respects. First, following Hume, they hold that agency can be divided into two distinct components: "reason" and "passion". On the "reason" side of this divide lie our cognitive faculties such as beliefs, and belief-forming mechanisms; whilst non-cognitive states such as desires lie on the "passion" side. Second, Humeans follow Hume in thinking that reason is "the slave of the passions" (1740, p. 415). Accordingly, Humeans deny that beliefs alone could show how an agent is motivated to act. For that we need a contribution from the passions: a desire. Humean accounts of moral motivation, then, hold that to explain what motivates an agent to act morally, we need to mention both a belief and a desire. Affinities with Hume's own view may end there.

The Humean view of moral motivation has not gone unchallenged. Opposition has come particularly from writers who think Aristotle provides the best understanding of moral agency, in which the idea of a virtue plays a central role. John McDowell has been a key player in promoting this Aristotelian view. In a number of papers McDowell (1978, 1979, 1981, 1985, 1987) has put forward arguments in which

he tries to show the Humean position to be mistaken. But whilst some aspects of the anti-Humean alternative are reasonably clear - for instance, McDowell (1978) denies that desires are required for moral motivation - other aspects of the position he prefers remain, at best, rather obscure. I shall be in a position to offer some clarification by the end of this chapter.

It will not escape the reader's notice that we have come across these two sets of writers before, in Chapter 5 in the discussion of whether habitual actions could be actions for reasons. I argued there that neither the reasons internalists, which include Smith and Blackburn, nor the reasons externalists, which include McDowell, could accommodate habitual actions as rational actions. Whilst I briefly rehearsed two arguments supporting externalism as a view of reasons (§6), there was no assessment of internalist defences. The arguments for the Humean position on moral motivation can be seen as part of such a defence.

What turns on this disagreement about moral motivation? Perhaps the most important related issue is that which divides *moral cognitivists* from *moral non-cognitivists*.[1] According to moral cognitivists, when, in our everyday discourse, we discuss ethical questions, such as what should and should not be done, or which things have or lack moral value, we are debating matters of fact. When we assert, for instance, that the woman should be allowed to choose whether she aborts her foetus, or that chimpanzees should be accorded some of the rights normally reserved for humans, we are saying something which is not just important, but also true. In contrast, moral non-cognitivists hold that these apparently factual claims are not genuinely factual, but are something else; perhaps, as Blackburn has it, sophisticated expressions of emotion. They are no less important for that, but according to the non-cognitivist, it is a mistake to think of these assertions as reflecting the way the world genuinely is, by some standard of genuineness.

The question of whether moral cognitivism or non-cognitivism is a correct view of our moral talk is intimately connected to the question of what it is to be

---

[1] I follow Wiggins (1991) in here preferring "cognitivist" (vs. "non-cognitivist") terminology to the "realist" (vs. "anti-realist") one.

morally motivated. If one thinks with the Humeans, that an agent always needs a desire to be morally motivated, it cannot be facts that motivate. Hence one will be hard-pressed to defend cognitivism. This is not to say that this has not been tried - indeed Smith attempts to do precisely this. But as we shall see in §2, his attempt fails. If, on the other hand, one denies that desires are essential for motivation, cognitivism follows much more naturally.[2]

It seems then that by offering an anti-Humean account of moral motivation, McDowell is making available a simple route to moral cognitivism. Unfortunately, however, he has yet to convince the Humeans that they have got it wrong. In recent books, both Smith (1994, Chapter 4) and Blackburn (1998, Chapter 4) have offered separate defences of their Humean views against McDowell's attacks. I think that if we approach the debate even-handedly, we will find that the Humean defences against McDowell's arguments are by and large successful, as I shall argue in §3 and §7. But if McDowell's arguments can be strengthened, whilst the Humeans might have won the battle, but they may yet lose the war.

In this chapter I suggest that the idea of a habit developed in earlier chapters could be the ingredient needed to make the anti-Humean case stronger in the face of these recent Humean defences. I shall consider the distinctive defences of Smith and Blackburn in turn. In each case I shall first describe the current Humean defence and its merits, before spelling out how, equipped with the idea of a habit, the anti-Humean can now respond.[3]

---

[2] Other theoretical motivations might also be at work. For if one is attracted by cognitivism, one will look first for an account of motivation that does not depend upon non-cognitive extras; and if, for some reason, one finds non-cognitivism attractive, one will not have any difficulty with the claim that such motivational items are necessary.

[3] I shall not be discussing the pair of defences of Humeanism by David Lewis (1988, 1996). Not only does he not engage directly with McDowell's arguments, but his defences are based on controversial mathematical models of belief and desire. As a result, I think that the version of anti-Humeanism I shall recommend will be immune to his arguments, since we can reject any reduction of habitual actions to such models.

As we shall see, the meta-ethical implications are significant. In particular, I shall be in a position to suggest that attention to habits promises to deliver a novel resolution to the cognitivism/non-cognitivism debate.

Finally I shall briefly discuss whether we can read McDowell as working with a notion akin to habit, in which case what I offer is merely a further (though necessary) articulation of what McDowell has already said; or whether he is working with a different notion, as Smith and Blackburn think. This discussion will show how the idea of a habit can shed light on how we should understand the work of this influential thinker.

2. *Smith's Moral Problem*

In *The Moral Problem* (1994) Smith formulates, and claims to resolve, what he takes to be "the central organising problem in contemporary meta-ethics" (p. 11).[4] The problem consists in the apparent incompatibility of three independently plausible propositions, as follows:

1. *Cognitivism.* Moral judgements of the form 'It is right that I ϕ' express a subject's beliefs about an objective matter of fact, a fact about what it is right for her to do.

2. *Internalism.* If someone judges that it is right that she ϕs then, *ceteris paribus*, she is motivated to ϕ.

3. *Humeanism.* An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume's terms, distinct existences. (p. 12)

Smith's solution is to propose a theory which allows these three propositions to be reconciled. He does this by combining Humeanism (3) with what he calls an "anti-Humean theory of normative reasons". According to the anti-Humean theory, statements such as "it is right that I ϕ", which Smith calls "normative reasons", are factual, and hence his position can be regarded as cognitivist, honouring (1).

---

[4] In this section and the next page numbers refer to this work unless otherwise stated.

However, these normative reasons according to (2), only motivate *ceteris paribus*. Smith claims that the Humean and anti-Humean theories can be joined together by employing the notion of a "fully rational agent", for which *ceteris* is *paribus*. He writes, "to say that we have a normative reason to ϕ in certain circumstances C is to say that, if we were fully rational, we would want that we ϕ in C" (p. 181). Thus, if the "fully rational agent" has a normative reason to ϕ, a desire to ϕ is generated, and at which point the Humean theory can explain how the agent is motivated to ϕ. If all this were right, the three premises would be reconciled and the moral problem solved.

Objections to Smith's position could take many forms, and I leave most of them to others.[5] I here merely note that his position has two undesirable features. First, it is highly theoretical. It requires interlocking theories, each making substantive claims, and each with accompanying technical terms, for each of its three aspects. Of course, such theories might be justified if non-theoretical solutions cannot be found, but I am not convinced that such solutions have been fully explored.

Second, Smith's account relies heavily on "psychological states", specifically beliefs and desires, conceived as elements of an "inner" reality; in Smith's terms, they are "psychologically real" (p. 96). Not only are there familiar anti-intellectualist doubts about such states so conceived, but we have already seen in Chapter 5, §6, how they are of doubtful use in rationalizing actions. These difficulties do not however mean that Smith's position can be dismissed, for he offers it in full awareness of a number of anti-Humean moves.

I shall have more to say later about how Smith formulates cognitivism in proposition (1), and about how the idea of a "fully rational agent" is supposed to resolve proposition (2). Smith's claims regarding both of these issues turn out to be controversial in the context of what I have to say about habits. For now, we shall focus our attention on Smith's defence of proposition (3), Humeanism.

---

[5] Critical notices include Dancy (1996) and Dreier (1996). Brink (1997), Copp (1997) and Sayre-McCord (1997) put their objections in an *Ethics* symposium, to which Smith (1997) replies.

3. *Humeanism Defended I*

Smith's defence of Humeanism is necessary in the face of anti-Humean moves made by McDowell (1978, 1979, 1981), amongst others.[6] For if the anti-Humeans were right and proposition (3) turned out to be false, there would be no inconsistency to reconcile, no moral problem, and no need for Smith's theoretical apparatus to save cognitivism. What, then, are the anti-Humean moves Smith is worried about? To see this we have to first learn something about Smith's initial position.

Smith's argument for Humeanism depends upon there being two, and only two, different kinds of psychological state. These are distinguished by having different *directions of fit*, a distinction attributed to Anscombe (1957, p. 56).[7] Roughly speaking, some states, the paradigm of which is belief, aim at fitting the world. As Smith wants to cash that out, "a belief that p tends to go out of existence in the presence of a perception that not p" (p. 115). In contrast, other states, the paradigm of which is desire, aim at the world fitting them. In Smith's terms, "a desire that p tends to endure, disposing the subject in that state to bring it about that p" (p. 115). There are complications with this distinction.[8] For instance, a desire that p normally entails a belief that not p, so is at least dependent on a state of the other kind. But let us accept with Smith that the distinction captures something important about the kinds of psychological states there can be. If we further assume, as I take it Smith does, that the direction of fit distinction is exhaustive, then a desire, or any other state which aims at the world fitting it, will be a necessary component of any explanation of what motivated an agent. For if we accept the distinction, it seems that only those states

---

[6] Smith also objects to anti-Humeans Nagel (1970) and Platts (1979, 1981), but since, by Smith's lights, McDowell represents the most convincing anti-Humean, I here focus on Smith's treatment of his arguments.

[7] The attribution to Anscombe may be unfair as she never uses the phrase.

[8] Humberstone (1991) discusses difficulties with characterising the distinction.

which aim at the world fitting them could account for an agent being motivated to do anything, speaking both normatively and perhaps also causally.[9]

Smith's concern, however, is that McDowell suggests that there may be states which have *both* directions of fit. Because such states would be both belief-like and desire-like, Smith follows J. E. J. Altham (1986) in calling them *besires.* Smith's example of such a state is "the besire that ϕ-ing is right" (p. 118). This is belief-like because if ϕ-ing turns out not to be right, the besire would be in error. It is desire-like because, I take it, it disposes the subject to bring something about.[10] If there were such states, this would undermine the direction of fit distinction upon which Smith's defence of Humeanism rests.

Now it is not entirely clear where in the McDowell paper Smith cites (McDowell 1978) the idea of a besire is to be found. McDowell certainly doesn't use the term in that paper, or in any other I have seen. However, I want to postpone exegetical matters until §10, and for now allow that Smith is justified in saying that McDowell posits besires. Hence Smith reads McDowell as holding what I call *the besire theory.*

Note that as it is specified here, the besire theory does not specify *when* a besire might be present. Hence the besire theory can come in various strengths: one might say that besires are expressed in *all* moral judgements, or one might say that besires occur rather less often than that. As long as one insists that there could be at least one possible case in which a besire is expressed in a moral judgement, one holds a besire theory. If it were right, and there are not just two kinds of psychological state, but three, this would undermine the direction of fit distinction.

Smith claims that the Humean can give the following "quite simple" response:

---

[9] In fact, this last only follows if we make the common assumption that direction of fit goes together with the causation of these states, which runs in the opposite direction. See e.g. Searle (1983, pp. 7-9, 96).

[10] Strictly, by Smith's characterisation, it disposes the subject to *bring it about that ϕ-ing is right*, whatever that could mean. I take it that Smith intends that the agent with such a state is disposed to bring it about that ϕ-ing *is done* (or even more simply, the agent is disposed to ϕ).

> it is always at least possible for agents who are in some particular belief-like state not to be in some particular desire-like state; ... the two can always be pulled apart, at least modally. (p. 119)

Smith tries to drive this point home by appealing to what Michael Stocker (1979) calls "depressions" in the following passage:

> Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that, there is less good to be obtained or produced, as in the case of a universal Weltschmerz. Indeed, a frequent added defect of being in such 'depressions' is that one sees all the good to be won or saved and one lacks the will, interest, desire or strength. (Stocker 1979, p. 744; quoted by Smith on p. 120)

Smith suggests that this is decisive, according to the following reasoning:

> It is a commonplace, a fact of ordinary moral experience, that practical irrationalities of various kinds - various sorts of 'depression' as Stocker calls them - can leave someone's evaluative outlook intact while removing their motivations altogether. The anti-Humeans' claim that moral judgements are expressions of besires is incompatible with this. The Humeans' claim that they are expressions of beliefs is not. The anti-Humeans' view must therefore be rejected in favour of the Humeans'. (pp. 120-1)

Smith's argument is not conclusive. The besire theorist can deny that the cases of depression should be explained in the same way as cases of motivated action, and claim that it is only the latter that are to be explained by means of a besire. However, as Smith points out, whilst this manoeuvre is "technically available" (p. 124) to McDowell, it does not provide good support for anti-Humeanism. Certainly, it is theoretically possible to claim that whenever an agent acts this is explicable not by a belief and a desire but by a besire, but whilst there might be independent grounds for attributing the belief and the desire to the agent (independent, that is, from the fact that the agent acts), it is quite unclear how grounds for attributing to her the hybrid state of besire could differ from these. McDowell offers no guidance on this matter. Indeed if a besire is just *defined* in terms of belief and desire, no such guidance *could* be

given. Of course McDowell's grounds might indeed be purely theoretical, as besires promise to deliver a way of defending cognitivism, but as Smith also offers a cognitivist position, this should not by itself persuade us. McDowell may have other theoretical benefits in mind. But whatever they are, they must outweigh the immediate explanatory benefits of the belief plus desire story of depressions over the besire story. McDowell has offered us precious little in this respect.

In sum, if McDowell's anti-Humeanism depends upon besires, and we lack distinct criteria of application and theoretical reasons to prefer such states, we should not be persuaded to adopt it. In contrast Humeanism is in good shape.

In the next two sections I want to describe how habits could do better than besires in presenting the anti-Humean case. In particular, I want to suggest that habits allow us to reply to Smith's objections to the besire theory.

4. Can Habits do the Job?

Recall that according to the anti-Humean position sketched above, besires were worth considering because they were items with both directions of fit. But they turned out to be indefensible as psychological entities because there were significant explanatory costs, and no clear explanatory benefits. But besires may not be the only states which straddle the direction of fit distinction, and other such states may not be subject to these objections. A habit may be just such a state.[11]

There is an initial difficulty with this suggestion. In Chapter 5, §4, I argued that habits, whilst they are states of the agent, are not psychological states. In particular, I denied that habits can have propositional contents, as beliefs and desires do. So since Smith's characterisation of the direction of fit distinction depends upon

---

[11] There may be others. For instance, Ruth Millikan's (1995) idea of a "Pushmi-pullyu representation" (PPR), so called because they face both ways at once. According to Millikan, "PPRs have both a descriptive and a directive function, yet they are not equivalent to the mere conjunction of a pure descriptive representation and a pure directive one but are more primitive than either" (p. 145). I here merely note that insofar as this idea is one of a "representation", it may suffer from intellectualism, from which the idea of a habit is free.

the states in question having propositional content, it is unclear how we can even understand the question of whether habits have one or the other directions of fit. We can't, for instance, ask whether a habit "tends to go out of existence in the presence of a perception that not *p*", or whether it "tends to endure, disposing the subject in that state to bring it about that *p*" (p. 115).

There seem to be two ways to go. The first possibility is to adopt Smith's characterisation of the distinction as it stands. Here it seems we have a state of the agent with *neither* direction of fit. Of course this will not worry Smith if the state is not psychological, but if habits have a role to play in the agent's motivation, as I think they do, this should worry him. The existence of such items could be significant in undermining Smith's assumption that the distinction is exhaustive. It must be admitted, however, that a state with neither direction of fit would not help the anti-Humean establish the kind of cognitivism expressed in Smith's proposition (1).

The second possibility is to be more flexible in our interpretation of the direction of fit distinction, and not restrict it to psychological states. Smith's understanding of the distinction is contentious, since it is not only habits that lack propositional content, but some psychological states lack it too. Consider, for example, a state of dizziness, a psychological state if ever there was one, but it does not have propositional content. I could of course be said to believe that I am dizzy, but my being dizzy does not entail my believing it. I could have been dizzy before I came to believe it. What is more, if asked, I could describe my state in many different ways. To think that a belief with a determinate content can always be attributed to an agent in such a state, would be to credit them with more reflective thought than is justified. The same thing could be said of many other psychological states: my experiences of the objects around me; upwellings of emotion; feelings of dizziness. It is a prejudice to say that these states come over us, as it were, already in propositional form. Though that is not of course to say that such states could not be expressed propositionally given a suitable requirement to do so.

If we drop the propositional requirement on the idea of a state of the agent, then, we can give some understanding of how habits could be said to have *both* directions of fit, rather than none. What could it mean to say that "a habit aims at

fitting the world"? I have two suggestions. The first is that habits, like beliefs, are a way of the agent being sensitive to the way the world is. A perceptual belief, for example, is gained by the agent entering into some epistemic relation with what is presented to her. Similarly, a habit is exercised when the agent is in some epistemic relation with normal circumstances. There is also an analogous idea of these states being "mistaken". The perceptual belief is mistaken insofar as the facts are not as the agent takes them to be. Similarly, the habitual action is mistaken insofar as the circumstances are not in fact normal, as the agent takes them to be. In this sense, then, a habit aims at fitting the world.

The second suggestion of how habits could be said to aim at fitting the world is that habits must "mesh" with the agent's usual environment. This is the idea that behaviours which do not fit in with the world I tend to come into contact with will not be repeated, so any habit of performing such things will tend to be lost. For instance, when I worked in an office, I had the habit of putting on a suit and tie every weekday. Now I no longer work in an office, I have lost that habit. My habit is in that sense responsive to the world around me, and in being so responsive, it can be said to share the direction of fit of belief.

What could it mean to say that "a habit aims at the world fitting it"? This can be understood by drawing a parallel with desires of a non-propositional kind. As we have seen, it is often said of desires that they dispose the agent to act in some way. Indeed Smith himself adopts a "dispositional" account of desire, according to which desires dispose the agent to bring about some definite outcome (pp. 104-11). It can analogously be said of habits that they dispose the agent to act in certain ways, though no particular outcome need be sought. Just as desires need the right conditions to obtain before they are indulged, so habits need the right conditions to obtain before they are exercised. When an agent is presented with a situation - normal circumstances - she engages herself in the world in a way which is, in virtue of her having that habit, automatic for her. It is in this sense that habits share the direction of fit of desire.

I don't want to suggest that much hangs on habits having both directions of fit. If I am right that they can be understood to have either neither or both directions of

fit, the point is to undermine the claim that the distinction is exhaustive, and so compromise its usefulness as a way of classifying states of agents. If the distinction is not useful, it hardly matters whether habits lie on neither side, or both. But for the argument against Smith, it is helpful to see that whichever way we go, habits show his interpretation of the distinction to be suspect.

The next question is, how do habits furnish replies to Smith's objections to the besire theory?

5. *Replying with Habits*

Recall that Smith makes two objections to the besire theory: that besires were incompatible with Stocker's depressions and that besires are merely a technical device, lacking distinctive criteria of application. I shall first show how habits furnish replies to these objections. I tackle them in reverse order.

Smith complains that there are no clear criteria of application for besires which are distinct from those for applying a belief plus a desire, so the idea is merely a technical device to save a theory. Unlike besires, however, habits do have clear and distinctive criteria of application. Habits are attributed on the basis of the agent's history of doing the same kind of thing in the same kind of context.

What is more, there is a difference in the sort of authority associated with besires when compared with habits. The difference between psychological states and habits in terms of an asymmetry in authority is something I described in Chapter 5, §4. The point here is that if beliefs and desires are not going to turn out to be purely theoretical in the way that besires were charged with being purely theoretical, then even if we do not expect an agent to self-ascribe such states, we must acknowledge some authority on the part of the agent to at least *deny* their presence. We will usually withdraw the attribution of a belief or a desire that the agent denies (after taking into account the possibility of insincerity or self-deception). Habits in contrast are, as I described in Chapter 5, second-person authoritative. So protestations from the agent of whatever kind will not affect the truth of the attribution to her of a given habit.

Positing a habit as an explanatory item is not, then, a dogmatic technical stipulation conceived only to rescue an otherwise doomed theory, but, as I have

argued, a reasonable and familiar explanatory strategy which locates the action in the agent's behavioural history. In this respect, then, appeal to habit is a clear advance on appeal to besires in making the anti-Humean case.

Recall now the first part of Smith's argument against the besire theory. Smith argues that whereas beliefs and desires can make sense of one of the facts of "ordinary moral experience", namely, Stocker's "depressions", besires cannot. I want to argue that habits are not vulnerable to this charge.

Smith claims that it is "a fact of ordinary moral experience" that Stocker's depressions "can leave someone's evaluative outlook intact while removing their motivations altogether" (pp. 120-1). Now we can acknowledge that in the course of our everyday lives we from time to time come across people who suffer from the kinds of complaints Stocker lists. Indeed most of us would admit to having first-hand experience of a number of these. If these are "facts of ordinary moral experience", we can probably all agree that any account of moral action must at least be consistent with the existence of these. However, that is not to say that the kind of motivation (or lack of it) that is going on in the depressive cases is in any sense "ordinary".

The idea of a habit provides us with a clear understanding of what ordinary moral motivation is like. When one acts both morally and habitually, one's motivation is ordinary. Of course we have a term for such actions from the Aristotelian tradition, *virtuous actions*, which we first met in Chapter 2, and I shall say more about them later on. Let us for now accept that at least some ordinary moral actions are habitual, and restrict our comments to those.

Thinking in this way gives us an account of moral motivation. Habits, as we have seen, have what I have called "normal circumstances", in which they are usually exercised. When this happens we can explain the action in terms of the habit, and, if the argument of Chapter 6 is right, provided the habit is rational, we need look no further - to beliefs, desires, or to any other psychological item for that matter - to understand how the action is rational. If we need reassurance on this matter at all, we should look instead to the action's lack of irrationality, and to it's reasonableness. This is the normative aspect of motivation.

What is more, if the suggestion of Chapter 6, §4 about how the etiology of habitual actions could be developed, is right, it follows that in saying that an action is habitual we have thereby specified what causes it, namely, the agent's having the habit, which is "triggered" by her encounter with the normal circumstances. If we understand ordinary moral motivation like this, we can see why there is no explanatory need for beliefs or desires at all.

But do habits help us to understand Stocker's depressions? The short answer is that they don't. But this is not a problem because it is a mistake to think they should. It is of course true that an agent with a given habit will not always exercise it in normal circumstances. What is doubtful is whether, when she does not exercise the habit, the habit must feature in the explanation of her (in)action. After all, in such cases, the agent does not act habitually, but does something else, or perhaps nothing at all. In such cases we look not to the habit to explain what she does; we seek a special explanation by looking for other factors. Indeed, it is the fact that this is an exceptional case - that the agent did not do what we expected her to do, that is, act habitually - that motivates our search for an explanation. That there is a way of acting habitually in this situation is a *presupposition* of this special explanatory enterprise, and therefore the habit has no place *within* the resulting explanation.

If we see ordinary moral motivation as habitual, then, there is no incompatibility between this and the claim that other factors might, on occasion, be needed to explain depressions. There is no *a priori* constraint on the kinds of special explanation which might be appropriate for any given instance of depression. So there is nothing to rule out the possibility that a given failure of a habitual action might best be explained in terms of a lack of "will, interest, desire or strength", whilst leaving cognition, where "one sees all the good to be won or saved", intact.[12] So whilst the besire theory may be incompatible with such an analysis - if, that is, "moral judgements" (in the sense of Smith's proposition (1)) are thought always to entail a besire - if we employ the idea of habit instead, we can see why there need be no such incompatibility. For the one thing that we will not expect to feature in the special

[12] The phrases in quotation marks are from the Stocker (1979, p. 744) passage quoted in §3 above.

explanation is the habit which is not now being exercised. What is more, if we employ habits in place of besires, it becomes questionable why we should accept that *any* "moral judgements" entail the presence of such items. Moral judgements have propositional form, and as we saw above, any "cognition" in habitual action will not be propositional. Hence the supposition that the agent who has the belief that it is right that she ϕs, must also be acting habitually, actually needs support.[13]

We can go further and subject Smith to a version of his own criticism. For if we substitute habits for besires, we can see that in Smith's attempt to refute the besire theory, he falls foul of his own criticism of that thesis, which is that in adopting it we lose analytical resources. Smith's attempted refutation depends on the assumption that understanding depressions essentially involves our being able to distinguish beliefs from desires, which then allows him to reason that desires are missing whilst beliefs are in place. Only by showing this to be true can he hope to refute the besire theory which does not allow such an analysis. But as a claim about the general form of an analysis of depressions, Smith's assumption is contentious.

Just as there was nothing to unify the plurality of explanations of irrational behaviours described in Chapter 6, §5 (some of which are on Stocker's list), there need be nothing to unify the different ways in which habits can be said to fail. So, to take one of Stocker's depressions as an example, when an agent is suffering from "inability to concentrate", it would seem highly misleading to explain their lack of motivation as mere lack of desire, with beliefs intact. If the inability to concentrate is to be understood in terms of belief and desire at all, it is more plausibly understood as a failure not of desire, but of belief. Smith's preferred form of analysis actually prohibits this understanding.

But even this analysis is not mandatory. For a failure to concentrate would seem to be an instance of the irrationality of inattentiveness, which was also described in Chapter 6, §5. I characterised that idea without reference to the notion of belief, and for good reason. For we can admit that inattentiveness is an epistemic failing, without conceding that this must be understood as a failure to acquire some

---

[13] I shall say more about the effects of habits on the thesis of moral cognitivism in §8.

propositional attitude. To insist on that would be another example of intellectualist prejudice. But Smith's preferred form of analysis obscures this difficulty.

The above merely outlines how we might explain one of Stocker's depressions. But it also clearly demonstrates that Smith's insistence on the belief-desire model, with only beliefs in place, to capture the diversity of complex states that play a role in understanding such maladies, grossly oversimplifies them.

Smith's defence of Humeanism might have succeeded against McDowell interpreted as a besire theorist. But this defence is ineffective against the anti-Humean who instead employs the idea of a habit.

Where does all this leave Smith's moral problem? Assuming that habits are a way of articulating a viable anti-Humeanism, proposition (3), for habitual moral actions at least, can now be rejected. Arguably also, some progress has been made towards rejecting the intellectualist way in which Smith characterises cognitivism in proposition (1).

In addition, I think that there is some promise for understanding proposition (2), internalism, non-propositionally, and in terms of normal circumstances, though I have not mentioned this so far. The idea is that a habitual action does not involve the agent in a "judgement that it is right that she ϕs", which, insofar as she is "fully rational", in turn generates her desire to ϕ, and, finally, her action. A much simpler story can be told. We can say that given the agent has a habit of ϕ-ing in C, and that she has encountered such circumstances, *ceteris paribus*, she automatically ϕs. *Ceteris* is *paribus* not when she is "fully rational", but when she exhibits no irrationality, and does not decide to opt out of the habit on that occasion. We need not spuriously attribute a judgement that it is right to ϕ to the agent, which only displays an intellectualist prejudice. And we can drop the similarly intellectualised character, the "fully rational agent". But we can still, I think, do justice to the central internalist thought, that coming into contact with certain things just does motivate us.

I cannot claim to have completely dissolved the moral problem, because I have not given an argument for the claim that *all* moral actions are habitual. I merely note that by drawing on Aristotelian materials, some of which were sketched in Chapter 2,

§3, there may be a case for thinking of them as the paradigms of moral action. Indeed one may be able to argue, in the spirit of Aristotle, that if the question occurs to one of whether to, say, act courageously, even if one still does it, one's action nevertheless falls short of being fully moral. But in the absence of such an argument more work needs to be done to show that Humeanism is false for non-habitual moral actions (if there be such) as well. If some moral actions are habitual, the moral problem does not arise in explaining those actions, and there is accordingly no longer any need to wheel in Smith's complex theoretical apparatus.

6. Blackburn's Expressivism

Blackburn's overall position, which he has developed over a couple of decades (Blackburn 1984a, 1994, 1998) is a version of *moral non-cognitivism.* Blackburn believes that moral facts are not genuine features of the world, according to some standard of genuineness, and in this respect it contrasts with both Smith's and McDowell's views which both purport to be cognitivist. According to Blackburn, when we say things like "you should not kill innocent people", whilst we appear to be saying something that importantly reflects the way the world is, we are in fact doing nothing of the sort. What we are in fact doing is "expressing" our moral emotions towards the world, which includes innocents and potential killers. Because of the centrality of this notion of expression, Blackburn calls his position *expressivism.*[14]

Blackburn denies that expressivism leads to Mackie's (1977) conclusion that talk of moral facts is an error (Blackburn 1998, p. 301; 1984a, p. 180). He offers a theory of how our expressions of emotions come to have propositional content, which Blackburn calls *quasi-realism.* Roughly the idea is this. We all have moral attitudes towards things, many of which we tend to share. We also have attitudes - such as admiration or disapproval - towards *each other's* attitudes. Hence the attitudes

---

[14] Although in his early statements of the position (e.g. 1984a), Blackburn used the term "projectivism" for this component of his position, in recent work he has dropped this term. The term "projectivism", he writes, "can make it sound as if projecting attitudes involves some kind of mistake, like projecting our emotions onto the weather, or projecting our wishes onto the world by believing things we want to believe." He adds, "This is emphatically not what is intended" (1998, p. 77).

themselves become subject to our rational criticism. This in turn means that we end up with moral talk which is couched in terms of moral "beliefs", which are capable of "truth" and "falsity", in virtue of representing the moral "facts". And because Blackburn thinks we can be "minimalists" about truth, this talk about facts does not fall short of the real thing. Thus, Blackburn claims, quasi-realism can give us everything the cognitivist could want, without actually *being* cognitivist. This represents a distinctive and challenging position in meta-ethics, and one with which any aspiring moral cognitivist must engage.

There are problems with Blackburn's position, not least of which is the tension apparent in saying that it is very like cognitivism without really being cognitivism.[15] A preferable position would at least avoid such talk, if it didn't simply accept moral truths at face value. Another difficulty is that the only genuine cognition that Blackburn allows involves an idea of reality stripped of all influence from the human mind, which science will ultimately reveal. But this idea of mind-independent reality is notoriously difficult to defend. And whilst Blackburn's project is nothing like as theoretical as Smith's, the quasi-realist construction may be based on a contentious picture of human nature and human practices.

Blackburn's position, like Smith's, has at its core a Humean division between reason and passion, and a Humean account of motivation. Blackburn's Humeanism differs in some details, and in particular is rather less precisely specified than Smith's proposition (3) (see e.g. Blackburn 1998, pp. 238-41). Nevertheless, like Smith, Blackburn holds that desires are the only things that can motivate agents to act, and are not the right kinds of things to disclose how the world is arranged. This Humean view does not, for Blackburn, motivate a theoretical project like Smith's to rescue cognitivism. Rather it drives Blackburn's quasi-realist construction and his denial that genuine cognition in ethics is achieved. Blackburn's position, if it works, supports the Humean theory. But equally, if the Humean theory fails, Blackburn's position is

---

[15] The stability of Blackburn's position in this respect has been challenged by Crispin Wright (1985), to which Blackburn (1993) replies. For a fuller discussion see Pollard (1998, pp. 50-53), in which, whilst expressing other reservations, I deny there is an instability.

undermined. Blackburn's repeated attempts to defend Humeanism against its critics, amongst whom McDowell has been prominent, show he recognises this vulnerability.[16]

7. *Humeanism Defended II*

On the assumption that Blackburn's most recent defence of Humeanism against McDowell's objections is his best to date, let us look at his arguments, which appear in *Ruling Passions* (Blackburn 1998, pp. 92-104).[17] This will enable us to show, in the two subsequent sections, how the notion of a habit can provide better replies to Blackburn's objections than those more readily available from McDowell.

Blackburn objects to McDowell in two ways which he summarises as follows:

> The bad things [about McDowell's account] include the unearned emphasis on 'receptivity' or the belief that some kind of cognitivism has been established, and the unfortunate hostility to the essential business of factoring out the inputs and the outputs of our evaluative practices. (p. 104)

Blackburn's first criticism is that McDowell thinks that "some kind of cognitivism has been established". This turns on the question of whether what we call "moral properties", say, the cruelty of a certain action, are genuine features of the world. Blackburn denies it; McDowell (at least as Blackburn reads him) asserts it. Blackburn contrasts the two explanatory stories as follows ("where the arrows indicate some explanatory story"):

> *Cognitivism*: Shapeless underlying class → shapely property M → perception of it by those with proper affective dispositions ≡ perception of a reason for action → action
>
> *Non-cognitivism*: Shapeless underlying class → attitudes in those with specific affective dispositions → action. (pp. 98-9)

---

[16] Criticisms are found in McDowell (1981, 1985, 1987); and replies in Blackburn (1981, 1985, 1998).

[17] In this and the next two sections, page numbers refer to this work unless otherwise stated.

Blackburn then asks "what extra explanatory weight is added by the mention of the shapely property at the moral level, and by the talk of perception?" (p. 99). He thinks that "there is no evident answer" (p. 99).

Blackburn's second criticism is of McDowell's "unfortunate hostility to the essential business of factoring out the inputs and the outputs of our evaluative practices" (p. 104). The idea of inputs and outputs is an idiom employed throughout Blackburn's book. He introduces the notion as follows:

> we can usefully compare the ethical agent to a device whose function is to take certain inputs and deliver certain outputs. The *input* to the system is a representation, for instance of an action, or a situation, or a character, of being of a certain type, as having certain properties. The *output*, we are saying, is a certain attitude, or a pressure on attitudes, or a favouring of policies, choices and actions. (p. 5)

Blackburn goes on to insist that "It is only by thus 'splitting' the input and the output that the reaction can be seen sufficiently clearly for what it is. And this is important because only then can the reaction itself be intelligently discussed, and perhaps ... seen as highly questionable" (p. 7). Blackburn develops this theme later in the book in his objection to McDowell, who he reads, as Smith does, as positing besires (p. 97).[18] Blackburn gives an example in which it is "*morally* vital" to "split" if we are to correct a pernicious chauvinistic use of the word "cute". Speaking critically of the besire theory in this regard, Blackburn writes,

> if the last word is that these people perceive cuteness and react to it with the appropriate cuteness reaction, whereas other people do not, we have lost the analytic tools with which to recognize what is wrong with them. (p. 101)

So for Blackburn, it counts against an explanatory theory if the materials required for moral criticism are inconsistent with it. This is the case with the besire theory. We want to be able to say what is wrong with the chauvinists who call women "cute",

---

[18] Like Smith, Blackburn finds besires in McDowell, but in McDowell (1981), rather than McDowell (1978). I take up this exegetical issue in §10.

and we want to be able to do so by pointing both to their mistaken attitudes as well as to various facts about women that they seem to have got wrong. Blackburn writes:

> What is wrong with them is along these lines: they react to an infantile and unthreatening appearance or self-presentation in women, or overt indications of willingness to be subservient to men, with admiration or desire (the men) or envy and emulation (the women). Cute things are those to which we can show affection without threat, or patronizingly, or even with contempt. Children and pets are quintessentially cute. Applied to women, this, I say, is a bad thing. (p. 101)

Hence Blackburn thinks that McDowell's purported failure, disguises "a conservative and ultimately self-serving complacency" (p. 102). According to Blackburn, then, positing besires is not just a theoretical problem, but also a moral failing.

This criticism is an advance on Smith's criticism of the besire theory discussed in §3 above, which was based on Stocker's depressions. It's one thing to say what is wrong with somebody who is not motivated: the depressives. But it's quite another to say what is wrong with somebody who *is* motivated, but in the wrong way: Blackburn's chauvinists. If a theory can satisfactorily deal with the first disorder, it doesn't follow that it can deal with the second.

Let us concede that Blackburn's objections may be fair given McDowell's rather limited articulation of anti-Humeanism. Still Blackburn's objections may be insufficient to defend Humeanism. In the next two sections I shall show how habits can provide replies to both of Blackburn's objections.

8. Cognitivism vs. Non-Cognitivism: A Way Out

Blackburn's objection to cognitivism depends upon construing cognitivism as entailing the idea that we must always be able to locate a moral property to which every moral action can be seen as a response. This may be a reasonable construal and one with which we need not quarrel here. However, the idea of a habit shows how cognitivism thus understood is more than we need to oppose non-cognitivism. For I think we can oppose non-cognitivism without making any claims about moral properties.

Recall from Chapter 5, §9, that the notion of a habit that I have developed allows habits to be acquired without any intellectual awareness of the kinds of circumstances that the corresponding habitual actions are a response to. That is to say, an agent might have the habit of ϕ-ing in C without being able to describe C or even ϕ. She need not have the conceptual or otherwise intellectual resources to disclose these things. This means that if we are explaining such a habitual action it will be a mistake to include the property under which C might fall, as if the agent herself had "recognised" that property in this intellectualised sense. So we get an explanatory picture that looks like this:

(HE) Agent with the habit of ϕ-ing in C + encounter with normal circumstances C $\rightarrow$ habitual action ϕ.

The first thing to note is that there is no mention either of moral properties, or of a "shapeless underlying class". The only "worldly" items in explanatory scheme (HE) are normal circumstances C. However, because there is no commitment to any particular description of C, we can be neutral on the question of whether (HE) represents cognitivism or not. We might specify C in terms of moral properties such as cuteness, so one of Blackburn's chauvinists can be said to have the habit of responding to cute things by calling them "cute". Or we might specify C in much more neutral terms, so the chauvinist has the habit of calling women who present an unthreatening appearance "cute". Whether this latter specification is "shapeless" I am not sure, but presumably we could specify, albeit in rather broad terms, what these men are reacting to without mentioning the objectionable moral properties, and hence not be committed to the cognitivism of which Blackburn is critical.

But to call explanatory schema (HE) *non*-cognitivist would be equally misleading. For unlike Blackburn's own schema, there is no mention of a non-cognitive attitude of the agent. Whilst we do not need to deny that the agent has such attitudes, they need not feature in an adequate explanation of her action either to show it to be rational, or to show how the agent is motivated. Habits can do that on their own. We saw in §3 how habits can be construed as sharing the direction of fit of beliefs, and hence equip the agent to respond to genuine features of the world. As we are not committed to any particular description of C, we can adopt whatever

description satisfies our craving for genuineness. And we can do this without implying that the agent is aware of the world under that description when she acts, though we might be able to convince her of the appropriateness of such a description in the future, if this were important to us.

What the notion of a habit gives us is a clear distinction between responding to features of the world, and specifying what those features are. The cognitivist and non-cognitivist alike are hung up on committing themselves to just one description of the world to which the agent can be said to be responding: respectively, the world as it appears in the agent's reasons; and the theoretical idea of mind-independent reality. But if we accept that moral actions are habitual, we need not commit ourselves either way. We can reject the intellectualist assumption that all actions are actions for reasons, without committing ourselves to the dubious notion of a mind-independent reality. When an agent acts habitually, she can be understood as responding rationally to the world. The temptation to add to that description may well arise from a flawed philosophy of action.

Understanding moral actions as habitual, then, allows us to side-step the debate between cognitivists and non-cognitivists, at least as these positions have been conceived so far. However, a version of cognitivism is arguably still available in which habits, being rational responses to the world, play a central role. The cognitivism that is available is not conceived propositionally, as in Smith's construal, or in terms of moral properties, as in Blackburn's, but in terms of the agent being disposed to respond to the world in certain ways, one such disposition being a habit. Viewed in this way, the traditional debate between cognitivists and non-cognitivists is a result of a shared, intellectualised conception of rational actions, which we can now avoid.

If this is right, there is no need for the anti-non-cognitivist to supply an account of the "extra explanatory weight" provided by talk of moral properties, as Blackburn claims there is. For the anti-non-cognitivist can be ambivalent about the

role of such talk in action explanations anyway.[19] However, as we shall soon see, it would be a mistake for him to deny that such talk is ever useful.

9. *Splitting and Habits*

We turn now to Blackburn's second objection, namely, the moral failing implicit in McDowell's refusal to "split" inputs from outputs. As I have said, I think that this objection might indeed be telling against an anti-Humean case built around the notion of a besire. But again the idea of a habit improves the anti-Humean case. I want to show that we can reply to Blackburn's second objection by making this substitution. Indeed, I want to go further, and suggest that by making this substitution, we can turn Blackburn's criticism against him. For, in a move similar to that made against Smith in §5, it is the Humean picture, and not the alternative, that is left wanting when it comes to the criticism of moral failures.

Blackburn's criticism of the besire theory is that we lose materials which are essential to the proper criticism of morally pernicious behaviour. However, if the anti-Humean employs a habit rather than a besire, it is not clear that any critical materials are lost. When an action is explained by a besire, this involves the denial of separable belief and desire components which may be of use in later criticism. However, when an agent acts habitually, however bad her habit may be, we can consistently maintain that her action is best explained by a habit without denying the existence of any beliefs and desires which might be helpful in criticising what she did.

To say this we must maintain a distinction between the best explanation of an action on the one hand, and the best materials with which to criticise that action on the other. This should not be problematic. Explaining an action and assessing it morally are different kinds of activity, and it would be a coincidence if the two happened to employ the same concepts. Blackburn's own example of the chauvinistic use of the term "cute" illustrates this. The actions are criticised not on the basis of beliefs and desires, which are Blackburn's preferred explanatory materials, but on the basis of how the chauvinists react, what they associate women with, and facts about

---

[19] Though this is not to say that such weight cannot be supplied. See, for instance Wiggins (1996).

children and pets. It seems highly implausible that any or all of these materials will feature in Blackburn's best explanation of the action. But even if this could be argued, it remains the case that we need not criticise an action on the basis of its best explanation, and nor need we explain an action in the terms that are best going to get across what is morally wrong with it. Of course, as Blackburn recognises, the two cannot be inconsistent, but other than that, there need be no *a priori* limit on their respective contents.

The idea of a habit also suggests how we can turn Blackburn's criticism against his own account. For if the Humean looks to his explanatory theory for his only source of critical materials, as Blackburn seems to, and thereby restricts himself to attitudes and attitude-independent facts, he will miss another kind of critical material that is made available through the notion of a habit. For whilst we might want to say that what is objectionable about the chauvinists application of "cute" to women is the attitude it expresses, we might also, and more importantly, want to say something else. Namely, that what is objectionable is that they don't just do this once, which would be bad enough, but they do it repeatedly, automatically and responsibly. In other words, they do it habitually. An occasional bad action is bad enough; a bad habit is much worse.

If the role of moral criticism is to say what is wrong with such actions with a view to correction, criticising a bad action as habitual will be importantly different from merely pointing out how disgusting we find the attitude expressed by that action. In fact, if we are just criticising the attitude expressed in a *single* action, it becomes puzzling why our concern is so great, and what we would hope to achieve by such criticism. For as yet we have no reason to think that the agent will ever express the same attitude again. In contrast if we criticise somebody's bad action whilst acknowledging that it is habitual, the likelihood of future transgressions makes criticism more urgent.

Not only that, but treating this kind of action as habitual suggests various corrective strategies which would not be visible were the materials for criticism restricted to attitude and fact, as Blackburn's account assumes. A first stage might be to make the agent aware of their habit, which might be enough on its own suffice to

motivate corrective strategies to lose the habit.[20] Harder cases might involve other kinds of criticism, such as reminders about what women are like, and indeed recommendations of more exposure to them, and rather less to groups of men in their locker rooms. Again change in lifestyle is what is suggested, not merely a change of mind. Thus any account of critical resources which leads to such suggestions would seem to be preferable to one that limits those resources to mere attitude and fact. The Humean who looks to action-explanatory concepts as the only source of critical resources is in danger of imposing just such a limit. Insofar as Blackburn does so, he is subject to a version of his own criticism intended for the besire theory.

This point is relevant to the issue of moral properties raised above (§7). For a similar criticism can be made of Blackburn's refusal to acknowledge the legitimacy of moral property talk. For this may be a helpful critical resource in some cases, as in, for instance, "it's wrong to call women 'cute' because they are not cute". Just who the criticism is directed at will affect the terms in which we put it, but it would be a mistake to rule out property talk *a priori*, as Blackburn does. And we can say this without making any claim about a specifically explanatory role for moral properties.

In sum, the notion of a habit can be used to make a stronger case against Blackburn's Humeanism than is readily available in McDowell's writing. This places Blackburn's version of moral non-cognitivism in doubt, though it may do nothing to improve the prospects for moral cognitivism if it is understood as essentially involving moral properties. This way of responding to Blackburn's defence of Humeanism, then, is a second use for habits in meta-ethics.

10. But is this McDowell?

I have suggested above that McDowell's rejection of Humeanism would fail to survive Humean replies if he holds a besire theory, which is how the Humeans read him. It also follows from what I have argued that if McDowell had exploited the notion of a habit defended in this thesis, my arguments would be not a corrective to McDowell,

---

[20] A brief account of habit loss can be found in Chapter 3, §8.

but rather a fuller articulation of he says. I will now discuss briefly which of these two views is the right reading of McDowell.

First, the evidence Smith cites is from McDowell's "Are Moral Requirements Hypothetical Imperatives?" (1978). The only portion of this paper that Smith actually draws on is the following passage:

> A view of how things are is a state or disposition of one's cognitive equipment. But the psychological states we are considering are to suffice, on their own, to show how certain actions appeared in a favourable light. That requires that their possession entails a disposition of the possessor's will. (McDowell 1978, p. 82; partially quoted by Smith 1994, p. 121)

Here there is clear evidence that McDowell is putting forward the idea of a "psychological state", and, for reasons that I have given earlier, this would seem to conclusively rule out the possibility that he is referring to a habit. We get more information about the kind of psychological state later on in that paper.

"What is questionable", writes McDowell, "is whether there need *always* be an independently intelligible desire to whose fulfilment a virtuous action, if rational at all, can be seen as conducive" (p. 84). The suggestion is that virtuous agents have "special way of conceiving situations" (p. 84), and that how they conceive a given situation might at times be enough to explain their motivation for the resulting virtuous act. In such cases, whilst we may be able to attribute a desire to such an agent, that, argues McDowell, would merely be a reflection of how she conceives of the situation, and is hence ascribed "in a purely consequential way", which is contrasted with being "independently intelligible" (p. 84).[21] This then is where, I take it, Smith finds the idea of a besire. We have the idea of a cognitive state of the agent, namely, how she conceives the situation, and the possibility that a desire might be ascribed only as a logical consequence of her being in this state, rather than that desire being "independently intelligible", as it would have to be were it a Humean "distinct existence". Combined with the claim that we are here talking about psychological

---

[21] This is a contrast McDowell attributes to Nagel (1970, pp. 29-30).

states, this would seem to give excellent support to Smith's claim that McDowell is in this paper positing besires.

Blackburn does not find the besire thesis in the same McDowell paper. He instead cites McDowell's "Non-Cognitivism and Rule-Following" (1981):

> The suggestion involves thinking of exercises of our affective or conative natures as themselves in some way percipient, or at least as expanding our sensitivity to how things are; and the eighteenth-century philosophy of and would purport to exclude this *a priori* (McDowell 1981, p. 200; partially quoted by Blackburn 1998, p. 97).

Now it is far from clear what McDowell means by "our affective or conative natures", but in view of the fact that McDowell is here trying to loosen our grip on "eighteenth-century philosophy of mind" - i.e. Humeanism - I think it is again reasonable for Blackburn to suppose that he means some state of desire which is at the same time cognitive. Certainly, McDowell's suggestion that such a state would be ruled out *a priori* by such a philosophy of mind would seem to be good evidence that he wants to leave room for nothing less than a besire.

However, whilst I think that these early writings of McDowell give good support for the besire reading, I think that in later work we find more evidence that this may be a mistake. In "Might there be External Reasons?" (1995b), for instance, McDowell distances himself from the view that possession of psychological states is all there are to being morally motivated. In his *Mind & World* (1996) McDowell emphasises what he calls *Bildung*, that is, the process of acquiring rational (and moral) natures through upbringing into a tradition. And in "Two Sorts of Naturalism" (1995a) McDowell makes a clear commitment to the Aristotelian idea that virtue is "second nature" to moral agents. McDowell writes,

> The practical intellect's coming to be as it ought to be is the acquisition of a second nature, involving the moulding of motivational and evaluative propensities. (1995a, p. 185)

I think this can be read as entailing the suggestion that trained-up moral agents exercise the virtues (as well as other rational habits) *automatically*. Not only that, but

there is also evidence that for McDowell that the exercise of second nature is also under the agent's rational control, through the capacity to "step back":

> moral education enables one to step back from any motivational impulse one finds oneself subject to, and question its rational credentials. (1995a, p. 188)

This would correspond to the capacity to opt out of a given habitual action, and act otherwise for reasons. These papers suggest that McDowell is less committed to the idea of a psychological state like a besire, and more committed to that of a rational habit, of which the moral virtues would be instances, as the crucial ingredient in his anti-Humean philosophy of mind.

Our conclusions must be tentative. McDowell leaves the question of whether he posits besires or habits rather open. Going by some of the earlier portions of McDowell's writings, the Humeans seem justified in reading him as positing besires. But other, later, textual evidence suggests that this reading is unjust, or that McDowell has revised his view. For in later work the idea of a virtue, as a kind of habit, plays a central role for McDowell.

If McDowell leaves this matter open, he should not have done. Doing so only invites spurious attacks by the Humeans, giving their position more credibility than it deserves. McDowell might have done better to distance himself from those early remarks, and oppose Humeanism in the manner developed here, by giving habit a central role in an anti-Humean philosophy of action. In clarifying the views of this important, though sometimes rather inscrutable, writer, habits contribute to current meta-ethical debate in a third way.

11. Conclusion

In this chapter I have outlined how the idea of a habit might contribute to contemporary debate in meta-ethics. So long as Smith and Blackburn remain influential proponents of Humean philosophy of action, this is a significant result. For without an understanding of habits they may with some justification think that Humeanism survives the attacks on it led by McDowell. However, I think that Humeanism remains vulnerable. Equipped with the idea of a habit, the anti-Humean

can mount new attacks on their positions. This opens up new avenues not only for accounts of moral motivation, but also to moral internalism and cognitivism. It also provides us with a fresh understanding of McDowell's work.

8

CONCLUSIONS AND FUTURE WORK

> For we can avoid ineptness or emptiness in our assertions only by presenting the model as what it is, as an object of comparison - as, so to speak, a measuring-rod; not as a preconceived idea to which reality *must* correspond. (The dogmatism into which we fall so easily in doing philosophy).
>
> Wittgenstein, *Philosophical Investigations*, §131

1. The Role of Habits

In this thesis I argue that habits should play a much more significant role in analytic philosophy of action than they currently do. This is because, just like Aristotle, Ryle and Wittgenstein before us, we are once again up against intellectualism, and habits can help us, like they helped them. Contemporary intellectualism is there in our very understanding of what an action is. The mistake is thinking that our actions are better thought out than they in fact are. As a result, our paradigms of action are those involving deliberation, and we think we always have a reason when we act.

But if we turn our attention to the actions which make up the majority of our active lives, in particular, to the things we do habitually in our most familiar surroundings, the mistake becomes clear. Deliberation and reasons do not play a role in such actions. So long as we are gripped by the intellectualist mistake, if we notice habitual actions at all, we will be liable to distort the true nature of such actions by fitting them into the intellectualist mould, to which they are ill-suited. We have, in Wittgenstein's terms, "a preconceived idea to which [the] reality [of action] *must* correspond" (1953, §131). Whether we ignore habitual actions, or distort them, we lose touch with the phenomenon of ordinary action that is recognisable to philosophers and non-philosophers alike.

This thesis has wider implications within philosophy too. Not least of these are the exegetical possibilities opened up, in particular for the three writers mentioned earlier. Aristotle's idea of action can now be examined in a new light.[1] Ryle's idea of a behavioural disposition can now be compared with the notion of a habit developed here. And the implications for interpreting Wittgenstein as offering habits as an account of rule following can be followed up, and contrasted with the views of commentators such as Winch (1958) and Pettit (1993).

Hume might be thought to have had a rough deal. But if he has, that is only insofar as he is a Humean. I think he may have much to contribute to our understanding of habits, though it is presented quite separately from his thoughts on motivation. For Hume was well aware that "Custom or Habit" is essential to understanding how we can make inferences from experience, "without any process of reasoning" (1777, pp. 42-3). He writes:

> For wherever the repetition of any particular act or operation produces a propensity to renew the same act or operation, without being impelled by any reasoning or process of the understanding, we always say, that this propensity is the effect of *Custom* (Hume 1777, p. 43).

Perhaps Hume's writing could be illuminated by a proper understanding of the idea of habit presented here. And equally perhaps there is much to learn from Hume on its applications.

In the rest of this final chapter I want to bring some themes together from the thesis, and indicate directions in which they might be developed. In §2 I say where I think this thesis leaves accounts of action. In §3 I describe the emerging importance of the role of normal cases. In §4 I suggest how, as a result, we might develop accounts of rational action. In §5 I draw out some implications for moral motivation. And finally in §6 I deliver on the promise made in Chapter 1, of saying how the idea of habitual action can enhance a naturalism of second nature.

---

[1] A recent paper by Mikael Karlsson (2002) makes a start in this direction.

This is not to deny that there may be a class of actions which can be understood as being done for some purpose, or as, in some sense, as goal-seeking, as has been proposed by Collins (1987, Ch. 6) and Stout (1996). But we can now see some of the challenges which such teleological views will have to meet. For if such theories are to apply to all actions, they had better not determine the purpose with reference to the agent's reasons (or they had better conceive of reasons rather differently to the ways considered here). And in the light of the idea of habitual action, I rather doubt we could say that all expressions of agency aim at something without strain. But perhaps the greatest challenge presented here is that we now have reason to doubt that conceiving of actions teleologically is the only, or the best, way of making sense of their normative character. We should act habitually not because its means and ends are justified, but because there is no justification for not doing so.

3. *The Priority of Normal Cases*

There is a theme running through this thesis that I think it would be helpful to bring out. It not only reveals links between some of the arguments, but it also suggests how a number of issues might be taken forward. I trace the theme back to some moves made outside the philosophy of action altogether.

In the philosophy of perception a view called *disjunctivism* is becoming popular. It is a kind of reply to a certain use of the argument from illusion, the classic statement of which is made by Grice (1961). That argument is intended to establish that since our subjective experiences of illusions are indistinguishable from veridical experiences, there must be some factor common to both veridical and illusory experiences. The common factor is supposed to be that of "appearance", or sense data. Such items are then supposed to mediate between us and the objects we perceive even in veridical experience.

Disjunctivists such as J. M. Hinton (1973), Paul Snowdon (1981) and McDowell (1982) resist this conclusion. They argue that the similarity of appearances are insufficient grounds for positing anything common to the two sorts of occurrence. We can instead say that *either* an experience is veridical, which does not fall short of genuine contact with its object; *or* it is a mere appearance, which of course does. This

sort of disjunction is where the response gets its name. Just because veridicality breaks down at times, we need not conclude that appearances mediate in veridical appearance.

Dancy (1995a) has noticed something of the generality of the disjunctive move. But I don't think he sees just how general the scope of its application can be. For as I read the disjunctive move, it asserts the right to a view of a phenomenon as it normally occurs, and refuses to regard aberrant cases of that phenomenon as informative on this matter. That would be as misguided as trying to understand veridical perception as an illusion that just happened to go normally. In other words, disjunctivism recognises *the priority of normal cases.*

We can understand why normal cases are prior to abnormal ones by considering how explanations work.[2] One might have a perfectly good explanation of why something behaves abnormally. But this would be a special explanation. The reason that such explanations will not be applicable to normal cases is that the very search for a special explanation *presupposes* some (perhaps inchoate) grasp of what normally happens. What we seek is an explanation of what is presupposed, and that cannot be given by an explanation of why it did *not* happen.

There is a second reason why normal cases are prior to abnormal cases, and that is because they are fundamental in the order of explanation. If the job of an explanation is to present some problematic phenomenon in terms which are less problematic, we have every reason to expect that explanations will stop when we reach truths which are both obvious and familiar. Relative to abnormal cases, normal cases are like this. Explanations stop at normal cases because there may be nothing more obvious or familiar with which could explain them. These two points, then, give something of a explanation for why disjunctivism as a strategy is promising in action theory.

I used a variation of a disjunctive argument in Chapter 3, §7. There I said that we should reject the arguments for saying that an agent always "tries" when she acts,

---

[2] I used a more specific form of this argument in opposing Smith in Chapter 7, §5.

since those arguments trade on intuitions about contexts which are abnormal, in which we would say that an agent "tries", and then persuade us to say the same thing in normal cases. But this is asking us to understand successful actions as failed actions that just happened to go well. And I rejected these as grounds for denying that habitual actions involve trying.

I also employed a kind of disjunctive argument in my rejection of Frankfurt-type cases in Chapter 4, §6. Frankfurt uses cases of alien manipulation in which agents "cannot do otherwise", yet still appear responsible for their actions, to try to persuade us that the idea of doing otherwise is quite generally irrelevant to responsibility. It is another case of inferring from features of abnormal cases to conclusions about the normal. The disjunctive move shows us that we need not draw any such conclusions. We should instead give priority to the normal cases.

In the light of the above two arguments I think there is potential here for a quite general critique of certain sorts of thought experiment. Engagement with Grice (1961) and others would undoubtedly be profitable in this respect. That is an avenue of future research. But I want to make further suggestions here about how the above kind of move might be applied to a possible difficulty with my account of rational actions.

4. Rational Actions

In Chapter 5 I rejected the orthodox view that all rational actions are actions for reasons, on the grounds that our understanding of this idea is too intellectualised to accommodate habitual actions. My solution, in Chapter 6, was to suggest two ways in which a good many such actions might be accommodated as being rational, and that is, either by adopting the permissive conception of rational action, or by understanding "rational" as meaning something closer to "reasonable". I have intended both conceptions of rational action to be inclusive, so that actions for reasons can still be regarded as rational whichever we adopt. That way, it is easy to retain the natural intuition that rational agents act rationally most of the time.

But left with this duality, the question remains of just how actions for reasons and rational habitual actions are related to one another. I now want to make a

suggestion about how this might go, by drawing on the principle of the priority of normal cases. The thought would be that habitual actions should be regarded as the normal cases of rational actions, whereas actions for reasons should be considered to be a kind of aberrant case. This reverses the orthodox view, which makes actions for reasons central. I think there is much to be said for the reversal.

Rational habitual actions deserve to be regarded as our normal expressions of agency, since they are perfectly in order when our surrounding are familiar. In such surroundings, when there is a normal thing to do, we don't ask "why?"; more likely, that question would occur to somebody else, and we could justifiably reverse the onus by replying "why *not*?" It is only when things go awry, that is, when the world presents itself to us in a new way, or perhaps when more than one natural course suggests itself that a question arises. We find ourselves unsure about what to do, and resort to deliberation before we act. Any resulting action will still, we hope, be a rational action (either not irrational or reasonable), but insofar as it is, it is a rational action of a rather more difficult and risky kind. Deliberation, reflection, decision, trying, and reasons themselves, have their home in such *ab*normal circumstances, for it is only in such circumstances that our rational habits give out.

The status of normal and abnormal is not fixed, but dynamic. If an agent comes across these same sorts of circumstances again, less thought will be required, and before long she will have added to her repertoire of habits. Previously abnormal circumstances will become normal for that agent with the new habit. Similarly, if she were to lose a habit, what would have been normal circumstances would become abnormal. A full picture of rational action will capture this interplay of habit and deliberation.

What is striking in this picture is that the role of deliberation and reasons is seen as marginal. Their role is to deal with the abnormalities as and when they arise, and to allow us to respond appropriately. But when things are normal, deliberation and reasons have no role at all to play in our actions. Viewed in this light, the reasons theory, as a general account of rational action, seems deeply misguided. The reasons theory involves trying to understand all rational actions as responses appropriate to abnormal circumstances. The picture is of a somebody pathologically unable to resist

reflecting on everything she does; somebody strangely alienated from her surroundings, rather than somebody well attuned to them.

Making habits central as I propose offers a fundamental challenge to the way the philosophy of action is currently conducted. For it may be that at the bottom of the reasons theory then lies a failure to recognise the priority of normal cases. When it comes to understanding rational actions, habitual actions could be more fundamental than actions for reasons in the order of explanation.

5. *Moral Motivators*

The priority of normal cases might also allow us to make progress on two issues that arose in Chapter 7, both of which concern how we should understand the idea of moral motivation.

First, in Chapter 7, §5, I left open the question of what proportion of moral, or virtuous, actions are habitual. If they could all be shown to be habitual, this would completely get rid of the moral problem with which Smith is concerned. But there it seemed unlikely that this could be done since it appeared plausible that deliberation should be no bar to an action's moral status. Indeed it would seem to be a moral failing if some deliberation did not take place in certain kinds of context. An example would be when the consequences of a putative action would be of great moral significance. One thinks of paradigm moral dilemmas of our age, such as abortion and euthanasia. A lack of deliberation in these contexts would be reckless.

But if the reversal suggested in §3 above is right, there is scope for investigating the following possibility. For one could still maintain that the paradigms of moral action are habitual, but it is just the hard cases that are not. In normal moral contexts, such as a situation requiring courage, the person with the right moral habits, in this case courage, just habitually acts courageously. In normal contexts, nothing further is required to show how she is motivated.

But in the hard cases, where it is unclear what virtue requires, since our current virtues either clash, or are not up to the task, we must deliberate. And here our motivations will be properly explained by the reasons we ultimately act upon. And here, of course, the moral problem might still arise. But there are now two sorts

of response available. First, the moral problem only arises in marginal cases of moral actions. So its claimed centrality is only a result of an intellectualist bias. And second, since the cases when the moral problem may still arise are hard cases, the question to press would be, does the Humean picture capture anything of the subtlety and complexity of the sorts of reasons which it would be appropriate to consider? I have argued it falls far short of this.

The second way in which the reversal might allow us to make progress on the topic of moral motivation is by engagement with some very recent writing. Jonathan Dancy (1995b) and Richard Norman (2001) have argued, in different ways, against the suggestion that we need a "theory of motivation." Such a theory is supposed to fill a putative "gap" between the agent's knowing what is good about doing the action, and her actually doing it. The Humean fills this gap in the familiar way, buy positing a belief-desire pair, which together both justify and cause the action. But Dancy and Norman claim that the "gap" which is supposed to be filled by such a theory actually presupposes the discredited Humean picture. Whilst I am sympathetic to this conclusion, I don't think they go far enough, and that is ultimately due to their intellectualism, which is itself a remaining vestige of Humeanism.

Norman gets closest to what I think should be said. He claims that "we do not need any further explanation of why, *in the normal case*, human agents act on their beliefs about what they have good reasons to do. They just do it" (2002, p. 11, my italics). He considers the possibility that this sort of explanation might be "dispositional". "This may be so", he writes, "but the 'theory of motivation' is not a plausible candidate for a further explanation at this level. I doubt whether any uncontentious view is available as to what *would* count as an appropriate explanation" (p. 11).

Norman is right about this, but he has no way of explaining why he is right. I think that is because of his residual intellectualism. What he needs to appreciate is the significance of the "normal case" of action. For if we understand such cases as those in which agents act *habitually*, we can see that reasons drop out of the picture altogether. There is no appropriate explanation in such cases, because there are no reasons that relate to the action. But that is not to say that there is no appropriate

explanation of why the agent does something. For the appropriate explanation is dispositional: "just doing it" is acting habitually. And as we have seen, the fact that this is not an explanation in terms of reasons does not mean that the action is not rational.

So I can agree with Norman (and Dancy) that insofar as the theory of motivation requires an explanation for the "gap" between belief and action, it is unnecessary. But we need not restrict ourselves to this intellectualised way of seeing the problem. There may be some gap between an agent's (inarticulate) apprehension of the world and her response to it. We can reasonably ask why she responded both at all, and in this particular way. And this is a gap that can be filled by habits. In that sense, we can regard habits as motivators.

This chimes with some other things Norman says. For earlier in the same paper he suggests that we can be "'motivated by pity' or 'motivated by ambition' or 'motivated by a sense of justice'" (2001, p. 7). If these character traits can be thought of as habits, we may be able to capture Norman's intuitions that they are motivators, without understanding them, as he does, as "ways of characterising the kind of reason for which the agent acts" (p. 7). Clearly there are implications here for the way virtues can be said to motivate.[3] But these matters must be taken up elsewhere.

6. Action Naturalism Enhanced

In Chapter 1, §4, I said I would return to the issue of action naturalism, and specifically, to the question of how McDowell's "naturalism of second nature" might be helped by this thesis. We saw there that McDowell thinks that our upbringing instils in us a second nature which consists in "habits of thought and action" (1996, p. 84). And I suggested that whilst I applaud McDowell's general naturalistic strategy, some of the details of how this transition, during which our "eyes [are] opened to

---

[3] I think the suggestion would be compatible with what Rosalind Hursthouse (1999) has recently concluded: "what is both necessary and sufficient for a virtuous act to be 'morally motivated' is that it is done from a state of character that adequately resembles the state of character from which the perfectly virtuous agent acts" (pp. 159-60). I am however uneasy about the idea of a "perfectly virtuous agent".

reasons at large" (p. 84), remain obscure. We are now in a position to make some headway in this respect.

The most obvious thing to say is that having defended the claim that many exercises of habits are rational actions, it is now much clearer how acquiring habits can result in rational actions. Hitherto, intellectualism in our theory of rational action has prevented us from understanding how habitual actions can be rational. Thus my claim that habitual actions can be rational makes McDowell's naturalism more plausible.

But there is a difficulty with reading McDowell's "habits of thought and action" in this way. For according to the account of responsible actions presented in Chapter 4, an agent's personal responsibility for a behaviour crucially depends upon the possibility that she could have acted otherwise for reasons. Without a naturalistic account of how *that* capacity came about, the only "habits" we in fact have a naturalistic story about are merely *proto-habits*, as we might call them, which enable habitual *behaviours*, but these fall short habitual *actions* proper. As things stand, then, the above proposal may not be as much help for McDowell's naturalism as we might have hoped.

I want to suggest a possible solution: the capacity for acting for reasons *just is* a rational habit which is acquired like all the others.[4] Admittedly, the capacity to act for reasons must be a particularly sophisticated kind of habit, in the sense that it will depend upon the agent's having a host of other rational habits, such as the habits constitutive of language use, and reasoning. But this should be no bar to it's being thought of as itself a special sort of habit.

The proposal is that the capacity for acting for reasons is the habit of deliberating in novel circumstances. It fulfils our criteria for being a habit. It is repeated, since we do it often in novel circumstances; we do it automatically, since it does not require deliberation about whether to do it, and does not require us to try to

---

[4] The possibility of an alternative understanding of this idea was mooted earlier in footnotes, see Chapter 4, §7, and Chapter 5, §6.

do it; and since it is possible for us not to deliberate at our choosing, we are responsible for exercises of this habit.

The idea of a habit such as this may seem paradoxical given the definitions with which we are working. After all, I have ruled out any sort of deliberation, which includes reflection on reasons, from what can count as a habitual action, so the idea of a "habit of deliberation" may seem paradoxical. However, this I think is to confuse a habitual *act* of deliberation with the *content* of such an act, or what is being deliberated upon. Just because a given behaviour is one of deliberating, it can still be the case that the agent does not deliberate upon whether to perform that very deliberation. And it is only a deliberation with that sort of content that could not be habitual. So deliberation, like other "mental" actions, can be said to be habitual. This would be one understanding of what McDowell means, in the quotation mentioned above, by "habits of thought". The habits involved in deliberation operate at a higher level than other habits, since all habits depend upon these habits if their exercises are to produce actions. Indeed the habits of deliberation would depend upon themselves, in the sense that it is always possible to opt out of it, and ask whether deliberation is the thing to do here. But high-level habits are habits nonetheless.

To fill out the picture just slightly we can identify particular habits upon which the habit of deliberation will depend, for instance, the habit of employing *modus ponens* in our reasoning. The employment of *modus ponens* deserves to be termed habitual because in all but the most unusual circumstances, when we do it, we do not think about whether to. Thus, in normal circumstances, which are that we think p and $p \rightarrow q$, we habitually think q. We might be thinking of both p and $p \rightarrow q$ before or at the time of this transition, but we shall not be also thinking about whether to perform the inference and conclude q, which would be the kind of deliberation which would disqualify it from being habitual. Rather, we *find ourselves* thinking that q, and that is an indication that the transition was habitual. In fact those who have never come across the idea of *modus ponens* (which will include those who have not studied logic) are unlikely to even realise that there is a question here. But that does not mean that even for them, there is no possible deliberation about whether to make the transition to q. The fact that it is unclear what sort of reasoning would lead to anybody deciding

to do otherwise just shows how fundamental to rational thought this particular habit is.[5]

So the naturalistic story would be one of an individual's acquiring first a set of proto-habits, which would eventually include more sophisticated proto-habits of reflection, deliberation, justification and criticism. Once these latter proto-habits are acquired, the individual can be said to be capable of actions proper, both habitual and otherwise. For it is now true to say that she can act otherwise for reasons, and thus that she has rational intervention control.

The striking difference between this and what I have said previously, however, is that the rational capacity to intervene when appropriate, hitherto glossed as acting for reasons, is now itself understood as consisting in a mutually supportive set of rational habits. This is how, as McDowell following Aristotle suggests, the acquisition of second nature can be seen a matter of habit acquisition. Agency itself is habits all the way up.

---

[5] This connects with the suggestion made in Chapter 6, §8, about certain rational habits being constitutive of agency, and thus a necessary part of a Davidsonian "principle of charity".

APPENDIX: SUMMARY OF DEFINITIONS

(R) An agent A's behaviour ϕ in circumstances C is *repeated* if and only if A has on numerous prior occasions encountered circumstances similar to C ("normal circumstances"), and when in such circumstances, A usually, and on most of the recent occasions, behaved in a similar way to ϕ. (p. 61)

(A) A behaviour ϕ is *automatic* if and only if the agent performing ϕ engages in no process of conscious deliberation about whether to ϕ, either before ϕ is performed, or during its performance; *and* the agent does not try to ϕ. (p. 73)

(HB) A behaviour ϕ is *habitual* if and only if it is repeated according to (R) and automatic according to (A). (p. 77)

(RIC) An agent has *rational intervention control* over a behaviour if and only if she has both a capacity to notice when an intervention on that behaviour is rationally appropriate, and a capacity to intervene appropriately on such occasions. (p. 96)

(HA) A behaviour ϕ is a *habitual action* if and only if it is repeated according to (R), automatic according to (A), and the agent has rational intervention control over ϕ. (p. 101)

(H) An agent has a *habit of ϕ-ing in C* if and only if the agent performed habitual action ϕ on most of her recent encounters with C. (p. 102)

(PC) An action is rational if and only if it is not irrational. (p. 144)

(RH) A habit is rational if and only if it is not irrational. (p. 158)

(IH) A habit is irrational if and only if its exercises are irrational. (p. 158)

(HE) Agent with the habit of ϕ-ing in C + encounter with normal circumstances C $\rightarrow$ habitual action ϕ. (p. 194)

Chisholm, Roderick M. 1976: "The Agent as Cause", in Myles Brand & Douglas Walton (eds.) *Action Theory*, D. Reidel, pp. 199-211.

Collins, Arthur W. 1987: *The Nature of Mental Things*. University of Notre Dame Press.

——— 1997: "On the Psychological Reality of Reasons", *Ratio*, 10, pp. 108-23.

Cooper, John M. 1986: *Reason and Human Good in Aristotle*. Hackett.

Copp, David 1997: "Belief, Reason, and Motivation: Michael Smith's *The Moral Problem*", *Ethics*, 108, pp. 33-54.

Crocker, Thomas P. 1998: "Wittgenstein's Practices and Peirce's Habits: Agreement in Human Activity", *History of Philosophy Quarterly*, 15, pp. 475-93.

Dancy, Jonathan 1993: *Moral Reasons*. Blackwell.

——— 1995a: "Arguments from Illusion", *Philosophical Quarterly*, 45, pp. 421-38.

——— 1995b: "Why There is Really No Such Thing as the Theory of Motivation", *Proceedings of the Aristotelian Society*, 95, pp. 1-18.

——— 1996: "Critical Notice of *The Moral Problem*", *Ratio*, 9, pp. 171-83.

——— 2000: *Practical Reality*. OUP.

Danto, Arthur C. 1973: *Analytical Philosophy of Action*. CUP.

Davidson, Donald 1963: "Actions, Reasons, and Causes", reprinted in Davidson 1980, pp. 3-19.

——— 1973: "Freedom to Act", reprinted in Davidson 1980, pp. 63-81.

——— 1980: *Essays on Actions and Events*. Clarendon Press: Oxford.

——— 1982: "Paradoxes of Irrationality", reprinted in Paul K. Moser (ed.) *Rationality in Action: Contemporary Approaches*, CUP, 1990, pp. 449-64.

——— 1984: *Inquiries into Truth & Interpretation*. Clarendon Press: Oxford.

Dreier, James 1996: "Review of *The Moral Problem*", *Mind*, 105, pp. 363-7.

Dretske, Fred 1988: *Explaining Behaviour: Reasons in a World of Causes*. Bradford: MIT Press.

Fischer, John M. & Ravizza, Mark 1998: *Responsibility and Control: A Theory of Moral Responsibility*. CUP.

Fodor, Jerry. A. 1987: "Why There Still Has to Be a Language of Thought", reprinted in William G. Lycan (ed.) *Mind and Cognition: A Reader*, Blackwell, 1990, pp. 282-99.

Fogelin, Robert J. 1987: *Wittgenstein*, Second Edn.. Routledge.

Foot, Philippa 1978: *Virtues and Vices*. Blackwell.

Frankfurt, Harry G. 1969: "Alternate Possibilities and Moral Responsibility", reprinted in Frankfurt 1988, pp. 1-10.

——— 1978: "The Problem of Action", reprinted in Frankfurt 1988, pp. 69-79.

——— 1988: *The Importance of What We Care About*. CUP.

Gert, Bernard 1998: *Morality: Its Nature and Justification*. OUP.

Gibbard, Allan 1990: *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. Clarendon Press: Oxford.

Ginet, Carl 1990: *On Action*. CUP.

Glock, Hanjo & Preston, John 1995: "Externalism and First-Person Authority", *Monist*, 78, pp. 515-33.

Goldman, Alvin 1970: *A Theory of Human Action*. Prentice-Hall.

Götz, Ignacio L. 1989: "Is There a Moral Skill?", *Educational Theory*, 39, pp. 11-16.

Grice, H. P. 1961: "The Causal Theory of Perception", *Proceedings of the Aristotelian Society Supplementary Volume*, 35, pp. 121-52.

Haldane, John & Wright, Crispin (eds.) 1993: *Reality, Representation, and Projection*. OUP.

Hamm, C. M. 1975: "The Role of Habit in Moral Education", *Educational Theory*, 25, pp. 417-28.

Hinton, J. M. 1973: *Experiences*. OUP.

Hooker, Brad (ed.) 1996: *Truth in Ethics*. Blackwell.

Hookway, Christopher 2000: "Epistemic Norms and Theoretical Deliberation", in Jonathan Dancy (ed.) *Normativity*, Blackwell, 2000, pp. 60-77.

Hornsby, Jennifer 1980: *Actions*. Routledge & Kegan Paul.

——— 1982: "Reply to Lowe on Actions", *Analysis*, 42, pp. 152-3.

——— 1983: "Events that are Causings: A Response to Lowe", *Analysis*, 43, pp. 141-2.

——— 1997: *Simple Mindedness*. Harvard UP.

Humberstone, I. L. 1992: "Direction of Fit", *Mind*, 101, pp. 59-83.

Hume, David 1740: *A Treatise of Human Nature*, 2nd edn., OUP, 1978.

——— 1777: *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, 3rd edn., OUP, 1975.

Hursthouse, Rosalind 1991: "Arational Actions", *Journal of Philosophy*, 88, pp. 57-68.

——— 1999: *On Virtue Ethics*. OUP.

Jones, O. R. (ed.) 1971: *The Private Language Argument*. MacMillan: St. Martin's Press.

Karlsson, Mikael M. 2002: "Agency and Patiency: Back to Nature?", *Philosophical Explorations*, 5, pp. 59-81.

Kenny, Anthony 1989: *The Metaphysics of Mind*. Clarendon Press: Oxford.

Kripke, Saul A. 1982: *Wittgenstein on Rules and Private Language*. Blackwell.

Lewis, David K. 1988: "Desire as Belief", *Mind*, 97, pp. 323-32.

——— 1996: "Desire as Belief II", *Mind*, 105, pp. 303-13.

Lovibond, Sabina & Williams, S. G. (eds.) 1996: *Essays for David Wiggins: Identity, Truth and Value*. Blackwell.

Lovibond, Sabina 1996: "Ethical Upbringing: from Connivance to Cognition", in Lovibond & Williams 1996, pp. 76-94.

Lowe, E. J. 1978: "Neither Intentional nor Unintentional", *Analysis*, 38, pp. 117-8.

——— 1980: "An Analysis of Intentionality", *Philosophical Quarterly*, 30, pp. 294-304.

——— 1981: "'All Actions Occur Inside the Body'", *Analysis*, 41, pp. 126-9.

——— 1983: "Reply to Hornsby on Actions", *Analysis*, 43, pp. 140-1.

——— 1984: "A Note on a Response of Hornsby's", *Analysis*, 44, pp. 196-7.

——— 1996: *Subjects of Experience*. CUP.

——— 2000: *An Introduction to the Philosophy of Mind.* CUP.

MacIntyre, Alasdair 1981: *After Virtue.* University of Notre Dame Press.

Mackie, J. L. 1977: *Ethics: Inventing Right and Wrong.* Penguin, Harmondsworth.

Makin, Stephen 2000: "Aristotle on Modality: How Many Ways Can a Capacity be Exercised?", *Proceedings of the Aristotelian Society Supplementary Volume,* 74, pp. 143-61.

Malcolm, Norman 1986: "Following a Rule". In *Nothing is Hidden,* Basil Blackwell, Chapter 9, pp. 154-81.

——— 1989: "Wittgenstein on Language and Rules", *Philosophy,* 64, pp. 5-28.

McDowell, John H. 1978: "Are Moral Requirements Hypothetical Imperatives?", reprinted in McDowell 1998, pp. 77-94.

——— 1979: "Virtue and Reason", reprinted in McDowell 1998, pp. 50-73.

——— 1981: "Non-Cognitivism and Rule-Following", reprinted in McDowell 1998, pp. 198-218.

——— 1982: "Criteria, Defeasibility and Knowledge", *Proceedings of the British Academy,* 68, pp. 455-79.

——— 1984: "Wittgenstein on Following a Rule". Reprinted in McDowell 1998, pp. 221-62.

——— 1985: "Values and Secondary Qualities", reprinted in McDowell 1998, pp. 131-50.

——— 1987: "Projection and Truth in Ethics", reprinted in McDowell 1998, pp. 151-66.

——— 1993: "Meaning and Intentionality in Wittgenstein's Later Philosophy", reprinted in McDowell 1998, pp. 263-78.

——— 1995: "Knowledge and the Internal", *Philosophy and Phenomenological Research,* 55, pp. 877-93.

——— 1995a: "Two Sorts of Naturalism", reprinted in McDowell 1998, pp. 167-97.

——— 1995b: "Might There Be External Reasons?", reprinted in McDowell 1998, pp. 95-111.

——— 1996: *Mind and World: With a New Introduction by the Author.* 2nd edn. Harvard UP.

——— 1998: *Mind, Value, and Reality.* Harvard UP.

McGinn, Colin 1984: *Wittgenstein on Meaning.* Basil Blackwell.

Mele, Alfred R. (ed.) 1997: *The Philosophy of Action.* OUP.

——— 1992a: *Springs of Action: Understanding Intentional Behaviour.* OUP.

——— 1992b: "Recent Work on Intentional Action", *American Philosophical Quarterly,* 29, pp. 199-217.

Millikan, Ruth G 1995: "Pushmi-pullyu Representations", reprinted in Larry May, Marilyn Friedman and Andy Clark (eds.) *Mind and Morals: Essays on Cognitive Science and Ethics,* Bradford: MIT Press, 1996, pp. 145-61.

Mixon, Don 1980: "The Place of Habit in the Control of Action", *Journal for the Theory of Social Behaviour,* 10, pp. 169-86.

Morton, Adam 1975: "Because he Thought he had Insulted him", *Journal of Philosophy*, 72, pp. 5-15.

Moya, Carlos J. 1990: *The Philosophy of Action: An Introduction*. Polity.

Mumford, Stephen 1998: *Dispositions*. OUP.

Nagel, Thomas 1970: *The Possibility of Altruism*. Clarendon Press: Oxford.

Norman, Richard 2001: "Practical Reasons and the Redundancy of Motives", *Ethical Theory and Moral Practice*, 4, pp. 3-22.

O'Shaughnessy, Brian 1973: "Trying (as the Mental 'Pineal Gland')", reprinted in Mele 1997, pp. 53-74.

Peters, R. S. 1958: *The Concept of Motivation*. Routledge & Kegan Paul.

——— 1963: "Reason and Habit: The Paradox of Moral Education", in W. R. Niblett (ed.) *Moral Education in a Changing Society*, Faber & Faber, pp. 46-65.

Pettit, Philip 1993: *The Common Mind*. OUP.

Platts, Mark de Bretton 1979: *Ways of Meaning*. Routledge and Kegan Paul.

——— 1981: "Moral Reality and the End of Desire", in Mark Platts (ed.) *Reference, Truth and Reality*, Routledge and Kegan Paul, pp. 69-82.

Pollard, Bill 1998: "Blackburn's *Ruling Passions*: A Partial Reply", *Philosophical Writings*, 9, pp. 49-64.

Puhl, Klaus (ed.) 1991: *Meaning Scepticism*. Walter de Gruyter.

Quinn, Warren 1993: *Morality and Action*. CUP.

Rhees, Rush 1954: "Could Language be Invented by a Robinson Crusoe? Reply to Ayer". In Jones 1971, pp. 61-75.

Rorty, Amélie Oksenberg 1980: *Essays on Aristotle's Ethics*. University of California Press: Berkeley.

Ryle, Gilbert 1949: *The Concept of Mind*. Penguin.

Sayre-McCord, Geoffrey 1997: "The Metaethical Problem", *Ethics*, 108, pp. 55-83.

Searle, John 1983: *Intentionality*. CUP.

Sellars, Wilfrid 1956: "Empiricism and the Philosophy of Mind", *Minnesota Studies in the Philosophy of Science*, 1, University of Minnesota Press, pp. 253-329.

Skinner, B. F. 1953: *Science and Human Behaviour*. New York.

Smith, Michael 1987: "The Humean Theory of Motivation", *Mind*, 96, pp. 36-61.

——— 1994: *The Moral Problem*. Blackwell.

——— 1997: "In Defense of The Moral Problem: A Reply to Brink, Copp, and Sayre-McCord", *Ethics*, 108, pp. 84-119.

——— 1998: "The Possibility of Philosophy of Action", in Bransen & Cuypers 1998, pp. 17-41.

Snowdon, Paul 1981: "Perception, Vision and Causation", *Proceedings of the Aristotelian Society*, 81, pp. 175-92.

Stocker, Michael 1979: "Desiring the Bad: An Essay in Moral Psychology", *Journal of Philosophy*, 66, pp. 738-53.

Stout, Rowland 1996: *Things That Happen Because They Should: A Teleological Approach to Action*. Clarendon Press: Oxford.

Stanley, Jason & Williamson, Timothy 2001: "Knowing How", *Journal of Philosophy*, 98, pp. 411-44.

Stoutland, Frederick 1998: "The Real Reasons", in Bransen & Cuypers 1998, pp. 43-66.

Strawson, Peter 1962: "Freedom and Resentment", reprinted in Strawson 1974, pp. 1-25.

——— 1974: *Freedom and Resentment and Other Essays*. Methuen.

Sutherland, N. S. 1959: "Motives as Explanations", *Mind*, 78, pp. 145-59.

Taylor, Richard 1966: *Action and Purpose*. Prentice Hall.

Toulmin, Stephen 2001: *Return to Reason*. Harvard University Press.

Van Fraassen, Bas C. 1989: *Laws and Symmetry*. Clarendon Press: Oxford.

——— unpublished: "Scientific Revolution as a Philosophical Problem", RIP Lecture, University of Durham, 23rd May 2001.

Walker, Arthur F. 1989: "The Problem of Weakness of Will", *Noûs*, 23, pp. 653-76.

Wallace, R. Jay 1994: *Responsibility and the Moral Sentiments*. Harvard UP.

Watson, J. B. 1919: *Psychology from the Standpoint of a Behaviourist*. Philadelphia.

Wiggins, David 1991: "Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs", *Proceedings of the Aristotelian Society*, 91, pp. 61-85.

——— 1996: "Objective and Subjective in Ethics, with Two Postscripts on Truth", in Hooker 1996, pp. 35-50.

Williams, Bernard 1981: "Internal and External Reasons", in his *Moral Luck*, CUP, pp. 101-13.

Williams, Meredith 1991: "Blind Obedience: Rules, Community and the Individual", in Puhl 1991, pp. 93-125.

——— 1994: "The Significance of Learning in Wittgenstein's Later Philosophy", *Canadian Journal of Philosophy*, 24, pp. 173-204.

Williams, Michael 1996: "Exorcism and Enchantment", *Philosophical Quarterly*, 46, pp. 99-109.

Winch, Peter 1958: *The Idea of a Social Science and its Relation to Philosophy*. Routledge & Kegan Paul.

Wittgenstein, Ludwig 1953: *Philosophical Investigations*. Basil Blackwell.

Wright, Crispin 1980: *Wittgenstein on the Foundations of Mathematics*. London: Duckworth.

——— 1985: "Review of *Spreading the Word*", *Mind*, 94, pp. 310-19.

——— 1996: "Human Nature?", *European Journal of Philosophy*, 4, pp. 235-54.

Wright, John P. 1994: "Butler and Hume on Habit and Moral Character", in M. A. Stewart & John P. Wright (eds.) *Hume and Hume's Connexions*, Edinburgh University Press, pp. 105-18.