THE UNIVERSITY OF CHICAGO

ACTING FOR A REASON: A WITTGENSTEINIAN APPROACH

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE HUMANITIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

BY
AMOS BROWNE

CHICAGO, ILLINOIS
AUGUST 2018

For my parents, Francis Browne and Gail Riminton.

How could human behaviour be described? Surely only by sketching the actions of a variety of humans, as they are all mixed up together. What determines our judgement, our concepts and reactions, is not what *one* man is doing *now*, an individual action, but the whole hurly-burly of human actions, the background against which we see any action.

Zettel §*567*

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I must first thank my committee: Jason Bridges, Jim Conant, and David Finkelstein. Though I came to Chicago with an interest in Wittgenstein, I could not have anticipated the extent to which it would be revised and deepened under their guidance. Interactions with Jason and David in classes during my first few years as a graduate student (as well as at the now defunct Philosophy of Mind Workshop) cemented my interest in the relation between Wittgenstein's ideas and contemporary issues in the philosophy of mind and action. Later, serving as a course assistant for Jim in a several classes transformed my understanding of Wittgenstein's work, and introduced me to a number of other authors whose influence runs through this dissertation (in particular G.E.M. Anscombe, Cora Diamond, and Rush Rhees). This dissertation project came out of a seminar on 'Reasons' taught by Jason in Autumn 2013. Without his intervention and encouragement at two key moments over the past few years, I would neither have conceived of this dissertation nor seen it through to completion. I would also like to thank my undergraduate advisor, Christopher Shields – without his mentorship I would never have considered applying to graduate school or pursuing a PhD in philosophy.

I knew very little about the department at the University of Chicago when I was offered a place here. After visiting I left with the impression of a lively and engaged community of graduate students, and accepted because of it. That impression proved accurate, and my philosophical education here came mostly from my fellow students. Conversations and workshops with Stina Bäckström, Nir Ben Moshe, Rachel Goodman, Katie Howe, Santiago Mejia, Gilad Nir, Daniel Rodriguez Navas, Rory O'Connell, Francey Russell, Will Small, Tuomo Tiisala, Jessica Tizzard, and Andrew Werner particularly informed this dissertation (or the ideas that led to it). I am grateful to Nicholas Koziolek most of all – I have learnt a great deal from our conversations over the years, but more than anything else I've depended on his unwavering interest in what I had to say about the philosophy I'd been reading, and his ability to help me develop my scattered thoughts into something coherent. There is noone with whom I would rather discuss philosophy. I would also like to thank Justin Spain and Jacob Swenson – though they left the program before I started this dissertation, their friendship is one of the things I value most from my time here.

In addition to the members of my committee, I would particularly like to thank Michael Kremer and Irad Kimhi for what I have learnt from them. Michael taught my First Year Seminar and a number of other classes to which I owe any understanding I have of the history of analytic philosophy. He also gave me work as a research assistant at a point when I was struggling to support myself financially. It has taken me a long time to begin to understand Irad's work, but my way into it has been through his comments on both the *Tractatus* and the *Philosophical Investigations* – he has a capacity to illuminate those works and connect them to other figures in the history of philosophy in ways I can never anticipate. I also feel particularly lucky to have discovered Cora Diamond's work during my time at graduate school, and to have attended an ALWS summer school taught by her and Jim during the summer of 2013. I would like to thank the department as a whole for providing me with a dissertation year fellowship during 2015-6. I would not have completed this dissertation without it.

Outside of the philosophy department, I would like to thank Maren Robinson, Hilary Strang, and all of the preceptors and staff at the MAPH program. They not only provided me with secure and stable employment while finishing my dissertation, but also with a supportive and stimulating work environment unlike anything else I've experienced in academia.

Perhaps the hardest thing about doing my PhD here has been the distance to my friends and family at home in England. This dissertation is dedicated to my parents – I hope I am always guided by what I have learnt from them. I now look forward to my time in England more than any other part of the year – nothing could be more fun than an evening in the pub with Simon, Jacob, Matthew, Nina, and Jess. I want to thank all of them for their support and encouragement. Sandi and Rick Tizzard welcomed me into their home and made family seem less distant. I'm also grateful to all of the friends who have kept up with, hosted,

and visited me while I've been here – particularly Liban Dirye, Chrissy Downing, Anna Gomberg, Stephanie Burnett Heyes, Beth Lawrence, Johnny Petevinou, Andrew Walters, and Alice Wishart.

Finally I want to thank my partner, Jessica Tizzard, for everything. Her love and company are what matter most.

CHAPTER 1

INTRODUCTION

The aim of this dissertation is to show the ongoing relevance of Ludwig Wittgenstein's ideas to contemporary discussions in the philosophy of action and philosophy of mind. My central contention is that Wittgenstein's discussions of simple kinds of intelligent behaviour such as reading or sign-following provide the basis for an instructive critical perspective on contemporary accounts of 'acting for a reason', along with a model for an alternative approach to this topic.

Reflection on these simple kinds of intelligent behaviour can lead to various philosophical puzzles. Imagine the following scenario: you see a friend come to a crossroad and follow a sign pointing to the left. In acting this way, your friend manifests an understanding of the sign: it pointed left, and she responded by following it. Though her response could have been quite automatic—she needn't have given any thought to it—if you asked her why she turned to the left, she could explain her action by pointing to the sign.

Reflection on this case might lead us to wonder what made her act of turning left a case of 'following the sign'. After all, we can imagine a companion case, in which a stranger who was unfamiliar with our practices of sign-posting came to the same crossroad, looked at what was to them merely a plank of wood with a tapered end, and then turned to the left. Perhaps if he were asked why he turned left, the stranger would explain his action by saying 'I looked at that piece of wood, and it made me decide to turn left'. For all that, what he did does not count as 'following the sign'.

Wherein lies the difference between these two cases? A philosopher might propose the following response. For the stranger, the sign was merely a *cause* of his action – he saw it, and it prompted him to turn left; but for your friend, the sign was a *reason* for her action – she saw it, and it prompted her to turn left, but she also took the sign as a reason to turn left, and would think of her action as a correct response to it.

However saying this much seems to merely redescribe the difference we want to explain.

For we can now ask what makes your friend's response count as an instance of 'acting for a reason'. After all, we said her response was quite automatic—she didn't give it any thought—so there seems to be nothing that happened at the crossroads to mark this difference. What then does this 'acting for a reason' consist in?

***

Reflections such as these provide the background to the philosophical questions that frame contemporary discussion of 'reasons for action'. The first I shall call a 'metaphysical question' (MQ):

Metaphysical Question: What is it to act for a reason?

The basis for this question is our puzzle about what differentiates certain kinds of intelligent activity from other kinds of behaviour. But now this puzzle is dressed up in more formal philosophical garb: (MQ) is a 'metaphysical question' because it asks 'what is it to be such-and-such?' – a question that could also be phrased 'what is essential to being such-and-such' or 'what does being such-and-such consist in?'. Here these metaphysical questions are directed at certain kinds of intelligent behaviour: not just the simple case of sign-following, but any case in which a person can be described as 'acting for a reason'. They ask us what makes some activity count as such, and thus what differentiates it from other kinds of behaviour that we would not describe in this way.

The second I shall call an 'epistemological question' (EQ):

Epistemological Question: How do we come to have first-personal knowledge of such acts?

The 'first-personal knowledge' described in this question is the knowledge that expresses the agent's understanding of what she is doing and why she is doing it. (EQ) is an 'epistemological question' because it asks 'how is it that we come to know such-and-such?'. For the observation we made in our simple case of sign-following holds quite generally: when a person acts for a reason, she can explain her action by citing that reason.

***

The contemporary accounts I shall focus on approach (MQ) and (EQ) by way of an investigation into 'rational' or 'reason-giving' explanations. These take such forms as,

> S's reason for ϕing is that _____,

or

> S ϕed because _____,

where ϕ is replaced by a description of the relevant act, and the *explanans* by a description of the person's reason for that act. The general notion of a 'reason-giving explanation' emerges from reflection on the explanatory dimension of our talk about people's actions and attitudes. Speaking in broad terms, they are concerned with *making sense* of the things people say and do. It is an essential feature of such explanations that the understanding they aim to provide should reflect the understanding from which the agent herself was acting.

Our simple example provided us with a specific version of such explanations:

> S turned left because the sign pointed that way.

This is intended to articulate the answer your friend would provide if we asked her why she turned to the left. Part of what it is to make sense of her act *as* an act of sign-following is to see it as subject to explanations of this sort.

Explanations in this general category might be thought of as representing something essential to the idea of 'acting for a reason', since to act for a reason *is* for one's act to be explainable in this way. Moreover, explanations of this form could be used to give expression to the first-personal knowledge that the agent has of her action. Thus, the general idea guiding this approach is that an account of how such explanations do their work will provide the basis for a response to both (MQ) and (EQ).

Such an account could therefore explain the difference between an explanation,

S1: S turned left because$_R$ the sign pointed that way,

where the subscript marks the fact that the *explanandum* describes an instance of acting for a reason, and the *explanans* gives that reason, and an apparently similar explanation,

S2: S turned left because$_C$ the sign pointed that way,

where the subscript marks the fact that the sign merely caused the action described in the *explanandum*, so that the *explanans* does not give S's reason for that action.

***

My primary goal in this dissertation is to outline two different forms of response to these questions and the puzzles that lead to them. The first approach aims to provide what I shall call a 'substantive response' based on a philosophical account of *what it is to act for a reason* and *how we come to have first-personal knowledge of such acts.* Such accounts understand these questions as placing a coherent demand on us, and aim to respond to it by specifying *in virtue of what* something counts (and is known) as an instance of acting for a reason. As we shall see, most contemporary accounts of 'acting for a reason' at least appear to be giving a response of this form.

The second approach derives from the ideas of Ludwig Wittgenstein, particularly as they were developed in the work of the philosopher G.E.M. Anscombe. This approach rejects the idea that there could be a substantive and general account of *what it is to act for a reason* or *how we come to have first-personal knowledge of such acts*, and sees the apparent pressure to provide such an account as having its origin in various philosophical confusions. In place of a general and substantive account, it instead answers the question 'what is it to act for a reason?' by describing particular examples which we would count as instances of acting for a reason. It thus has the form of a reminder: to act for a reason is to act like *this*, or like *this*, or .... One of the key questions that I aim to address in what follows is how a response of this form might help dissipate the puzzles from which we started: for instance, the question of what distinguishes acting for a reason from other kinds of behaviour.

***

Appeal to the work of Wittgenstein and Anscombe in relation to this topic introduces a further issue. The consensus among contemporary accounts of 'acting for a reason' is that both philosophers were committed to the following claim: a reason for an action is not a cause of that action. Their apparent commitment to this claim is understood to have had a deleterious effect on their overall approach to the topic of 'acting for a reason' – for instance, by committing them to the further claim that (S1) above cannot be understood as an instance of causal explanation.

A final goal of this dissertation is to provide grounds for questioning this judgement. This task has two parts. First, I show that the passages of Wittgenstein and Anscombe's work that are supposed to commit them to this claim have in fact been misunderstood by most of the contemporary literature. Second, I present my own approach to the topic of 'acting for a reason' through Wittgenstein's discussion of a particular class of acts of this kind: acts of reading, along with other kinds of response such as following a sign or obeying an order.

Focusing on such acts has several advantages. First, since they are instances of 'acting for a reason', if a general and substantive response to (MQ) and (EQ) is possible, it ought to apply to them. Thus if I can show that no such account is possible for these cases, I will have shown that no general account is possible either.

Second, by focusing on such acts I can show that a Wittgensteinian approach can explain what distinguishes, say, a case of sign-following from some other response to a sign, and thus what distinguishes

S1: S turned left because$_R$ the sign pointed that way,

from

S2: S turned left because$_C$ the sign pointed that way,

while also showing that this approach does not commit us to the claim that (S1) does not represent a causal transaction. Doing so provides a small contribution to showing how acts of reason might fit into the causal order.

***

The structure of this dissertation is as follows.

Chapters 2-4 provide the foundation for my central discussion. In §2 I provide a survey of contemporary accounts of reason-giving explanations, and show that these accounts appear to be framed by a shared conception of (MQ) and (EQ). In §3, I begin to sketch an alternative Wittgensteinian approach to this topic, partly through a discussion of work by G.E.M. Anscombe. In §4, I show that once we understand this approach, we can see that Wittgenstein and Anscombe are not in fact committed to the strict differentiation between reasons and causes that is attributed to them by contemporary accounts.

Chapters 5-7 provide a critical assessment of these contemporary accounts. §5 introduces Wittgenstein's discussion of *reading* in §156-§178 of the *Philosophical Investigations*, and shows that this discussion provides the basis for a critical assessment of substantive responses to (MQ) and (EQ). §6 then applies this criticism to one form of contemporary account: the causal-psychologism defended by philosophers such as Kieran Setiya and Wayne Davis. §7 considers a different form of account: the normativism defended by Eric Marcus and Sebastian Rödl. Ultimately, this form of account turns out to be an unsuccessful attempt to apply the Wittgensteinian approach outlined in §3 to (MQ) and (EQ).

Finally, in §8 I show that the dissertation as a whole has provided the basis for a different kind of response to (MQ) – though not to (EQ).[1] Here I aim to develop insights from earlier chapters to demonstrate what is involved in taking a Wittgensteinian approach to the idea of 'acting for a reason'.

[1] This is partly because rejecting the idea of a substantive response to (MQ) relieves some of the pressure to provide a substantive response to (EQ). Ultimately, however, a development of the ideas in this dissertation would have to show how an alternative response to (EQ) was also possible.

CHAPTER 2

ACCOUNTS OF REASON-GIVING EXPLANATION

2.1 Introduction

As we saw in §1, contemporary discussion of reasons for action is framed by the following two questions:

Metaphysical Question: What is it to act for a reason?

Epistemological Question: How do we come to have first-personal knowledge of such acts?

In this first chapter, I will provide a survey of contemporary responses to these questions. As well as outlining the overall state of the discussion, my goal is to bring out a common understanding of (MQ) and (EQ) that shapes what is counted as an adequate response. This will lay the foundation for a comparison with Wittgenstein' and Anscombe's treatment of these topics in subsequent chapters.

The chapter itself falls into three parts. In §2.2, I show how (MQ) and (EQ) came to frame contemporary discussion, by showing how contemporary accounts emerge out of earlier debates from the mid-twentieth century. Next, in §2.3, I introduce three different kinds of account of rational or reason-giving explanation, showing how each responds our central questions. In §2.4, I outline the common understanding of (MQ) and (EQ) that underlies these different accounts, and show how this shapes their various conceptions of what counts as an adequate response to them.

2.2 The Background to the Contemporary Debate

2.2.1 The Wittgensteinian Orthodoxy

The contemporary debate about reason-giving explanations can be seen as emerging in response to particular claims about ‘reasons for action’ traceable to Wittgenstein’s work, along with the writings of a number of philosophers influenced by his ideas. While we shall need to look more closely at Wittgenstein’s treatment of this topic in subsequent chapters, the following sketch of what I take to be a common reading of his work will serve for present purposes (though note that the reading of Wittgenstein and Anscombe’s work presented here will be challenged later):

> In his mid-period works such as *The Blue and the Brown Books*, Wittgenstein claimed that the concepts ‘cause’ and ‘reason’ had distinctive grammars, and suggested that confusion about these grammars was a common source of philosophical problems. In general, Wittgenstein was coming to see confusions of this form as a key source of philosophical puzzlement. In *The Blue Book* he states that “[w]hen words in our ordinary language have *prima facie* analagous grammars . . . we are inclined to try to interpret them analogously; i.e. we try to make the analogy hold throughout” (BB7). The “ambiguous use of the word ‘why”’—sometimes asking for a cause of our actions, sometimes for a reason—leads us to confuse the two cases. Taking a case where the question ‘why did you do such-and-such?’ asks for a *cause* for our action, we are inclined to think that the ‘grammar’ of this particular case must apply to all others, including those in which a question of the same form asks for a *reason.* However, Wittgenstein asserted that there were important differences between these cases, such as the fact that “we can only conjecture the cause but we know the motive [or reason for our action]” (BB15).

Wittgenstein’s insistence on the distinction between reasons and causes, together with his emphasis on the kind of knowledge that one has of one's reasons, inspired many readers of his work to investigate the ways in which the 'grammar' of reason-giving explanation differed from that of causal explanation. G.E.M. Anscombe frames her investigation in her monograph *Intention* with the question "what distinguishes actions which are intentional from those which are not?", and continues

> The answer I shall suggest is that they are 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. §5

As Anton Ford has noted, it can seem as though this passage has the following form: "first [Anscombe] poses a question, the one her account will address; and then she announces her answer to it" [21]. This then gives the impression that, on Anscombe's account, what makes an action intentional is its being done for a reason. If one took this to be one of the central theses established by Anscombe's investigation, then the question "what is it to act for a reason?" would seem particularly pressing. For it now seems as though "acting for a reason" must be at the heart of any account of intentional action, such that if we could get clear on *that* then the rest would fall into place.[1]

Wittgenstein and his followers were therefore taken to be committed to some version of the following two claims:

1. We must distinguish between answers to the question 'Why?' that give a reason for an action, and those that give a cause for an action.

2. We have first-personal knowledge of the reasons for our actions, whereas we can only conjecture about the causes.

Since answers to the question 'Why did you do such-and-such' provide explanations for that action, (1) appears to entail that reason-giving explanations must be distinguished from

[1] This is in fact the form taken by Setiya and Davis' accounts: both provide an analysis of *acting for a reason*, and then tentatively suggest that this analysis provides the basis for an account of *intentional action*. See [33] and [18].

causal explanations. (2) can then be understood to mark a further difference between these kinds of explanations: we have first-personal knowledge of reason-giving explanations for our actions, but not of causal explanations. Together, these claims have an obvious bearing on both (MQ) and (EQ).

2.2.2 Davidson's Response

In the introduction *Essays on Actions and Events*, Donald Davidson describes his 1963 paper *Actions, Reasons, and Causes* as "a reaction against the widely accepted doctrine that the explanation of an intentional action in terms of its motives or reasons could not relate reasons and attitude as cause and effect", and further traces the origins of this doctrine back to the portions of *The Blue Book* identified above [17, xii]. The question that opens Davidson's paper should then be understood as posing a challenge to this orthodoxy:

Davidson's Question: What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did?

The relation Davidson asks about is precisely that which holds between a reason and an action when the reason can be given in answer to Anscombe's distinctive sense of the question 'Why?'. It is this relation that is represented in the 'because' of rationalizing or reason-giving explanations of the form '*S* did A because _____'. Davidson's challenge to philosophers who insist on a strict differentiation between reasons and causes is that they characterize this relation, and in doing so explain the force of the 'because' in reason-giving explanations. To Davidson, the bare notion of a reason for action seems insufficient for such purposes:

> [F]or a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action *because* he had the reason. Of course, we can include this idea too in justification;

> but then the notion of justification becomes as dark as the notion of reason until we can account for the force of that ‘because’. [15, 9]

Davidson’s own proposal—which directly contradicts the Wittgensteinian orthodoxy of (1) and (2)—is that “rationalization [i.e. reason-giving or rational explanation] is a species of causal explanation” [15, 3]. This has the advantage of subsuming reason-giving explanation under a familiar pattern:

> One way we can explain an event is by placing it in the context of its cause; cause and event form the sort of pattern that explains the effect, in a sense of ‘explain’ we understand as well as any. If reason and action illustrate a different pattern, that pattern must be identified. [15, 10]

On Davidson’s own account, a reason-giving explanation represents a causal relation holding between a pair of psychological states and the action that is the target of that explanation. As he puts it, “[i]n order to turn the first ‘and’ to ‘because’ in ‘He exercised *and* he wanted to reduce and thought exercise would do it’, we must, as the basic move, [allow that] . . . [a] primary reason for an action is its cause” [15, 11-2]

2.2.3 Anscombe’s Criticism

However, Davidson’s response to the Wittgensteinian orthodoxy was not without its problems, which the Wittgensteinians (and, to his credit, Davidson himself) were quick to pick up on. Considering a question akin to Davidson’s—i.e. what sort of connexion might there be between a want and thought on the one hand, and an action explained by them on the other—Anscombe is abruptly dismissive of his proposal that the “psychological ‘because’ is an ordinary *because* where the *because* clause gives a psychological state”:

> True, not only must I have a reason, it must also ‘operate as my reason’: that is, what I do must be done *in pursuit* of the end and *on grounds* of the belief. But

> no just any act of mine which is caused by my having a certain desire is done in pursuit of the object of desire; not just any act caused by my having a belief is done on the grounds of the belief. [6, 110]

Davidson in fact provided a clear example of the problem that Anscombe describes:

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. The belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally. [16, 79]

The problem here is that, though it would be true to say that the climber acted because of his belief and desire, and though this 'because' would describe a familiar causal relation, it remains the case that the climber did not act *on grounds* of the content of his belief nor *in pursuit* of the object of his desire.

At a minimum, this seems to provide a counter-example that Davidson's account needs to accommodate: his account must be modified to rule out deviant causal chains of this sort.[2] But for Anscombe, the problem is deeper: she says that Davidson and his defenders can "do no more than postulate a 'right' causal chain in the happy security that none such can be found" [6, 110]. In effect, Anscombe appears to double-down on (1), asserting that no causal connection could explain an action in the way that a reason-giving explanation does. For "[i]f a causal connexion were found we could always still ask: 'But was the act done for the sake of the end and in view of the thing believed?"' [6, 110].

2.2.4 Framing the Contemporary Debate

The questions framing contemporary discussion of reason-giving explanation can be seen as emerging from these exchanges. For it looks as though all three authors share a common

[2] As we shall see, this is how it is understood by causal-psychologists influenced by Davidson's work. See below.

question, which is a version of our (MQ), i.e. ‘What is it to act for a reason?’, and propose to approach that question via the topic of reason-giving explanation. Moreover, they seem to reach directly contradictory conclusions. The Davidsonian response is that reason-giving explanations are a kind of causal explanation, and that a movement counts as an instance of acting for a reason in virtue of a causal connection to a pair of psychological states. Wittgenstein and Anscombe reject that claim, and appear to assert that the explanatory connection between a reason and an action it explains cannot be causal. In the current context, this suggests that the account they propose involves some other kind of non-causal explanatory relation *in virtue of which* a movement will count as an instance of acting for a reason.

The use of reason-giving explanations to frame the question combines with the problem of causal deviance to create additional pressure to provide a response to (MQ). To see why, we can supplement Davidson’s climber case with a counterpart scenario, as we did with our simpler example of sign-following in §1: this time, the climber decides to loosen the grip on his companion’s rope in order to rid himself of the weight and danger. In both cases, we could apply the following explanation to the ‘act’ of the climber:

> *S* loosened his grip on the rope because he wanted to rid himself of the weight and danger of his companion, and knew that he could do so by such an action.

But only one of these two explanations would count as a reason-giving explanation. Since it looks as though the basic movement, and what explains it, are common to the two cases, this suggests that we must find something else in virtue of which one counts as a rational-explanation, whereas the other does not. This would then be that *in virtue of which* one movement counted as an instance of acting for a reason, and thus an answer to our question (MQ).

A concern with (EQ) emerges out of this general background. For, as we have seen, it is a mark of acting for a reason that the agent can answer Anscombe’s sense of the question ‘Why?’ just insofar as they are acting. This suggests that she must have first-personal knowledge of such explanations in virtue of being the agent of the acts they represent. So whatever it is that explains what it is to act for a reason should also be able to explain how it is that the agent comes to have a particular kind of first-personal knowledge of those acts. In other words, an answer to (MQ) might also be an answer to (EQ).

2.3 Contemporary Accounts

The contemporary accounts I shall survey in this section can each be understood as proposing a particular kind of response to our question (MQ), and in some cases to (EQ) as well. I shall be concerned with three different kinds of account:

- The first are what I shall call 'primitivist accounts', instances of which include the view defended by Jonathan Dancy. These claim that the explanatory unity represented in reason-giving explanations depends on a *sui generis* relation that must be taken as primitive. Since such explanations have features that distinguish them from causal explanations, primitivists claim that the *sui generis* relation cannot be causal.

- The second category are what I shall call 'reductive accounts', instances of which include the 'causal-psychologism' defended by philosophers such as Kieran Setiya and Wayne Davis. Reductive accounts claim that the explanatory unity represented in reason-giving explanations can be explained in independent terms, and so should not be taken as primitive. For causal-psychologists, this unity is to be explained in terms of causation by particular psychological states. They further claim that we can characterize all three terms in reason-giving explanations (act, cause, causal relation) without appeal to the idea of 'acting for a reason'. The reductivist thus hopes to 'build up' an account of the explanatory unity depicted in such explanations from independently available parts, and in doing to to provide a response to (MQ).[3]

[3] Setiya further aims to answer our epistemological question (EQ) by specifying a self-referential psychological state which both plays the role of cause, and represents itself as playing this role. Insofar as what

- The third category are what I shall call 'non-reductive accounts', which reject the reductivist's demand that we explain the explanatory unity represented in such explanations in terms that are independent of 'acting for a reason', while also rejecting the primitivist's demand that we treat this unity as primitive. The non-reductive account I shall be primarily concerned with is the 'normativist' account defended by Eric Marcus. I call his account 'normativist' because its central claim is that what it is to act for a reason is to represent one's action as inheriting a particular normative status. Describing intentional actions, Marcus states that "[a]cting-for-a-reason is representing ϕing as to be done as a consequence of the to-be-done-ness of ψing" [24, 111]. He adds that reason-giving explanations appeal to an ability to represent one's actions in this way. This account is non-reductive because its characterization of the ability whose acts constitute the unity represented in reason-giving explanations openly deploys notions that depend on the idea of 'acting for a reason', i.e. the idea that one action is to-be-done because another is to-be-done. Nevertheless, it still purports to provide an account of what it is to act for a reason, insofar as it claims that the explanatory unity represented in these explanations is constituted by a distinctive 'rational causation'.[4]

All three categories of account should be understood to be concerned with the explanatory unity represented in reason-giving explanations (i.e. *explanans*, *explanandum*, and the explanatory relation between them), but understand the nature of this unity in different ways. The reductivist thinks that we can 'build up' to the relevant unity from independently specifiable parts, and that once we have done this we will have shown in virtue of

it is to act for a reason involves a cause of this sort, the agent will have knowledge of the causal-explanatory relation just insofar as she is acting for a reason.

[4] Since this explanatory relation is constituted by the acts of this representational ability, it also serves to explain the first-personal knowledge that the agent has of her acts:

> Because acting-for-a-reason consists in an agent's so representing, agents can say, as Anscombe emphasized, what they are doing and why, yet not on the basis of observation or evidence. The ability to do what is to be done as a consequence of another action's being to be done is. . . necessarily self-conscious. [24, 7-8]

what one of those parts—what is described in the *explanandum*—counts as an instance of acting for a reason. In contrast, the non-reductivist and the primitivist both treat the unity represented in reason-giving explanations as a whole whose intelligibility precedes its parts. The primitivist claims that this whole must be taken as primitive, whereas the normativist claims that it should be understood as consisting in an act of representing on the part of the agent.

2.3.1 Primitivist Accounts

This first category ultimately involves a rejection of the claim that we need a philosophical account to provide resources for a response to (MQ) (and by extension—though this is not made explicit—to (EQ) as well). This does not mean that such an account has nothing to say about the the character of reason-giving explanations, or the acts that they represent. But it does abjure any claim to be providing an account of the explanatory unity that characterizes those explanations.

Jonathan Dancy provides a version of such an account in his book *Practical Reality* [13]. The motivation for his account emerges out of a feature of Davidson's proposed response to the Wittgensteinian orthodoxy, and his overall argumentative strategy has parallels with the one attributed to Anscombe and Wittgenstein in our sketch above.

Normative vs. Motivating Reasons

To understand his view, it is helpful to start from Davidson's claim that the relation described in DAVIDSON'S QUESTION is in fact a causal relation between a pair of psychological states and an action. The plausibility of this suggestion stems in part from the fact that every reason-giving explanation has a counterpart that cites the relevant psychological states. For instance, the explanation

[S1:] S turned left because she the sign pointed that way has as a counterpart

[S2:] *S* turned left because she believed that the sign pointed that way.

Indeed, given Davidson's account, it seems that it is actually [S2] that makes the underlying form of reason-giving explanations explicit, since unlike [S1] it explicitly describes one of the psychological states that the explanations represent as causing the action. Further plausibility is given to this idea by the fact that sometimes—if, say, the agent is wrong about the direction the sign points in—it would be much more natural to give an explanation like [S2]. Indeed, we could perhaps *always* give an explanation of this form, whereas we can only give an explanation like [S1] when the agent gets things right.

The last point is sometimes explained in terms of a distinction between *motivating* and *normative* reasons. Appeal to 'normative reasons' is supposed to reflect the justificatory and evaluative dimensions of our talk about actions: in the trivial case we are focusing on, the fact that the sign points left might be called a 'normative reason' insofar as it made turning left the correct response to the sign. In contrast, appeal to 'motivating reasons' is supposed to reflect the explanatory dimension of such talk. In our case, if the sign in fact pointed right, then it would provide a 'normative reason' for turning in that direction. If our agent somehow thought it pointed left, and acted accordingly, there would be no 'normative reason' for her action, since she got things wrong. Nevertheless, we can still explain her action by an appeal to her belief, since it is this that motivated her to turn to the left. Thus, in some cases at least, the notions of normative and motivating reasons seem to come apart.[5]

This distinction between normative and motivating reasons—and the attendant focus on psychological versions of reason-giving explanations—provides essential background for Dancy's treatment of this topic. His guiding idea is that we we need to ensure our philosophical treatment of reason-giving explanations does not have the paradoxical result of

[5] In this case, the 'motivating reason' *would* be a 'normative reason' if it were true, which means that our understanding of the former involves a grasp of the normative relations that would be involved in the latter. As we shall see, some philosophers understand the distinction to encompass cases where the motivating reason is not to be understood in these terms.

undermining the 'idea that it is possible to act for a good reason' [14, 25]. This depends on seeing that the distinction introduced above between 'motivating' and 'normative' reasons "should not be taken to suggest that there are two sorts of reason, the sort that motivate and the sort that are good"[13, 3]. Dancy argues that, if we make the distinction sharper, and assume that these are two distinct kinds of reason, we risk losing sight of the idea that *the reasons for which we act* ought to be—and, in most cases, are—*reasons that speak in favour of the action.* For instance, if we identify motivating reasons with psychological states, and normative reasons with the facts specified by the content of those states, it looks as though the reasons that explain our actions are distinct from the reasons that justify them. This "makes it impossible . . . for the reason why we act to be *among* the reasons in favour of acting" [13, 103], since it makes it impossible for our motivating reasons to be anything other than the fact that we are in particular psychological states. Though we might occasionally justify an action by citing the fact that we belief such-and-such (e.g. the by now familiar example of going to the psychiatrist because of a belief one knows to be delusional), it is not the usual case.[6]

This leads Dancy to claim that the agent's understanding of normative relations between the consideration specified as a reason, and the action it purports to explain, are fundamental to the explanatory character of reason-giving explanations. He takes this to entail that we must maintain a strict identity between normative and motivating reasons, even in cases where the agent is mistaken or confused about her reasons for acting. This means that, on Dancy's understanding, it is a feature of reason-giving explanations that their *explanans* needn't describe something that is actually the case:

> When we give the agent's reasons for doing what he did, the sort of light that is thereby cast on his actions does not seem in any way to require that things

[6] Dancy takes this to rule out a particular version of the causal-psychological view, viz. that the psychological states that such accounts claim reason-giving explanation relies on *are* our reasons for acting. Dancy calls this view *psychologism.* His critique of this view has been broadly embraced in the literature, and since none of the philosophers I will focus on defend any version of it, I will not explore it further.

> should actually have been as he took them to be. For the purposes of these explanations, then, there is no reason to insist that the *explanans* should be the case. [13, 25]

Thus Dancy would claim that the following reason-giving explanation legitimately describes the situation in which our agent thought the sign pointed left, though it in fact pointed to the right:

S3: S turned to the left because the sign pointed to the left.

The *Sui Generis* Character of Reason-Giving Explanations

This provides the basis for the key primitivist claim: that reason-giving explanations appeal to a *sui generis* and non-causal explanatory relation. Dancy begins from a claim about an essential feature of causal explanations, and argues that since reason-giving explanations lack this feature, they cannot be causal:

1. The *explanans* of causal explanations is factive, and so must be true.
2. The *explanans* of reason-giving explanations is non-factive, and so can be false.
3. Therefore, reason-giving explanations are not causal-explanations.

This argument purports to establish that reason-giving explanations cannot be causal explanations. In fact, the second premise notes a feature of reason-giving explanations that marks them off from all other kinds of explanation. This provides the basis for Dancy's second argument:

1. The *explanans* of reason-giving explanations is non-factive.
2. No other kind of explanation has this feature.
3. Therefore reason-giving explanations, and the explanatory relation they represent, are *sui generis*.

Dancy further takes it that the *sui generis* character of reason-giving explanations absolves him of the need to provide a substantive account in response to Davidson's Question:

> The most direct response to Davidson [is] that the difference between those reasons for which the agent did in fact act and those for which he might have acted but did not is not a difference in causal role at all. It is just the difference between the considerations in the light of which he acted and other considerations he took to favour acting as he did but which we not in fact ones in the light of which he decided to do it. [13, 163]

In other words, Dancy is suggesting that we take the relation of *doing A on the grounds that p* as primitive, which would mean that it simply *could not* be explained in other terms, and that Davidson's challenge can be ignored.[7] The relation between a reason and an action, when the reason explains the action by giving the agent's reason for doing what he did, is just that the agent did what he did for that reason. There is nothing further to say by way of explanation of that relation, and so nothing much to be said in response to (MQ), besides the fact that to act for a reason is for one's action to be explainable in this *sui generis* manner.

2.3.2 Reductive Accounts

Dancy's rejection of Davidson's challenge has not satisfied those who would follow Davidson in arguing that reason-giving explanations must depend on appeal to causal relations. Among these are the **causal-psychologists** that belong in our second category of response. Philosophers who provide such accounts aim to explain *what it is to act for a reason*, and, in doing so, to show that the explanatory character of reason-giving explanations depends on causal relations between psychological states and the actions they explain. As such, they

[7] See [33, 145] for this formulation of Dancy's claim.

reject the claim, characteristic of responses in the first category, that *acting for a reason* is a primitive idea that cannot be explained further.

The guiding idea behind such accounts comes from Davidson's work: that what it is to act for a reason is for one's action to be caused by psychological states that specify that reason. So, to return to our simple example, what it is to follow a sign is for one's action to be caused by psychological states that specify the direction of that sign—perhaps *seeing* or *believing that it pointed to the left*—together with a desire to follow it. Thus, an explanation such as

[S1:] *S* turned left because she the sign pointed that way

is true in virtue of causal explanations that are more explicitly articulated in explanations like

[S2:] *S* turned left because she believed that the sign pointed that way

which specify causal relations holding between her belief and her action. In keeping with Dancy's concerns, reductivists do not maintain that *explanans* of (S2) is a description of the person's reasons for acting; rather, it makes explicit that which constituted her acting for a reason, i.e. her action being caused by a state of this sort.

The form taken by the reductive accounts that we will be concerned with involves specifying a notion of causality that is independent of the targets of reason-giving explanations, and claiming that the character of those explanations can be explained by reference to this causality. Our trivial example (S2) involves *psychological states* and *actions* as the terms of the causal relation, and it is an important part of the reductivist account that causal relations of the same sort can be found in other cases as well—for instance, in cases of deviant causation in which the same belief caused a similar action without specifying the agent's reason. According to the reductivist, the causal relation between psychological state and movement is the same in these two cases, which means that the causality in question can be characterized independently of any specifically normative notions associated with normative reasons.

Reductive accounts therefore involve a further claim that Dancy (and others) would reject. For while such accounts are not committed to the claim that our beliefs *are* our reasons for acting, they do claim that *what it is to act for a reason* is to be caused to act by a certain belief. To make this clear, alongside the distinction between 'motivating' and 'normative' reasons discussed above, we can introduce a further idea of an 'explanatory reason', which would apply to the belief that plays a causal role in bringing about the action. The 'explanatory reason' would not be among the agent's reasons for acting, but it would nevertheless play a fundamental role in bringing about her action, and would be implicitly appealed to in any reason-giving explanation. A 'motivating reason' would then be the content of the belief that played the role of 'explanatory reason' for a particular action—and it would be a further question whether the agent must consider the 'motivating reason' to be a 'normative reason' for what she is doing.[8]

The Motivation for Reductive Accounts

As stated above, many of the contemporary philosophers offering such accounts take their lead from Davidson, whose characterization of reason-giving explanation as a kind of causal explanation suggested answers along these lines. However, this brief outline should make it clear that these same philosophers are often motivated by ambitions that go beyond Davidson's treatment of the topic. Whereas Davidson doubted the possibility of providing a reductive account of concepts such as *intention* without appeal to other action concepts—and despaired of spelling out what it was for psychological states to cause an action 'in the

[8] Since the causal relation between belief and action is, in principle at least, independent of any of the normative relations that might be associated with evaluation and justification of actions, it follows that *any* belief can play this causal role, regardless of whether its content is taken by the agent to justify the action in question. Reductivists see this as a virtue of their account: people can act for all kinds of strange reasons, without thinking that those reasons in any way speak in favour of what they are doing. We will return to this aspect of the reductivist's view in §6.

right way'—contemporary philosophers hope to develop Davidson's core ideas in precisely this direction. Kieran Setiya, whose work will be the main target of §6, provides a carefully worked out statement of the motivations behind such accounts: they aim to provide the resources to explain various truths about *what it is to act for a reason.*

These explanatory ambitions relate directly to our orienting questions. Setiya frames these ambitions by considering conditionals of the following form:

> MT1: If A ϕs because$_R$ p, then A ϕs because she believes that p.[9]

According to Setiya, this conditional expresses a necessary truth about *acting for reasons*: whenever a person acts for the reason that p, that person must (at the very least) believe that p and act because of that belief.[10] This is an essential feature of our talk about actions and reasons, one that (we might think) expresses something fundamental about *what it is* to act in this way.[11]

With truths of this kind in view, Setiya motivates his view by way of a "metaphysical edict against brute necessities":

Setiya's Metaphysical Edict: If it is metaphysically necessary that p, the fact that p must be explained by the nature of things; it must follow from what they are.[33, 139]

If we accept this edict, then any necessary truths about *acting for a reason* must be explained by the nature of *what it is to act for a reason.* If the answers to our orienting

[9] See [33, 134]. The subscript appended to 'because' indicates that we are dealing with an instance of reason-giving explanation.

[10] In the best case, a person knows that p, and acts on the basis of that knowledge.

[11] As we shall see in §6, Setiya is also concerned to explain another metaphysical truth that sheds light on our epistemological concerns:

> MT2: When someone is acting intentionally, there must be something he is doing intentionally, not merely trying to do, in the belief that he is doing it.[12]

Setiya claims that explaining this truth will also shed light on how it is that someone who is acting for a reason necessarily knows both what she is doing, and why she is doing it, i.e on our question (EQ).

questions are based on such necessary truths, then we need an account that explains these truths by reference to this nature.

In the case of (**MT1**), the challenge is to explain the sense of the 'because' in the consequent of the conditional, in a way that makes it clear why the truth of the consequent follows directly from the truth of the antecedent. In other words, Setiya is asking why, if a person ϕs for the reason that p, it must necessarily be the case that they ϕed because they believed that p. According to Setiya, this question brings out a pressure to provide an account of *what it is to act for a reason* that goes beyond Davidson's concerns. Setiya, like Davidson, is interested in the explanatory relations expressed by the 'because's of both the antecedent and consequent; but his interest stems from the fact that relations of the sort appealed to in the consequent seem to be a *necessary* consequence of explanations like the one that figures in the antecedent: when we act for a reason, we act because of various psychological states that specify that reason. His thought is that we need a 'metaphysical definition' [33, 139] of what it is to act for a reason that doesn't just clarify these explanatory relations, but that also explains the necessary connections between them.

This puts an additional pressure on Dancy's claim that the relation of *doing A on the grounds that p* should be treated as primitive. What Setiya does here is to show that we can formulate necessary truths about this relationship—in particular, that whenever a person does A on the grounds that p, they do A because they believe that p[13]. He then shows that a causal-psychological account is in a position to explain these truths, whereas an account like Dancy's, which treats the relations involved as primitive, cannot.

This move is characteristic of a broader response to primitivist accounts, i.e. one made by people who do not necessarily share Setiya's metaphysical outlook. The complaint is that in treating the relation of *acting on the grounds that ...* or *acting for a reason* as primitive, Dancy has given up too early, treating something as primitive that can in fact be explained

[13] The particular psychological state doesn't actually matter for Setiya's purposes. All he needs is that whenever someone ϕs for a reason, they ϕ because of some state that specifies that reason. This claim can encompass belief, knowledge, desire, and other candidate states as well.

in independent terms:

> A theory like Dancy's, which does not tell us what it is to act for a reason, must simply postulate that actions can be explained by citing the reasons for which they are done. This fact may be inexplicable. Not everything can be explained. But other things equal, a theory that explains a fact is better than one that leaves it a mystery. Moreover, the need to provide an explanation is especially pressing in the case at hand because of the non-factive nature of motivating reasons, noted above. *How can citing something that is not a fact explain why an action occurred? Why doesn't it matter whether the reasons are true or false?* These are serious questions that must be answered by a satisfactory theory of reasons. The causal theory has an answer. Dancy's theory does not. [18, 78-9]

This is a further sharpening of the challenge posed by Davidson's Question, since it highlights an important feature of reason-giving explanations, one that reflects something about *what it is to act for a reason.* Proponents of reductive accounts maintain that we need to be able to explain these facts, and that Dancy's account is inadequate because it fails to provide the requisite explanation.

The Form of Reductive Accounts

The motivation for a specifically reductive account in response to (MQ) and the metaphysical truths pertaining it goes beyond dissatisfaction with primitivism. To see why, we need to understand a constraint that reductive accounts place on any 'metaphysical definition' of 'what it is to be F' that is to provide a basis for explaining necessary truths:

> A pivotal constraint on explanations of necessity that draw on metaphysical definitions of this kind is *non-circularity.* When the account of what it is to be F has any structure—not just 'to be F is to be G—the materials on the right hand side cannot be metaphysically defined in terms of being F. Otherwise, the

> demand for explanations of necessity would be trivialized. We could explain why it is necessary that all Fs are Gs by this definition: to be F is to be F and G. [33, 139]

This gives us the general form taken by reductive accounts:

> A reductive account of F explains *what it is to be* F in terms that are independent from F.

The 'causal-psychological account' offered as a reductive account of *what it is to act for a reason* hinges on a 'metaphysical definition' that explains "what it is to act on the grounds that p in terms of doing ϕ because one believes that p, in a sense of 'because' that also applies to deviant causation" [33, 146]. Since this explanation depends on a sense of 'because' that applies to cases that are *not* instances of acting on the grounds that p (i.e. cases of deviant causation), it follows that it can be defined independently of 'acting on the grounds that p', meaning that the definition as a whole meets the constraint described above.

Proponents of reductive accounts see only two further options when it comes to answering these questions: taking the idea of *acting for a reason* as primitive (i.e. refusing to provide a metaphysical definition), or relying on some other source of necessity—besides metaphysical definition—to account for the relevant necessary truths. From this perspective, the primitivist accounts considered above fall into the first category of response, failing to provide resources for explaining the relevant necessary truths and thus leaving them a mystery. In contrast, while the non-reductive accounts that we will consider shortly reject the possibility of 'metaphysical definition' (since they do not purport to explain what it is to be F in terms that are independent from F), they at least hold out the promise of providing some other kind of explanation for the relevant necessary truths. For otherwise the circularity at the centre of their accounts will effectively mean they collapse into a version of primitivism.

Thus, whatever the ultimate plausibility of reductive accounts, the concerns presented here raise genuine questions for alternative approaches that belong in our first and third categories. Setiya repeatedly suggests that a reductive account is *required* if we are to provide explanations of the metaphysical truths he cites regarding intentional action, which is what he takes to be involved in answering our orienting questions. The only alternative on the table is to take the terms involved in these questions as 'primitive' or 'given', which makes the topics of these questions 'mysterious', or to provide a non-reductive explanation, whose force is as yet unclear. If the first category of response is to retain its plausibility, it must either embrace the claims about mystery, or demonstrate that it can shed philosophical light on these topics without providing 'explanations'. And if the third category of response is to retains its plausibility, it must show that its circularity does not undermine its explanatory ambitions.[14]

2.3.3 Non-Reductive Accounts

However, this is not to say that reductive accounts do not face their own problems, the most pressing is coming up with a solution to the problem of causal deviance.[15] For, as with Davidson's suggestion, these accounts are open to a version of Anscombe's criticism. In fact, as I shall argue in subsequent chapters, there is reason to believe that this problem is a product of the form taken by reductive accounts, and that no solution can be provided to it on the terms they lay out.

This difficulty serves as an introduction to the third category of response: non-reductive accounts. If the problem of causal deviance stems from the form of reductive accounts, and is

[14] However, it is worth noting that the form of reductive accounts leaves them with an analagous problem. From a narrow perspective focused solely on our topic, they are free to attempt to explain these truths in terms of ideas that are independent from an understanding of *what it is to act for a reason.* However, from a broader perspective, one could presumably press similar questions about any of these independent concepts that they apply in their accounts. The causal-psychological theorist can defer explanation of these notions while his topic is *acting for a reason*, but this deferral cannot last indefinitely. To stay true to his own methodology, the causal-psychological theorist must either provide an account that explains the various further metaphysical truths that might be formulated with these additional concepts, or else admit that they are not subject to such explanations, and must be taken as primitive or given. Of course, a causal-psychological theorist might embrace this position, perhaps with an explanation as to why taking these further concepts as primitive is not problematic: his primary complaint is that proponents of non-reductive accounts of *acting for a reason* embrace it too soon.

[15] Setiya is clear on this point [33]; other causal-psychologists see the problem as less pressing [18].

insurmountable given the resources they have available, a plausible line of response would be to abandon these reductive ambitions altogether. Non-reductive accounts are characterized by the fact that they still aim to provide explanations that shed light on our orienting questions (i.e. they do not simply reject this demand), but want to do this in a way that does not involve articulating what is involved in *acting for a reason* in independent terms.

As we've seen, reductive accounts follow Davidson in claiming that reason-giving explanations are causal explanations, and claim that the kind of causation in question is independent of actions that are subject to such explanations, as it must be if it is to provide the basis for a reductive account. The non-reductive accounts I will focus on agree with their reductive counterparts in the claim that reason-giving explanations are a kind of causal explanation, but disagree with the claim that the kind of causation can be specified without appeal to the normative relations that are reflected in the justification and evaluation of action.

The position I call **normativism**, which is defended by Eric Marcus and Sebastian Rödl, is an example of such an approach, and will be the topic of §7. Marcus agrees with Davidson that reason-giving explanations are a kind of causal explanation, but argues that the causation in question is a special kind, which he names 'rational causation'. Thus, when Marcus comes to explain what it is to act for a reason, he allows himself to appeal to a kind of causation that cannot be specified without an appeal to the idea of acting for a reason, explicitly rejecting the reductivist's constraint on circularity. Furthermore, an important element in his account is the idea that the distinctive character of this causation is reflected in the distinctive character of reason-giving explanations, in a way that helps resolve the fundamental tension we saw in Davidson's view: spelling out what is distinctive about this causation helps us to see how reason-giving explanations can be classified with other kinds of causal explanation, while still differing from them in various fundamental ways.

The contrast between these approaches is perhaps clearest when we look at the problem posed by cases of causal deviance. We have just seen that these cases present a difficult, perhaps insurmountable, challenge to reductive accounts, since they require an account that differentiates them from actions that are subject to reason-giving explanation, while holding on to the idea that the same kind of causation is involved in both. Non-reductive accounts do not face this problem. Taking Marcus as an example, it is a central claim of his account that instances of 'rational causation' of action are constituted by acts of an ability 'to do what is to be done because something else is to be done'.[16] The acts of this ability consist in representings of the normative relations between actions, so that "[t]o act for a reason is to represent the to-be-doneness of an action as following from the to-be-doneness of another action" [24, 7-8].

On this account, any instance of 'rational causation' *already* involves the idea that the action in question is done 'for the sake of the end', since the exercise of this ability will *consist in* an act of the ability to do what is to be done because something else is to be done—or, in other words, pursuit of a particular end as 'to-be-done', and belief that the present action is 'to-be-done' because of this. So whereas, with a reductive account, we can always ask, of any behaviour that involves the relevant kind of causation, 'But was it done for the sake of the end and in view of the thing believed?', with a non-reductive account the question cannot arise at all, since the relevant kind of causation just *is* an action's being done in pursuit of the end and in view of the thing believed, consisting as it does in the act of a capacity to do just that. This means that the 'problem of causal deviance' is no problem at all for these accounts.

Indeed, since the causality involved in reason-giving explanations consists in acts of representing normative relations, it also follows that it cannot be specified independently of those relations. The cases that Marcus focuses on are those where the agent's action is explained by another action to which it is instrumental, i.e. where one action is to-be-done because another action is to-be-done. The causal-explanatory relation involved in the agent's 'acting for a reason' then consists in the agent's representing these normative relations. Indeed, though Marcus does not discuss his view in these terms, this same idea can be applied to the

[16] For this formulation, see [24, 107]

simple cases we have been concerned with as well. For instance, we might say that following a sign consists of representing a particular action as inheriting a normative status from the sign: turning left is to-be-done because the sign points to the left. If the causal relation between the sign and the action consists in an act of representing this normative relation, then the causality in question cannot be specified independently of the relevant normativity.

Problems with Non-Reductive Accounts

Non-reductive accounts aim to distinguish themselves from responses in our first category, both because they claim that reason-giving explanations *are* instances of causal explanation and, more importantly, that we can say something substantive about the nature of this causation, and thus the character of the causal relations represented by reason-giving explanations. Indeed, Marcus' response to Dancy's position is along the same lines as the responses made by proponents of reductive accounts like Davis and Setiya, involving a claim that a non-reductive account can say something substantial at precisely the point at which Dancy's account leaves things mysterious:[17]

> [T]here is plainly a connection between the fact that "S is ϕing because p" entails—indeed gives the same explanation as—"S is ϕing because she believes that p" and the fact that "S is ϕing because p" entails "S believes that p". One would think that a satisfactory explanation of the latter entailment would (at the very least) also shed light on the former. But the "enabling condition" account does not. [24, 110]

But as we've seen, non-reductive accounts necessarily involve a circularity that, from the perspective of reductive accounts, threatens to make them equally explanatorily empty. To better understand this circularity, it is helpful to start with a fuller version of Marcus' description of his ability:

[17] Though, as we shall see, it is actually somewhat unclear how the little that Marcus says on this point differs from Dancy's claim here. See §7

> Rational action-explanations appeal to the exercise of an ability to do what is to be done because something else is to be done. To say that someone is ϕing because she is ψing is to say that she represents the to-be-doneness of ϕing as a consequence of the to-be-doneness of ψing. [24, 107]

We could easily imagine an 'ability' describable with these phrases, that would be part of the aetiology of particular actions, and yet not 'rational' in the sense Marcus intends. Consider the following case:

> Suppose that every Tuesday I have to take out the trash and feed my sourdough starter. Suppose further that, since I am forgetful, I train myself to use the fact that I have to do one of these things as a reminder that I have to do the other. On one particular Tuesday, I remember that I have to take out the trash, and so begin to gather up the bag and take it down stairs. In the middle of doing this, I remember that the fact that I have to take out the trash means that I also have to feed my sourdough starter, and I decide to do that before heading down to the bins.

Here I both represent feeding my sourdough starter as to-be-done because taking the garbage down is to-be-done, and am in fact feeding my starter because I am taking down the trash. My 'ability' to do this has a generality such that it could also be the source of other actions of the same type on other occasions. Of course, the 'because' of the representation, and the causal connection between the two acts, is not the kind that Marcus is concerned with (as is clear from the fact that the two actions are not instrumentally related to each other), and it seems odd to describe my habits of memory as 'abilities'. But a reductivist will push back on precisely these points, insisting that Marcus only avoids the counter-example by characterizing his ability in terms of the relations it is supposed to explain, and adding that the instrumental relations between my actions (or my representations of those relations) are no more obviously ‘acts’ of an ‘ability’ than my habits of memory. The complaint here is two-fold:

(A) The ‘abilities’ to which the non-reductivist appeals play no genuinely explanatory role in the account of reason-giving explanations.

(B) The nature of the ‘acts’ of these abilities, and the causal-explanatory role they are supposed to play, are both mysterious.

(A) suggests that non-reductivists provide an account that is explanatorily empty, insofar as it appeals to the notions it purports to explain. This follows directly from the fact that non-reductivists avoid the problem of causal deviance by providing an account of reason-giving explanations that makes explanatory appeal to normative relations involved in the acts represented in those explanations—e.g., by defining the causality involved in reason-giving explanations in terms that explicitly involve the idea of intelligent pursuit of an end. The apparent circularity of such accounts should thus be clear: asked to explain the character of an end-specifying explanation of the form,

S is doing A because she is doing B,

they respond that explanations of this sort work by appeal to an ability to do what is to-be-done because something else is to-be-done. Or asked to explain explanations like our (S1),

S1: *S* turned left because the sign pointed that way,

they respond that explanations of this sort work by an appeal to an ability to do what is to-be-done because a sign is to-be-followed. This amounts to saying that such explanations work by appealing to abilities to do the kinds of thing the explanation explains: we explain acts of sign-following by an ability to follow signs, or acts that are instrumentally-related by an ability to do what is to be done because something else is to be done. From the reductivist's perspective, this seems explanatorily empty. Thus, until we have a clearer understanding of why a non-circular account is impossible, and of how a circular account might be genuinely explanatory (or otherwise responsive to our philosophical puzzles), any proponent of a reductive account could be forgiven for continuing to look for a non-circular solution to the problem of causal deviance for lack of a viable alternative.

(B) raises a more specific concern about the way the non-reductivist ensures a close relation between normative and explanatory relations. As we've seen, they do this by claiming that the causality represented in reason-giving explanations *consists in* acts of representing normative relations: to do A because one is doing B *is* to represent doing A as to-be-done because doing B is to-be-done. The reductivist might reasonably respond to this with puzzlement about the nature of this 'act': how can a *representing of normative relations* constitute a genuine causal-explanatory relationship between what is represented? Surely causal relations, and our representations of those relations, are two distinct things? If the non-reductive account is to be a viable candidate, it must clarify the nature of these 'rational abilities', and the role that representation plays in them.

Together, (A) and (B) gives us a clear picture of the requirements that must be met by any non-reductive account of reason-giving explanations. Besides showing that a reductive account is not available, it must also clarify the specific details of its claims, and show how they amount to an explanatory response to our orienting questions. The danger here is that non-reductive accounts will simply collapse into a version of Dancy's position: the explanatory relation at the centre of reason-giving explanations 'is what it is, and not another thing' [13, 163].

2.3.4 Summary

As I hope to have shown in this survey, all three categories of account agree in aiming to provide a response to our metaphysical question (MQ) via an account of reason-giving explanations. The primitivist response is to assert that reason-giving explanations depend on a *sui generis* non-causal relation, and that this ultimately means that our response to (MQ) must be fairly minimal: to act for a reason is for one's reason to be related to one's action in this manner. The reductivist and the non-reductivist response both reject primitivism, and assert that we can provide a more substantive answer to (MQ). Causal-psychologists and normativists agree that reason-giving explanations depend on causal relations, and that *what it is to act for a reason* should be explained by an account of those relations. But they disagree insofar as the causal-psychologist claims that these causal relations must be explained in independent terms, whereas the normativist rejects this constraint.[18]

2.4 A Substantive Response to (MQ) and (EQ)

While the primitivist, reductivist, and non-reductivist provide different specific responses to (MQ) and (EQ), they appear to agree in providing what I shall call a 'substantive response' to (MQ) and (EQ). Although I shall ultimately provide some grounds for questioning whether anyone besides the reductivist *means* to be providing a substantive response, in this section I shall argue that the impression that this is a common goal is generated by a shared conception of what those questions demand from us, and of what we must do to resolve the puzzles from which we started. This conception can be traced to the role that 'reason-giving explanations' play in framing (MQ) and (EQ). The outline I provide here will be essential background for the alternative Wittgensteinian approach that I will begin to develop in the next chapter.

2.4.1 The Need for a Substantive Response

A 'substantive response' is one that specifies *in virtue of what* something counts (and is known) as an instance of 'acting for a reason'. The apparent need for a response of this form

[18] The causal-psychologist and normativist accounts we are concerned with also purport to give a response to (EQ) that follows directly from their account of *what it is to act for a reason*. Kieran Setiya argues that acting for a reason involves causation by a self-referential psychological state, which entails that the agent has knowledge of the relevant causation just insofar as she is in that state. Eric Marcus argues that the causation involved in acting for a reason consists of an act of representing on the part of the agent, and adds that this act of representing explains one's knowledge of this causation

can be traced to the way in which (MQ) and (EQ) are approached through an investigation into 'reason-giving explanations'.[19] The central question that all of the accounts in this chapter seek to answer can be posed by asking what difference there might be between explanations such as

S1: S turned left because$_R$ the sign pointed that way,

where S's act was an act of sign-following, and the *explanans* gave her reason for that act, and an apparently identical explanation

S2: S turned left because$_C$ the sign pointed that way,

where S's act was not an act of sign-following, but was caused by the sign in some other way. This quickly became a general puzzle about *any* instance of reason-giving explanation. Whenever we have an explanation of the form,

RA: S ϕed because$_R$ _____,

where the subscript $_R$ marks the fact that the *explanans* gives the agent's reason for her act, we can always ask ourselves what differentiates these explanations from apparently similar ones of the form,

NR: S ϕed because _____,

where the *explanans* does not give the agent's reason. For, as we saw from Davidson's climber-case and our example of sign-following, the descriptions of the *explanans* and the *explanandum* could be identical in both cases.

(MQ) is therefore understood as asking for what differentiates explanations that we would gather under (RA) from explanations that we would gather under (NR). This framing explains several common features shared by primitivist, reductivist, and non-reductivist

[19] Ultimately, I shall argue that they can be traced back further, to the way in which our initial puzzles arose from 'zooming-in' on particular acts. See §5

accounts. First, all of them understand (MQ) and (EQ) as calling for a response that is pitched a high level of generality: they assume that it will be possible to say, in general and (for the reductivist and non-reductivist) informative terms, what it is to act for a reason, and how it is that the agent comes to have knowledge of that act. This is because all three kinds of account are framed to cover *any* case in which we can offer an explanation of the form,

S ϕed because _____,

where the *explanans* gives S's reason for ϕing, and S can offer such an explanation just insofar as she ϕs for this reason. This in turn presupposes that 'acting for a reason' is one thing – or at least, that it is amenable to a general account that could be applied to any instance of reason-giving explanation.

Secondly, each response abstracts away from any local differences between particular instances of 'acting for a reason', or even between particular kinds of act, to provide an account with the right level of generality. From such a perspective, a substantive characterization of the particular act described in the *explanandum*—as, say, a handstand, or an act of reading—is irrelevant to the questions at hand. This is again because various determinate explanations can be thought of as gathered together under the general notion of reason-giving explanation, which is captured through our schemata (RA). What then seems to be required is a way of characterizing *any* instance that fits this schemata—and thus any instance of 'acting for a reason'—in a way that makes explicit why it counts as such an act, and how it is known as such by the agent.

Thirdly—and most importantly—approaching these questions by way of reason-giving explanations suggests that the basis for a response will lie in what is immediately described by those explanations, i.e. the specific act and its relation to whatever is described in the *explanans*. Somewhat paradoxically, this means that despite abstracting away from the distinctive features of particular acts, responses to (MQ) and (EQ) nevertheless centre around abstract characterizations of the specific act and its aetiology. To put this point another way, the implicit assumption behind this approach is that if we are asking why something counts as a reason-giving explanation, and how it is that the agent comes to be in a position to provide such an explanation, the answer must lie directly in the happenings that explanation describes: in what this particular person is doing here and now.[20]

It is this framing that produces the pressure to provide a substantive response to (MQ). For it now seems that in order to explain the difference between explanations that we would gather under (RA), and superficially similar explanations that we would gather under (NR), we need to specify something about what is represented by explanations in the first group that grounds our counting them as instances of reason-giving explanation. Since we look for this 'something' in what is immediately represented by the explanation, it makes sense to locate that something in what is represented by one of the terms of the explanations: the *explanans*, the *explanandum*, or the explanatory relation between them expressed by the 'because'. Moreover, since this 'something' needs to be there in *any* instance that fits the generic schemata (RA), it must be some quite generic feature of what is represented by such explanations, one that can be described in abstraction from the specifics of the act described in the *explanandum* or the reason described in the *explanans*.

2.4.2 Contemporary Accounts as Providing a Substantive Response

Such a response could then be summarised in a statement of the form,

> To act for a reason is . . .

where the '. . . ' are replaced by a specification of that something *in virtue of which* some behaviour counts as an instance of acting for a reason.[21]

This form is clearest in reductivist accounts such as causal-psychologism, which explicitly aim to identify some independently specifiable condition present in all and only instances

[20] Cf. my epigraph from *Zettel* §567.

[21] It is possible that what is specified will provide grounds of the agent's first-personal knowledge of her acts, and thus the basis for a response to (EQ). See below.

of acting for a reason. Thus a summary characterization of the reductivist response would state:

> To act for a reason is for one's action to be caused (in the right way) by a particular kind of psychological state, or set of such states.

This entails that explanations are gathered under (RA) in virtue of representing happenings that could be also described by explanations of the following form:

> S is doing A because she is in psychological state σ,

where the content of σ involves whatever is represented in the *explanans* of the original explanation.

Non-reductivist and primitivist accounts can also be summarized to suggest that their account takes the same form. Thus the normativist maintains that,

> To act for a reason is for one's act to be caused by (or an instance of) rational causation,

whereas the primitivist would say

> To act for a reason is for one's act to be explainable by a *sui generis* non-causal relation.

In both cases, explanations are gathered under (RA) in virtue of the distinctive relationship represented by the 'because' (i.e. rational causation, or some other *sui generis* relation).

On this reading, though each account has its own views both on the character of that something *in virtue of which* an act counts as an instance of 'acting for a reason', and on what is required for a philosophically helpful specification of it, they all agree that the basic form of a response to (MQ) must involve specifying *something* present in all and only instances of acting for a reason.

The reductivist and the non-reductivist agree that we can provide a positive and informative characterization of this *something*, and direct their respective accounts to this end. The work of the causal-psychological account lies in specifying the relevant kind of psychological state(s) σ, and showing that it is present in all and only instances of acting for a reason. The work of normativist accounts lies in characterizing what is involved in the 'rational causation' represented by the 'because', and explaining what it is for an act to be caused in this way. In contrast, the primitivist account rejects the idea that we can provide a positive and informative characterization of this *something*, resting content with a negative characterization of a *sui generis* non-causal explanatory relation.

Building on this general form of response, an account can use its response to (MQ) as the basis for a substantive response to (EQ). Such a response would be one that explains *how (i.e. on what grounds)* the agent comes to have first-personal knowledge of her acts expressed in her explanations of her actions. This involves specifying some ground that justifies the agent's knowledge of these explanations, and is present just insofar as she acts in a particular way. If that something *in virtue of which* an act counts as an instance of acting for a reason is epistemically available to the agent just insofar as she acts, then it could also provide this ground for her knowledge.

Here too, both causal-psychologism and normativism appear to agree in the basic form of their response, while disagreeing about the specifics. For Kieran Setiya, our grounds for our knowledge of our acts comes from the special psychological state that causes them. This state is a self-referential desire-like belief, which has the subject both represent herself as acting for a particular reason, and by the same token causes her to act for this reason. Since this state is always present when we act for a reason, the agent always has grounds to explain her knowledge of what she is doing and why. In a sense, our knowledge of our acts is self-grounding, since the state in virtue of which the act counts as an instance of acting for a reason is also the state that provides the grounds for our knowledge of it as such: as Setiya puts it, "in acting for reasons, we have beliefs about the psychological explanation of our actions" [32, 47].

For Eric Marcus, both acting intentionally and acting for a reason are to be identified with acts of representing the relevant action as having a particular normative status. Since to act *is* to represent one's action as to-be-done,[22] this representing could be understood as the grounds for our knowledge of what we are doing and why. As with Setiya's account, there is a sense in which our knowledge of our acts is self-grounding, since the 'act of representing' in virtue of which the act counts as an instance of acting for a reason is also the 'act of representing' that provides grounds for (or perhaps is itself) our knowledge of it as such.

2.4.3 An Alternative Approach?

In the previous section I suggested that both the primitivist and the non-reductivist might appear to be providing a 'substantive response' to (MQ) and (EQ). But on closer inspection, certain aspects of their views seem to undermine this reading. This was brought out in the reductivist's criticisms discussed in §2: from the perspective of the reductivist, both forms of account seem to involve a form of circularity that undermines the impression that they are giving a substantive response to our questions.

This is clearest in primitivist accounts. Above I described them as having the form,

> To act for a reason is for one's act to be explainable by a *sui generis* non-causal relation.

But their view might alternately be summarized as follows:

> To act for a reason is for one's act to be explainable by a reason-giving explanation,

since, as we said above, the only positive characterization of the *sui generis* relation that they provide amounts to saying that the relation between a reason and an action, when the

[22] At least, this is view I take Marcus to want to defend, though his view actually has one represent the action-concept (i.e. something general) rather than one's particular action (something particular) as to-be-done. As we shall see in §7, this points towards a fundamental difficulty with his view.

reason explains the action by giving the agent's reason for doing what he did, is just that the agent did what he did for that reason. Analogous concerns were raised about the normativist response as well: from the perspective of the reductivist, the circularity at the heart of their account seemed to render that account explanatorily empty (see §2.3.3 above).

This undermines the claim that the primitivist and the non-reductivist aim to be providing a substantive response to (MQ) and (EQ). For, on this reading, the 'content' of this response says nothing more than the following:

> To act for a reason is for one's act to be an instance of acting for a reason.

Although its surface form is the same as the causal-psychological response, this is not a substantive response to (MQ), since it explicitly fails to answer the question.[23] Indeed, it suggests that once we reject reductivism, no substantive response to (MQ) will be possible—any completely general 'account' we might provide in response to the question 'what is it to act for a reason?' will have the form of a tautology.

This thought provides important background to the alternative Wittgensteinian approach I shall outline in the next chapter. For it should lead us to ask the following questions:

1. If we cannot provide a substantive response to questions like (MQ) and (EQ), how can we resolve the puzzles that led to them? Is there an alternative approach to these puzzles, and will it provide us with the basis for a response to (MQ) and (EQ)?

2. If there is an alternative approach to these questions, are primitivism and normativism best understood as attempts to realize it? If so, are they successful?

I shall begin to suggest an answer to (1) in the next chapter; a full answer to (2) will only be possible at the end of this dissertation.

[23] For comparison, imagine you asked me the way to the library, and I replied that 'the right way to go is the one that leads to the library'. Such a response fails to engage directly with your question: it is either a riddle, a joke, a refusal to engage with you, etc.

CHAPTER 3

A WITTGENSTEINIAN APPROACH

In this chapter I provide an initial sketch of an alternative Wittgensteinian approach to our topic. As well as showing what is involved in this approach, I shall be concerned to show how it gives us a different perspective on questions such as (MQ) and (EQ). The accounts we surveyed in the previous chapter agreed that these questions demanded an account that could differentiate reason-giving explanations from other forms of explanation. They further maintained that any such account should be pitched in general terms so that it applied to any instance of reason-giving explanation, and that it should work by specifying something about what was immediately represented by those explanations *in virtue of which* they count as reason-giving explanations.

The approach I shall outline in this chapter rejects all these points. First, it sees the pressure to provide an account of this form as emerging from philosophical confusions. Second, it rejects the possibility of a substantive response to (MQ) and (EQ), and seeks to provide an alternative approach to characterizing the difference between reason-giving explanations and other forms of explanation. Finally, it maintains that philosophically useful characterizations of this difference needn't be pitched in general terms—i.e. as immediately applying to *all* instances of reason-giving explanation—nor work by specifying something about what is immediately represented by such explanations.

The chapter falls into three parts. In §3.1 I show how Anscombe's method in her monograph *Intention* provides a model for a different approach to questions such as (MQ). In §3.2, I connect this method to Anscombe's discussions of various ideas from Wittgenstein, and show how these provide a different perspective on questions such as (MQ) and the demands they place on us. Finally, in §3.3, I draw outline some general consequences for our discussion of (MQ) and (EQ) in subsequent chapters.

3.1 Grammar and Philosophy

3.1.1 Anscombe's Method in Intention

One of the central questions framing Anscombe's investigation in *Intention* is "[w]hat distinguishes actions which are intentional from those that are not?" (§5). This question has the same form as the puzzle we are started with – indeed, the impression that Anscombe provides the basis for a substantive response to (MQ) arises from the fact that she seems to provide a substantive response to this question that directly implicates (MQ):

> The answer I shall suggest is that they are 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. §5

As we saw in the previous chapter, this gives the impression that Anscombe is proposing that what distinguishes actions which are intentional from those which are not is that the former are explained by a reason for acting. This seems to give extra importance to (MQ), since it now appears that whatever makes an action count as an instance of 'acting for a reason' will also be what makes it an instance of intentional action.

That Anscombe is *not* proposing a response along these lines can be seen from the fact that explicitly disavows any ambitions to provide a substantive account of her topic, as is clear from her summary description of §19: "an action is not called 'intentional' in virtue of any extra feature which exists when it is performed". This should be heard as a rejection of the whole notion of a substantive response – that is, of any account that explains what distinguishes actions that are intentional from those that are not by reference to some mark or feature *in virtue of which* the former count as intentional. Given this general claim, it cannot be right to think that Anscombe is proposing that an action count as intentional *in virtue of* being explained by a reason for acting. As Anton Ford puts it, "Anscombe denies that an action is intentional in virtue of being caused by a reason because she denies that an action is intentional in virtue of standing in any relation to a reason, or in virtue of standing in any relation to anything, or in virtue of having any property whatsoever" [21, 140].

Looking at how Anscombe in fact treats her question will therefore prove instructive in outlining an alternative approach to problems such as the ones framing our topic. What follows is a quick and partial summary of those parts of her argument necessary to understand her own account of what distinguishes actions that are intentional from those that are not.

Anscombe's Account

Anscombe states that to clarify "the proposed account" she will "both explain this sense [of the question why?'] and describe cases shewing the question *not* to have application" (§6). She does this by laying out various uses of language:[1] on the one hand, responses to the question 'why?', asked in the relevant sense, that show the question has been rejected; and on the other, examples of a question 'why?' whose sense is different from the one she is concerned with.[2]

Once she has isolated the relevant sense of the question, she turns to discussion of a particular example, and proceeds to show that repeated application of her question 'why?' yields a series of descriptions of an action – in her example, we get four descriptions of the action of a gardener pumping water, which "form a series, A—B—C—D, in which each description is introduced as dependent on the previous one, though independent from the following one" (§26). Each description in the series expresses an intention with which previous acts were undertaken: the gardener is moving his hand up and down because he is operating the pump, and he is operating the pump because he is replenishing the water supply, etc.

[1] Anscombe says that she will complete the task of elucidating the relevant sense of the question 'why?' in three stages. These can be broken down as follows. In SS6-8 she describes various cases in which the question is rejected; in SS9-15 she describes different possible senses of the question 'why?'; in §16 she summarises her results so far, before going on to describe further cases in which the question is rejected, along with cases in which the question is accepted but no positive answer is provided, in SS17-18.

[2] Among these is one that asks for what she calls the 'mental cause' of her action. We will return to this sense when we consider her discussions of reasons and causes in the next chapter.

The A-D order that we come to see in this particular example is one that can be seen in other intentional actions as well. Wherever we can say something of the form,

S is doing A because she is doing B,[3].

we are describing something with the order Anscombe has characterized.

Following her discussion of this example and the A-D order that can be see in it, further reflection on practical reasoning—and, in particular, on practical syllogisms—reveals that one use for such syllogisms lies in illuminating an action's place in an A-D order by showing the good or point of that action relative to that order. Thus, if I have an intention to do D, a syllogism of the form,

Doing C is a way of doing D
To do C, do B
I can do B by doing A
Here's an opportunity to do A!

reveals the good or point of my doing A, B, and C, relative to my intention to do D. The premises of the syllogism articulate considerations that could be given in response to Anscombe's question 'why?'.

At a pivotal moment towards the end of the book, Anscombe concludes that "the term 'intentional' has reference to a *form* of description of events" and that "[w]hat is essential to this form is displayed by the results of our enquiries into the question 'Why?"'. I take it that this is intended as a summary of Anscombe's answer to her question "what distinguishes actions that are intentional from those that are not?". Its force can be seen in her discussion

[3] Or related forms such as,

S is doing A because she wants to do B

or

S is doing A because she is trying to do B

See [36, 99]

of action-concepts used in the description of intentional actions. That a description has the form is shown by the fact that it is “formally characterized as subject to our question ‘Why?’, whose application displays the A-D order we discovered” (§48). Indeed, “[e]vents are typically described in this form when ‘in order to’ or ‘because’ (in one sense) is attached to their descriptions”, i.e. when they are represented as explainable by reference to their place in an A-D order.

It follows from this that a particular description, e.g. ‘sliding on ice’, will sometimes be used of actions that are intentional, but sometimes not. Thus, a description such as

> James slid across the ice,

could be used on an occasion where it would also be true to say,

> [I:] James slid across the ice to get his coat,

or on a different occasion on which it would be true to say,

> [NI:] James slid across the ice because Jill pushed him.

In the first situation, the description has the relevant form – it is subject to the relevant sense of the question ‘why?’, and one could lay out James’ grounds for his action in a syllogism that showed the point of his action, such as,

> My coat is on the other side of the lake
> I need to get over there to retreive
> Since the lake is frozen, I can slide across it

Part of what it is to understand a description of this form is to see that it makes sense to ask ‘why did you do such-and-such?’ in response to it; or again to see that James could respond by specifying some further end, or various considerations to show the good or point he saw in it – considerations that could play the role of a premise in a practical syllogism, or form the basis for an evaluation of James’ action.

This contrasts with the role of 'James slid across the ice' in (NI). Here, if one understood the original description, it could only be a joke to ask for James' reasons for sliding across the ice, or to demand that he provide considerations that showed what good or point he saw in it. The description is subject to *a* sense of the question 'why?', but here the answer merely states the cause of the behaviour, and does not purport to specify something James took to be a reason for acting so.

Note that the form of a description needn't be clear from the phrase itself considered in abstraction—for there is a sense in which the same phrase is applied to both cases. What matters is how the phrase is used. That 'James slid on the ice' is, on a particular use, a description of an intentional action, is shown by e.g. the further things it makes sense to say in response to it. This might seem to be a purely linguistic point – but it in fact shows us something about the intentional action described by the phrase. For if we focus in on some particular such action—a snapshot moment of James sliding across the ice, say—there need be nothing about what happens in that moment that shows that James' action is intentional. The further things it makes sense to say in relation to the original description that show that we are talking about an intentional action are mostly not at all concerned with that particular moment. For instance, if one were asked why James slid across the ice, what one would describe might be what James went on to do once he reached the other side of the lake (retrieved his coat), or why a person (any person) might decide to slide across a lake rather than walk around it, etc. That James' action was intentional is shown in these further things it makes sense to say about it – but that it makes sense to say these things can't be seen if we zoom-in on the action itself, and consider it in isolation from its circumstances, since this thought and talk is concerned with more than that isolated movement.

Anscombe understands these points to have crucial implications for what it is to provide an account of the term 'intentional'. For "[i]f one simply attends to the fact that many actions can be either intentional or unintentional, it can be quite natural to think that events which are characterisable as intentional or unintentional are a certain natural class, ‘intentional’ being an extra property which a philosopher has to describe” (§47). This is the urge that makes us zoom in on the act itself, and look for some mark or feature *in virtue of which* it is correct to talk about it in the ways described above. But if one understands the claim that the term ‘intentional’ has reference to a ‘*form* of description’, one will give up on trying to identify something *in virtue of which* particular events count as intentional actions, and instead seek to provide another kind of account.

Indeed, Anscombe’s comments suggests that she takes herself to have provided some such account over the course of her book. This has involved showing that a mark of the description of intentional action is to be formally subject to her question ‘Why?’, and that a person acting intentionally knows their action under descriptions of this form; that the application of this question reveals the A-D order that represents actions as parts of teleologically-articulated wholes; that this order is the same order as is described in giving an account of the practical syllogism, i.e. representations of fragments of practical reasoning concerned with calculation of how to achieve a specific end; and finally, that this is “an order that is there whenever actions are done with intentions” (§42). Indeed since, in acting intentionally, a person knows their actions under a descriptions of the form Anscombe has identified, they represent them in terms of the order that characterizes practical reasoning, meaning that their knowledge of it is knowledge of it as an act of reason.

3.2 Philosophical Grammar

To understand the character of this ‘account’, it is helpful to set it alongside some of Anscombe’s discussions of Wittgenstein’s work – particularly of his conception of ‘grammar’. Anscombe reports that, far from denoting some technical or esoteric notion, Wittgenstein insisted that what he had in mind when he spoke of grammar was “what you heard lessons in at school”, i.e. grammar in a familiar sense.[4] An example of this appears in §137 of the

[4] See [7, 133]; also e.g. [9, 200]. The development of linguistics complicates this claim, as Anscombe notes:

Philosophical Investigations, where Wittgenstein describes learning to determine the subject of a sentence by means of the question "Who or what . . . ?". Anscombe takes up this example in several of her papers:

> In grammar we acquire the concepts of subjects and objects of verbs by being presented with sentences containing verbs, say 'He ran', and being trained to answer the question 'Who or what ran?' by repeating the word 'He'. Or, given the sentence 'John kicked Jim' and the question 'Whom did John kick?' we answer, 'Jim'. [2, 96]

Echoing the discussion in *Philosophical Investigations*, she notes that we may compare learning to answer this sort of question to learning to answer the question 'What comes after the letter L in the alphabet?' – only, she adds, "there is more intelligence in it":

> [F]or it is not a matter of merely repeating, say, the whole of what precedes the verb in a sentence or the whole of what succeeds it. Thus, given "Smith is a ninny and James is a dolt", and asked "Who or what is a dolt?" the learner answers "James", not "Smith is a ninny and James", But, given "The man next to James is a dolt" he answers not "James" but "The man next to James". [2, 97]

Mastering grammatical concepts such as 'subject' or 'direct object' thus involves a certain kind of reflective perspective on our understanding of language. The capacity to answer the

> The difference of opinion about what belongs to grammar arises from belief in and practice of a 'formal' science of grammar on the one hand, and the study of what a given use of words amounts to or achieves on the other. The former belief leads to an examination of the ways words occur together and an attempt to formulate rules and explanations of this, always in terms of purely linguistic structures. The latter leads to consideration of contrasts between say 'For how long did you forget that' and 'For how long did you reflect on that?' or between intermission of intention and intermission of attention. There is nothing obscure about calling 'grammatical' the observation of the different temporalities involved in these cases. but it is not a kind of observation that we expect from formal grammarians. [9, 201]

No doubt modern linguists might reject her claim that interest in 'purely linguistic structures' cannot encompass interest in the kind of differences she outlines in the second half of her paragraph.

question ‘who or what is a dolt?’ described here has its basis in our capacity to understand the original sentence. It can also be further developed into a capacity to answer such questions as ‘who or what is the subject of this sentence?’. We therefore display the basis for the acquisition of such grammatical concepts just insofar as we show ourselves to understand what is said, or what it makes sense to say.

An example like the one above gives us very general grammatical categories like ‘subject’ and ‘direct object’, but these admit of further specification in each particular case. For instance, in the sentence ‘John kicked Jim’, both subject and direct object are also *proper names*—indeed, *names of persons*[5]—with the former denoting the *agent* and the latter the *patient* of the *causal transaction* described by the verb and represented by the sentence as a whole. As I understand it, Wittgenstein’s conception of grammar would count all of these as grammatical categories, along with a considerable range of further determinations within them.[6]

Our capacity to understand what people say, along with our sense of what it makes to say in response to them, show that we in some sense grasp the grammar of what is said (even if we do not deploy the ‘reflective’ grammatical concepts named in the previous paragraph). Thus, on being told that “John kicked James”, we know that this is the kind of thing to which one might ask “And why did he do that?”, or “And was it on purpose or by accident?”, or perhaps “And how did Jim react?” or “And did he apologize?”. The applicability of these questions reflects the kind of thing we are talking about – to see this, we might compare it

[5] Anscombe reports Wittgenstein remarking: ‘It is a great deal of information about a word that it is a proper name, and still more, what kind of thing it is a proper name of – a man, a battle, a place, etc., etc.’ [8, 162]

[6] Anscombe also expresses the thought that Wittgenstein would recognize a wide range of categories, and admit further categorial differences within particular kinds:

> If ‘proper name’ is a grammatical category, then so in his conception is ‘numeral’ and so is ‘colour-name’ and so is ‘psychological verb’. But by Wittgensteinian considerations even all of these turn out to be somewhat generic: that is, there are ‘categorial’ differences within each kind. [9, 201]

with a sentence of the same general form[7] to which none of these questions are applicable, e.g. "The tree struck the window". If someone responded to this by asking "And was it on purpose or by accident?", or "And did it apologize?", we would take what they said as e.g. a joke, or suppose that they had not understood what had been said.

If one thinks of 'intentional action' as a 'grammatical concept' in this sense, then one can see Anscombe's investigations in *Intention* as a contribution to clarifying the grammar of the uses of language to which it applies: descriptions of actions that are intentional. It belongs to such a description that it have the form Anscombe has characterized, which means e.g. that it can be a term in either the *explanans* or *explanandum* of an explanation representing part of an A-D order,

S is doing A because she is doing B,

or can be the starting point or conclusion of a practical syllogism that shows the good or point of an action relative to its place in that order.[8]

We have a pre-reflective mastery of this grammar just insofar as we can understand talk about intentional actions – understand what it makes sense to ask about them, or understand how certain statements can provide justifications, explanations, evaluations for them, etc. For instance, even if I rarely say something of the form,

S is doing A because she is doing B,

I understand that it makes sense to ask of someone 'why are you doing A?', and further understand that 'I'm doing B' could be an intelligible response. Likewise, I needn't ever formulate (or even have the concept of) a practical syllogism, nor be able to articulate the idea that what the premises of such a syllogism do is show the good or point of a particular

[7] Ignoring the use of a proper name; for a direct comparison, replace "John kicked James" with "The president kicked James".

[8] If there is such a thing as primitive actions, then they could only figure in the *explanans* of such explanations. But descriptions of them are still characterized by the fact that they describe the kind of thing that can play this role.

action, relative to some end. But I still understand what is said when, on asking 'why are you doing A?', someone responds with a consideration that could play a role in such a syllogism.

Anscombe's investigation aims to give us a reflective perspective on this mastery, and with it a reflective perspective on what it is to act intentionally. The discussion of particular examples, and the discussion of the practical syllogism, are complementary ways of doing this. For through the discussion of the example of the gardener, we come obtain a reflective perspective on the A-D order, which we can then see explicitly in other kinds of case; we can further see that certain kinds of reason-giving explanations provide a form of representation of this order. Again, through the discussion of the practical syllogism, we come to see that such syllogisms can specify possible grounds for an action that show its good or point relative to some part of an A-D order; and we can see such syllogisms as formalizing part of our ordinary thought and talk about such actions.

3.2.1 Metaphysics and Grammar

The attention to what it makes sense to say that characterizes this approach—i.e. attention to the ways in which the use of a word might enter into conversation and discussion—indicates that what is meant by 'grammar' is not restricted to the ways words combine with each other in particular sentences.[9] This should already have been clear from the discussion of 'James slid on the ice' above – for the phrase could figure in identical sentences, only one of which would describe an intentional action.

The 'grammar' we are concerned with is to be seen in the use of language – the way that words find application and make a difference in the lives of people who can speak. This is

[9] Here is Rush Rhees on this point:

> When one speaks of the use of a word . . . one may primarily mean the words with which it is connected and the sort of connexions it has with them in sentences. On the other hand, one may mean what is done with these sentences, how they enter into actual conversations and discussions, the part they play in the activities of the people who do use them. It is clear that the use of a word, in this latter sense of the expression, might be very different in two connexions even though there were little difference in what one might call the external appearance of the sentence in each case. [27, 104]

why description of a particular intentional action involved circumstances broader than the immediate bodily movements that constituted that action. The uses of language that belong to our thought and talk about intentional action depend on features of our lives that go far beyond any one such moment.

Clarification of the grammar of particular terms—say, proper names of people, or descriptions of intentional actions—thus involves attentiveness to how words that have this role are used in human life. Understanding this point can help mitigate that impression that, in talking about 'grammar' and language, we are changing the subject and ignoring the metaphysical questions from which we started. For it is an important part of Wittgenstein's later philosophy that in describing particular cases to which we might apply our grammatical concepts—say, 'proper name of a person'—we are describing lives in which there are e.g. *people.* This is the point of Wittgenstein's statement at §371 of the *Philosophical Investigations*: *essence* is expressed by grammar. We therefore approach 'metaphysical questions' of the form 'what is it to be X?' by looking at what we would call thought and talk about X, and trying to characterize the grammar of such thought and talk.[10] This is the sense in which Anscombe's investigation provides an answer to the question 'what distinguishes

[10] One way of getting clear about the various aspects of the grammar of this thought and talk is through the description of simple language-games that emphasize particular aspects of it. For instance, a simple language-game in which players are to raise their hand when their name is spoken might draw our attention to one aspect of our use of proper names; another game in which players are to name people who are absent could draw our attention to another. Our mastery of the grammar of such names is manifested not only by our capacity to respond to descriptions that deploy them, but also in our understanding of the variety of uses that names have in our lives.

Reflective attention to how names are used—through attention to what we want to count under the grammatical category 'name'. or through attention to the various uses of words that fall in this category—develops a kind of clarity about the grammar of names that takes us beyond our pre-reflective mastery. For instance, such reflection reveals that what we count under a particular grammatical category can encompass a motley of different related uses, all of which are a part of what it is to be e.g. a proper name of a person. Thus, names of people are used in all kinds of descriptions of people and their activities and attributes, including descriptions of people who are long dead and with whom we have had no interactions (cf. PI §79, §39-40, etc). They are also used to call or summon people, to address them when they are present, and to invoke them when they are absent (including when they are dead). All of this is a part of what it is to be a name of a person, and thus (on the reading I am proposing) indicative of *what it is to be a name of a person* (and therefore of *what it is to be a person*).

We give names to other kinds of object as well, e.g. animals, toys, places, etc. In each case, our use of names might inherit some of the grammar of proper names, though it needn't inherit all of it. Which parts of the grammar it inherits, and how it inherits them, will reflect what it is we are talking about.

actions from intentional from those that are not' – by helping us see the differences in the grammar of our thought and talk about intentional actions and other kinds of behaviour.

Nevertheless, from the perspective of someone seeking a substantive response to questions like (MQ) and (EQ), this Wittgensteinian approach will appear disappointing. For to them it must seem that outlining the grammar of our thought and talk about a particular subject can only be preparatory for the real philosophical work. They do not deny that we mark a difference between, say, intentional action and other kinds of behaviour. But they want to know what grounds or justifies our drawing that distinction. Outlining the character of our thought and talk, or describing the place that it has in our lives, looks like drawing our attention to exactly what needs to be justified via a substantive account!

If we are to understand the role that questions such as (MQ) and (EQ) play in framing contemporary discussion, we need to understand where the pressure to provide a substantive response comes from, and how this Wittgensteinian approach aims to dissipate it. Here too Anscombe's discussions of Wittgenstein's work will prove helpful: her paper *The Reality of the Past* provides an example of how such pressure emerges, and how a Wittgensteinian account aims to engage with it, that will provide an important reference point for discussion in subsequent chapters.

The paper begins with a presentation of two simple but related language-games:

> Let us imagine that someone is taught (1) to say "red" when a red light is switched on before him, "yellow" for a yellow light, and so on; and (2) next to say "red", "yellow", etc., when lights of the appropriate colours *have* been switched on but are now off. [3, 103]

Anscombe has us imagine a spectator who finds (2) and its relation to (1) unintelligible since in (2), rather than being corrected for saying "red" when there is no red light (as they would be in (1)), the speaker's utterance is acknowledged as correct. From the perspective of such an observer, "[i]t is as though the learner were taught first to act according to a certain rule and then to break it" [3, 104]; for if we just consider each utterance of 'red' by itself, there seems to be nothing to differentiate an incorrect response to the present situation from a correct response to a past situation.[11]

This initial puzzle, which emerges from 'zooming-in' on these particular utterances, is compounded by the interlocutor's attempts to take a reflective perspective on our use of names in the present and past tense. Achieving such a perspective ought to help in resolving our initial puzzle; but the interlocutor's attempts to find it are thwarted by his attachment to a particular picture of how we use language to name something in the present tense, and his attempt to apply it to the past:[12]

> When I think of my acquaintance *A*, and think that he is in Birmingham, it is he, *A*, the very man himself, and Birmingham, that very place, that I mean, and not some intermediate representation of them. I might try to emphasize this by going and finding and pointing to the man and the city; not that I imagine that I should thereby make clear what it is to mean them, but I should then be

[11] Despite important differences, some analogies can be drawn between the puzzles that Anscombe is concerned with and the topic of this dissertation. The philosophical perplexity that Anscombe seeks to capture could also be posed in terms of explanations such as,

PRES: *S* said "red" because the light was on,

and,

PAST: *S* said "red" because the light *had been* on.

In (PRES), S's utterance is a response to something present or actual at the time of her utterance, i.e. the red light. But in (PAST) , S's utterance is a response to something that is no longer present or actual, but in the past. We can imagine our interlocutor wanting to understand how explanations like (PAST) do their work, when what they describe are correct utterances of "red" as part of (2).
One way of bringing out what puzzles this interlocutor is by saying that he doesn't understand what difference there can be between an utterance that can be explained by (PAST), and an apprently identical utterance that happened to be made after the light had been on, but was not counted as a correct response to that light. Such an interlocutor could be understood as looking for something *in virtue of which* an explanation counts as representing an utterance of the sort described in the *explanandum* of (PAST), i.e. an utterance that is explained by, and correct in virtue of, the fact that the light had been on. Given such a concern, it seems besides the point to state that such an utterance is one made *because the light had been on.* For this response simply repeats the fact that the utterance is subject to this kind of explanation, whereas what the interlocutor wants is a justification for our taking it as such.

[12] Cf. PI §455-7:

> We want to say: "When we mean something, it's like going up to someone, it's not having a dead picture (of any kind)." We go up to the thing we mean. (455)

> exhibiting them themselves, and I want to insist that I mean them as directly as that. . . . The name or thought of something past seems to point to its object in just the same way as the name or thought of any other actual thing: yet how can it, since its object does not exist? [3, 103].

This picture of how names work might seem helpful relative to the use of language described in game (1):[13] the speaker says ‘red’ when a red light is there before him as he makes his utterance – it is as though the utterance pointed to that red light. But this compounds the puzzle about game (2); for there there is nothing for the utterance to point to, yet it is counted as correct. The interlocutor is fixated by a picture of one particular use of language (a name as pointing to an object), and gets into difficulty when he tries to apply that picture to other (related) uses. He has a picture of how names work in present tense descriptions that accompanies or is somehow part of his pre-reflective use of names, and becomes puzzled when he tries to understand other related uses of names in the same terms.

Note that the picture itself could be quite innocuous;[14] our philosophical perplexity emerges from the way in which we apply it in our reflective thought.[15] In this particular case, because the picture that seemed to give the essence of naming doesn’t apply, the use

[13] Note that he might also become puzzled through an attempt to apply this picture to various aspects of our present tense use of names as well. For instance, this picture of a name as pointing to its bearer might lead one to conclude that ‘this’ is the “only *genuine* name; so that anything else we call a name was one only in an inexact, approximate sense” (PI §38).

[14] Cf. PI §424: “The picture is *there*; and I do not dispute its *correctness*. But *what* is its application?” See also the surrounding sections, and e.g. §374.

[15] Cora Diamond makes this point in her discussion of pictures we may have of ‘logical necessity’, and their role in our reflection on such topics as inference:

> That it can be misleading is not to say that we should really give up using the picture, stop talking that way. It is no accident that we do; no mischievous demon has been at work in our language putting in misleading analogies which the philosopher can simply discard when he has seen through them. To give up thinking and talking in such terms would be to give up the life in which these figurative expressions do have an application . . . The picture of a necessity behind what we do is not the to be rejected, but its application looked at. When Wittgenstein attacks a picture, it is the picture thought of as giving us an idea of a use, and elaborations of the picture which emphasize just what is misleading. [19, 259]

of names to describe the past comes to look mysterious: how is it possible to use a name to refer to something that is no longer there? This leads us to look for something to ground that use: something that will explain *in virtue of what* an utterance counts as playing this role, and *how it is* that the speaker can be justified in making such an utterance. This ground would thus explain how our use of names in the past tense is possible.

Anscombe goes on to point out that:

> It is of no use to say to someone philosophically perplexed at this that "red" is said when there has been red. For he is looking for, and cannot find, a difference between there having been red and there not having been red. We say "He is looking in the wrong place, in the present and not in the past". But what is it to look in the right place? He is looking for a justification; for surely "was red" is not said without justification when it is said rightly. But there seems to be no justification unless he finds himself reiterating the very thing he is trying to justify. [3, 104-5]

In the body of her paper Anscombe considers a range of responses that try to give the interlocutor what he thinks he wants, i.e. to specify something *in virtue of which* an utterance counts as a correct description of a past event. But it turns out that any such response cannot give the interlocutor what he wants; for they either seem to miss what is essential in that use of language (e.g. that it refers to something *past)*, or they involve an appeal to the idea they purport to explain, making the whole account circular.[16]

[16] The main candidates she considers for such a response all depend on an appeal to memory, which (to continue the analogy to our central topic) we might put by relating (PAST) to an explanation like,

REM: *S* said "red" because he remembered a red light.

On this picture, one could propose that an utterance counts as a description of a past event *in virtue of* being caused by the speaker's memory of that event. But here too any such account cannot give the interlocutor what he thinks he wants, since in explaining what it is to remember it will turn out that "either we have left out all reference to the actual past, or we have surreptitiously introduced it into an explanation that proposed to do without it" [3, 110]. Anscombe summarizes the results of this approach as follows:

> I was tempted to use the idea of memory to explain how the learner who named the coloured lights after they had been switched off was able to mean them. If he said "red" etc. *remember-*

Reflection on this failure puts pressure on the whole idea that we could state an 'independently intelligible condition' that grounds this use of language, since our attempts to specify such a condition fail to provide us with what we think we want. The alternative approach Anscombe considers is one she takes to be an application of Wittgenstein's "ideas and methods of discussion".[17] She imagines such a response beginning along the following lines:

> To speak of something as past is to have the kind of practice that you have made the learner acquire in your example [i.e. (1) and (2)]. The example is, of course, artificial, but a description similar in principle, but far more extended and complicated, would tell you how words are used in speaking of the past; and to use words in this way *is* to speak of the past. [3, 117]

Anscombe's descriptions of simple language-games such as (1) and (2) above can thus be seen as clarifying aspects of the grammar of our use of names in the past tense, and the way they relate to other uses of names. In this sense they are akin to the descriptions of simple language-games that run throughout Wittgenstein's later work. One purpose of such games is (as he puts it at PI §130) to serve as "objects of comparison" whose role is to "throw light on the facts of our language by way not only of similarities, but also of dissimilarities". Describing these simple uses of language helps us to see various aspects of the grammar of our own language, which are open to view in the ways we use words—in the present case, aspects of our use of names to describe things that happened in the past. Part of that

ing, then his behaviour was intelligible and intelligent; his use of "red" in (2) was not a misuse; he was applying the colour names to their appropriate objects as he learned to do in (1), only he was now applying them to past, not present objects. But now it appears that I cannot understand what it means to say he said "red" *remembering*, without understanding what it means to say that he meant the past event; hence my explanation is not an explanation but contains within itself the thing that it purported to explain; nor can I think that if I say he remembers I am saying that he has something that *shows* him the meaning of the past tense. I am saying that the learner acted intelligently if his saying "red" in (2) was an expression of knowledge of the past showing of the red light; I am not stating an independently intelligible *condition* on which his utterance would express such knowledge. [3, 109]

[17] See footnote 3 of [3, 114]

grammar will lie in the connection we see between these two games: that it is natural to us to see game (2) as an extension of game (1), i.e. as what we would count as 'going on in the same way' with our use of names, rather than a new and unrelated use.

This mode of response therefore works by giving a description of the way we use words to describe past events (or simpler versions of that use such as (2)) in a way that aims to clarify that use. Anscombe goes on to point out that a response of this sort cannot tell us *in virtue of what* something counts as an utterance about the past:

> [T]he purpose of the description . . . is not to show one what one is really knowing when one knows something about the past. Indeed, if that were supposed to be its function, it must fail; noone could understand, e.g., the description of the learner that I gave in connection with the coloured lights, if he did not *already* understand the past tense; for it was used in the description of what the learner did. The purpose of the description is rather to make us stop asking the question "What is it that I really know?", and stop looking for a foundation for the idea of the past. I spoke earlier of the perplexity which arises from the fact that one looks for a *justification* of "red" in stage (2) of the procedure I described, or of "was red", and cannot find what one is looking for. The purpose of answering the question "How does the past tense have meaning?" by giving a description of use is to make one think that this search for a justification is a mistake. [3, 117-8]

The method Anscombe describes here should be understood as a version of the approach she applies in *Intention*: she describes particular examples in a way that is intended to highlight (part of) what we call 'intentional action', but without purporting to ground or justify our application of that concept.[18] Part of the purpose of such descriptions is to stop

[18] Thus one can see Anscombe's description of the gardener—which is the example through which we come to see her A-D order—as playing the same role as the description of a language-game to highlight some aspect of our grammar

us from looking for a substantive response to our questions: one that provides a justification for our concepts, or tells us e.g. *in virtue of what* an utterance counts as a description of an intentional action and *how it is* that a speaker is justified in making such an utterance. But it is precisely this aspect of Wittgenstein's method that Anscombe states is 'most difficult to accept':

> [N]amely, the fact that he attacks the effort at justification, the desire to say: "But one says 'was red' because one knows that the light *was* red!" One says "was red" in these circumstances (not:*recognizing* these circumstances) and that *is* what in this case is called knowing the past fact. To say this is *not* to profess to give an analysis of what one really knows. [3, 118]

3.3 'Reason for Action' as a Grammatical Concept

Part of my purpose in presenting Anscombe's discussions, and their relation to material to Wittgenstein, is to provide a different perspective on (MQ) and (EQ), and the puzzles that led to them. In the previous chapter, we saw how (MQ) arises from the fact that whenever we have an explanation of the form,

RA: S ϕed because$_R$ _____,

where the subscript $_R$ marks the fact that the *explanans* gives the agent's reason for her act, we can always ask ourselves what differentiates these explanations from apparently similar ones of the form,

NR: S ϕed because _____,

where the *explanans* does not give the agent's reason. (MQ) asked us to explain this difference.

The accounts surveyed in the previous chapter all agreed in that a response to this question had to be given in general terms, and had to focus on what was immediately described by the terms of these explanations. This gave the impression that all three of them were trying to provide a substantive response – one that specified *in virtue of what* we counted something as an instance of reason-giving explanation, and so as describing an instance of acting for a reason.

The discussion in this chapter has suggested a different approach to characterizing this difference. Rather than looking for some mark or feature of what is immediately represented by the terms of these explanations, we should seek to characterize the different grammar of those terms. For it belongs to the idea of a reason-giving explanation that its *explanandum* describes an instance of acting for a reason, and its *explanans* describe a reason for acting. We can therefore seek to clarify the grammar of descriptions that play these roles – partly, perhaps, by considering how it differs from the grammar of apparently similar descriptions that play a role in other kinds of explanations. This is to treat 'acting for a reason' and 'reason for action' as grammatical concepts.[19]

This already suggests a significant shift from the conception of (MQ) and (EQ) described in §2.4. For as well as rejecting the idea of a substantive response, we are also questioning the more general thought that a characterization of this difference must be given in terms of what is immediately represented by these explanations.

But it also introduces an additional complication. We saw in the previous chapter that all three accounts assumed that it would be possible to provide a general account of these explanations in abstraction from the specific acts that they described. But once we begin to attend to the particular uses of language that we might count as descriptions of 'acting for a reason'—or to the explanations we might count as 'reason-giving'—what we see is not a single uniform category, but rather a motley of related forms. This means we need to be

[19] Sebastian Rödl expresses a different version of this point as follows, focusing on a related use of language:

> The concept of a reason is a formal concept; it applies to what it describes with regard to its role in reasoning. A reason for acting is something from which one may reason to an action; it is something that may serve as a premise of a practical inference. [30]

What Rödl calls a 'formal concept', I am calling a 'grammatical concept'.

open to the possibility (which is anyway implicit in Wittgenstein's treatment of 'grammatical concepts – cf. PI §65, and Rush Rhees' discussion of this and surrounding passages in [28]) that a our 'grammatical concepts' apply to a multiplicity of cases that are related to each other in a variety of ways, rather than a single uniform category.

So far, we have been working with the broad idea of reason-giving explanations, which might be made explicit through the schema,

S's reason for ϕing is that _____,

or

S ϕed because$_R$ _____

However, this general form encompasses a range of related kinds of explanation. An example will help illustrate this point. Imagine the following scenario: you walk into the kitchen to find a friend stretching out a sticky mass of flour and water, and ask her, 'Why are you doing that?'. We can imagine a range of possible responses here, each reflecting a different aspect of your friend's understanding of what she is up to. Here are some examples:

1. I'm making bread.
2. James is coming to dinner.
3. We're almost out of bread.
4. I'm developing the gluten.
5. I think this is a better method for developing the gluten in wet dough.
6. The timer just went off.
7. I don't think its fully developed yet.

All of these possible responses are supposed to fit the same action: we are imagining a case in which the agent is knowingly stretching out the dough, in response to the timer that indicated it was time to check on it, based on her judgement that the dough wasn't fully developed, with the purpose of stretching the gluten, this being her preferred means of achieving that, all with the goal of making bread, because she had run out and her friend James was coming to dinner.

This extended description brings out various facets of the situation, with each imagined response providing an explanation that reflects a particular way of being interested in the central action, along with different kinds of ignorance about the activity in question and the circumstances in which it takes place.[20] What's significant, for our purposes, is that each answer can be thought of as potential explanation for this action—one that specifies an end the person is pursuing, or a consideration from which they acted and, in that minimal sense, something we could call her reason for acting. What's more, it looks like each of these answers can be used to fill out our general schema for reason-giving explanation of intentional action:[21]

S's reason for stretching the dough was that...

1. she was making bread.
2. James was coming to dinner.
3. she was almost out of bread.

[20] The variety encompassed in this list illustrates the audience-relative character of reason-giving explanation. Which of these responses was appropriate in a given situation would depend on the knowledge and interests of the enquiring party. For instance, (2) presupposes knowledge that what the agent is doing, here and now, is (part of) making bread, whereas (1) does not. Likewise (4) and (5) presuppose a certain kind of interest in, or knowledge of, the process of making bread, whereas (1) might be asked by someone with nothing more than a casual interest in what the agent was up to.

[21] Note that that I have only specified descriptions of the action under which it is intentional, and explanations that acknowledge it as such. This means that the following case does not belong in our list: "Why are you using my flour?" "I didn't know it was yours." Nor would: "Why is she making bread?" "She's nervous about James' visit". That is, I have deliberately chosen explanations that cite considerations that, in some broad sense, the agent knows are a basis for her action.

4. she was developing the gluten.
5. she found stretching the dough to be the best way to develop the gluten in wet dough.
6. the timer just went off.
7. it wasn't fully developed.

This list brings out the relation between the idea of reason-giving explanation and the agent's own understanding of what she is up to that was stressed above. I deliberately picked an example in which the agent has a robust understanding of what she is doing: she knows that she is stretching out dough, in response to the timer that indicated it was time to check on it, based on her judgement that the dough wasn't fully developed, and so on. Requests for explanation can be seen as attempts at coming to share some part of this understanding, which in turn provides the measure of their success: the explanations are successful to whatever extent they help us understand the action as the agent does.

End-Specifying Explanations

We can use our baking example to give us examples of explanations fitting the various forms described above. First, explanations (1) and (4) work by specifying some particular end that the agent is pursuing. In each case, the action of stretching the dough is teleologically related to the pursuit of that end: in the first case, by being a *stage* or *part* of the process of making bread, and in the second, by being a *way* of developing the gluten, which is in turn a *part* of the process of making bread. Normativist accounts of reason-giving explanation treat explanations of the form 'S is doing A because she is doing B' as their central case of reason-giving explanation, and build their account around them.

From our example, we can already see that this kind of explanation is apt for capturing various aspects of our explanatory talk about action: wherever an action is made intelligible by being somehow related to a further end pursued by an agent, we can make that action intelligible by an explanation of this form.[22] For purposes of our discussion, I am going to categorize such explanations in a larger group that I shall call 'end-specifying explanation'.

Speaking in general terms, we can say that end-specifying explanations show the point of the agent's action by relating it to whatever is described in the *explanans.* In case (1) from the list above, we see the point of the action by seeing that it was instrumentally-related to some end that the agent was pursuing (i.e. baking a loaf of bread), and further grasp that the agent understood herself to be pursuing this end in undertaking this particular action. In other cases, an end-specifying explanation might show us the point of the action in other ways, as in the examples below:

- *S* is working the dough with her hands because the sensations are pleasurable.
- *S* is working the dough by hand because it is part of the training of an apprentice baker.

Here too both explanations make the action described in the *explanandum* intelligible by

[22] As should be clear from even these two examples, there may well be a *variety* of kinds of teleological relations between actions that can be expressed in explanations of this sort, but we can provide an abstract statement of what unites them by simply saying that they show the action in question to be teleologically-related to some end.

Explanations of this form are part of the basic background to our talk about people's actions, since it is only insofar as we represent people as engaged in the deliberate pursuit of ends that other explanatory talk about actions makes sense. (For arguments about the centrality of this form of explanation, see [36], which is a development of key ideas from [4].) One way of bringing this out is by noticing the way that ground-giving explanations fundamentally depend on these end-specifying explanation: understanding an explanation of the form '*S* is doing A because p' will typically involve knowledge of, or speculation about, the end that the agent is pursuing. Indeed, if one knows what ends an agent is pursuing, one can usually make an informed guess about how a particular fact might shape her pursuit of those ends. By contrast, if one doesn't know what further ends an agent is pursuing, it might be very difficult to make an informed guess as to how they might respond to a particular fact, and without further background an explanation of the form '*S* is doing A because p' might fail to make S's action intelligible. The following example, adapted from Anscombe, makes this point. Supposing one knew that a particular person were suicidal, it would be easy to make an informed guess as to how they might act on the following considerations:

> Strong alkaloids are a deadly poison to humans
> Nicotine is a strong alkaloid
> What's in this bottle is nicotine

But without this background knowledge, one would be unlikely to guess that a person would respond to these considerations by drinking the contents of the bottle—and should they behave in that way, one might be puzzled by an explanation that represented their action in these terms.

showing the good or purpose of that action, not by connecting it to some further finite action, but rather by connecting it to some other end that belongs to the agent. Specification of this end will still be an essential part of the agent's understanding of what she is doing and why, meaning that these too could be cases of reason-giving explanation. Despite the differences in what is specified in their *explanans*, from a certain level of abstraction they can be seen as working in the same way as the instrumental example above: they explain a particular action by representing it in terms of some end towards which the agent understands her action to be directed.

Indeed, even more specific cases of rational-responsiveness can be seen to work in the same way:

S1: *S* turned left because she was following the sign.

RU1: *S* wrote '16, 18, 20, 22' because she was following the rule '+2'.

O1: *S* turned left because she was obeying T's order.

R1: *S* said "CAT" because she was reading the word on the card.

Here too the explanations do their work by redescribing the action specified in the *explanandum* in a way that brings out its purpose or point: it was, for instance, an act of reading, or an obeyance of an order. By redescribing the action in terms of the concepts deployed in the *explanans*, these explanations reveal it to be a way of performing a particular action (e.g. reading a word, following an order), and so as directed towards that end.

Taken as a whole, end-specifying explanations do their work by representing the act described in the *explanandum* as, in some sense, directed towards realizing whatever is specified in the *explanans.* In some cases, the *explanans* will specify some particular end, with the act represented in the *explanandum* done *for the sake of* that end, either as preparation for realizing it, or as a stage in that process. In such cases, the first action can be understood as a *part* or *stage* of the more general action described in the *explanans.*[23] In other cases,

[23] See [4], and [36]

the action described in the *explanandum* will be a way (perhaps the way) of realizing that end, as is the case in the final list of examples above. An end-specifying explanation is then one that explains an action by representing the agent as directed towards the end specified in the *explanans*, and as performing that action because she understands it in terms of that end.[24]

Ground-Giving Explanations

Ground-Giving Explanations state considerations that provided the grounds for a particular action. Broadly speaking, the cases our authors are concerned with fall two categories: those that specify general knowledge that grounded the agent's action, and those that specify particular knowledge of the circumstances of action that explain why the agent acted as she did.

Explanations in the first group will include those that exhibit knowledge belonging to a skill or craft—e.g. that developing the gluten in the dough is an essential stage in making bread, or that stretching the dough is an effective way to do this—along with other kinds of non-specialist general knowledge, e.g. how one gets around a modern city or how one procures goods and services. In other words, these are explanations that appeal to various kinds of 'practical knowledge' that involves 'knowing one's way about', which Anscombe described as being fundamental to intentional action:

> Although the term 'practical knowledge' is most often used in connexion with specialised skills, there is no reason to think that this notion has application only in such contexts. 'Intentional action' always presupposes what might be called 'knowing one's way about' the matters described in the description under which an action can be called intentional, and this knowledge is exercised in the action

[24] As will become clear, an action can have more than one end. For instance, if I am following the signs to the coffee shop because I am trying to find the bathroom, falsely believing that they are next to each other, then when I turn left following a particular sign, my action will be correct relative to the end of following that sign, but incorrect relative to the end of finding the bathroom.

> and is practical knowledge. [4, §48]

Explanations in this first group will thus express general knowledge of how things are done (i.e. knowledge associated with skill or know-how), or of the principles or norms associated with a certain practice. In our bread example, the knowledge represented in these general claims is *knowledge of how to bake bread*, but other kinds of general knowledge will be involved in our other examples. For instance, [S1] involves knowledge of the norms of signage in our culture, i.e. that if a sign tapers in a particular direction, one is to go in that direction to follow it. Likewise, [RU1] involves mathematical knowledge (specifically, knowledge of how to apply a particular rule), and [R1] involves knowing how to read written English.

Explanations in the second group will include reference to specific facts of the immediate or mediate circumstances, e.g. those that made a particular action apt, necessary, attractive, and so on, *here and now*. For instance, the fact that James is coming to dinner might explain why our agent is going to the trouble of *making* bread, rather than simply buying it; the fact that this dough is insufficiently developed explained why she is doing *this* turn of the dough, instead of leaving it to proof; the fact that the timer went off indicates that now is the time to check on the bread; etc. Explanations in this second group will typically express immediate or mediate knowledge of the circumstances of action (e.g. knowledge based on perception of those circumstances, testimony, inference, etc.). Rather than citing general facts about how things are done, they will cite specific facts about the circumstances of action, i.e. considerations of the form *that such-and-such is the case*. The consideration specified in these particular grounds-giving explanations shows why the action described in the *explanandum* was required relative to the relevant end in the present circumstances.

A significant subset of explanations in this category will therefore represent the action as a *response* to some fact in the immediate environment. The counterparts to our first list of examples belong in this category:

S1": *S* turned left because the sign pointed that way.

RU1": *S* wrote '16, 18, 20, 22' because the previous number was '14'.

O1": *S* turned left because T said 'Left turn!'.

R1": *S* said "CAT" because "C-A-T" was written on the page.

This category of rational response will prove to be important in what follows, especially in relation to the further group of explanations discussed below.

Taken as a whole, ground-giving explanations do their work by representing the act described in the *explanandum* as based on, or grounded by, the consideration specified in the *explanans*, which shows why the agent understood the act in question to be directed towards some particular end. For instance, an explanation such as

GG1: *S* is stretching the dough because that is a way of developing gluten

explains S's action by representing it as based on the consideration that stretching the dough is a way of developing gluten. [GG1] will help us to make sense of S's action when we see that, based on this consideration, *S* understood her action to be directed towards the realization of her end of baking a loaf of bread, specified in the counterpart explanation

ES1: *S* stretching the dough because she is baking a loaf of bread.

Particular grounds-giving explanations will show features of the immediate or mediate situation on the basis of which the agent understood her action to be directed towards the realization of some end, or show her action as a response that she understood as directed towards that end. For instance, an explanation such as

O2: *S* turned left because T said 'Left turn!'

explains S's action by representing it as a response to the fact that T said 'Left turn!'. [O2] will help us to make sense of S's action when we see that *S* understood her response to be a way of realizing her end of obeying T's orders, specified in an explanation such as

O1: *S* turned left because she was obeying T's order.

As these two examples show, while the end is not explicitly cited in these ground-giving explanations, it still plays a role in our grasp of the explanatory role played by the considerations that are specified in them, since they are considerations that the agent understood to relate the action to that end.

Explanation by (Mental) Cause

Grounds-Giving Explanations suggest a further kind of explanation—one that is distinct from, but closely related to, the reason-giving explanations that are our primary topic. For reasons that will become apparent, I will call these 'explanations by (mental) cause'. Such explanations cite some happening prior to the event whose causal-explanatory role is related to the agent's grounds for acting, but which is not itself among those grounds. I use the term 'happening' in a broad sense here: it will include particular events external to the agent, but also aspects of the agent's psychology, such as her *noticing*, *remembering*, *hearing*, *realizing*, that such-and-such. Indeed, in most cases, if an external happening could figure in such an explanation, then that explanation will have a counterpart that mentions a psychological event.[25] In some cases, the external happening will be only contingently related to the action it explains: an agent might see or hear something that reminds her to do something quite unconnected. In such cases, we could say that the explanation specifies a *cause* for the action, but not a *reason*. But in other cases, this distinction is less clear cut. For example, consider the following explanations, related to (6) above:

MC1: *S* is stretching the dough because the timer went off,

whose psychological counterpart might be:

MC2: *S* is stretching the dough because she heard the timer go off.

[25] Thus the *explanans* in this category of explanation will be what Anscombe calls a 'mental cause' [4, §10].

Both explanations represent the agent's action as a *response* to an event, and in that sense specify a *cause* of that action. Furthermore, both can be related to a consideration on the basis of which she acts: *that the timer went off* indicates *that it is time to turn the dough*, which explains the agent's act of stretching the dough by showing its timeliness. But while the *explanans* of (MC1) might be thought of as a reason for the action[26], the *explanans* of (MC2) is not. Indeed, it is rarely the case that the agent's reason for acting was *that she heard* such-and-such—it will rather be *that such-and-such*, i.e. whatever it is that she heard.[27] Nevertheless, the efficacy of the cause specified in (MC2) depends on the normative relation expressed in the closely related reason-giving explanations: the agent responds in this way because she recognizes what she hears as a *timer*—and one must understand what a timer is, and why the agent set this one, to fully understand the causal relationship between her hearing the timer and turning the dough.[28] Explanations such as (MC2) are thus intimately to reason-giving explanations, to the point where they cannot play their full explanatory role independently of the understanding that is cultivated by such explanations; but they are not themselves instances of such explanations.

[26] Nothing in particular hangs on how we characterize this case, but we will return to the notions of a *rational response*, *cause*, and *reason for action* in more detail later.

[27] For example, one might participate in an experiment where the instructions were to raise one's hand when one *heard* a particular buzzing sound, they point being to see when one heard it. Here one is responding to the fact that one heard it, which is different from a case where one e.g. looks around because one hears a buzzing sound, responding to the sound itself, rather than the fact that one heard it.

[28] This is a contentious claim: one might argue that one can fully understand the causal relation without understanding its normative significance. This is a complex case, since the explanation locates the act and its cause within a practice of using timers, and my claim is that the causal relation cannot be fully understood independent from that practice. The character of acts that belong to practices, and the way that certain causes figure in their explanation, will be an explicit topic in §5.3 when we consider Wittgenstein's discussion of reading. (A more straight-forward example would be one in which the agent heard something unrelated to her bread making, which for some reason prompted or reminded her to check on the dough. Here the happening could be part of the aetiology of her action, but it would not be related to it in the same way as the beeping of the timer.)

My Focus

This overview has suggested that the general category of reason-giving explanations, i.e. explanations of the form,

> S is doing A because . . . ,

encompasses two further very general categories: what I have called 'end-specifying' and and 'ground-giving' explanations. Each of these categories also admits of further specification. For end-specifying explanations, this will be a matter of understanding the character of the end and its relation to the action it purports to explain. Thus, three explanations,

E1: *S* is working the dough because she is making bread;

E2: *S* is working the dough because it is pleasant to do so;

E3: *S* is working the dough because she is training to be a baker,

all represent the action described in the *explanandum* as directed toward realization of the end described in the *explanans*, but the character of that end, and its relation to that action, differ from case to case. A generic characterization of an 'end-specifying explanation' will abstract away from these specific differences. Analogous points can be made for ground-giving explanations, which will themselves encompass an array of different cases.

This provides important background for our question of what it would be to apply this Wittgensteinian approach to our topic. For it suggests that any attempt to characterize the grammar of concepts such as 'acting for a reason' or 'reason for action' that aims to cover *all* of these cases will have to abstract away from all of these differences. The result will therefore be the peculiarly empty claim that seemed to capture the accounts of the primitivist and perhaps the normativist,

> To act for a reason is for one's act to be explainable by a reason-giving explanation,

where that encompasses *any* of the explanations described above. As we shall see, an account framed at this level of generality is philosophically unhelpful – both because it seems peculiarly substanceless, but also because by encouraging us to ignore the details of particular cases, it tends to engender philosophical confusions.

The alternative would be to begin with particular cases—as Anscombe does with her gardener in *Intention*—and see how we would apply concepts like 'acting for a reason' or 'reason for action' in our descriptions of them. What we learn from any such case might not immediately apply to every other application of these concepts – but it might help us begin to see certain similarities and differences between the cases to which we'd apply these terms. This slow and somewhat piecemeal process would therefore *be* our attaining a reflective perspective on what we call 'acting for a reason'.

In the positive portions of what follows I therefore focus on one particular kind of case: specific ground-giving explanations that represent the act they describe as a *response* to something, such as those on our earlier list:

S1": *S* turned left because the sign pointed that way.

RU1": *S* wrote '16, 18, 20, 22' because the previous number was '14'.

O1": *S* turned left because T said 'Left turn!'.

R1": *S* said "CAT" because "C-A-T" was written on the page.

Indeed, I shall be particularly concerned with explanations like (R1) – that is, with instances of 'reading'. In positive terms, what I shall provide by the end of §5 is therefore only a partial clarification of the grammar of 'acting for a reason'. But just doing this much will prove valuable. First, it will provide a different model for responses to questions such as (MQ). Rather than aiming to provide a general account that specifies *in virtue of what* something counts as an instance of acting for a reason, this approach works by describing particular kinds of case with a view to providing a reflective perspective on their grammar.

A reflective perspective on the grammar of reason-based responses cannot claim to survey everything we might count as an instance of acting for a reason – but it does survey part of it. More importantly, focusing on these kind of cases helps contribute to a question that provided the background for much of the contemporary discussion: should we think of our reasons for action as among the causes of our actions? For responses such as those described above are cases where our application of the concepts 'reason for action' and 'cause of action' seem to overlap with each other.

To this end, in the next chapter I shall discuss Wittgenstein and Anscombe's treatment of these cases – and the controversial claims about reasons and causes that provided the background for the contemporary discussion.

CHAPTER 4

WITTGENSTEIN AND ANSCOMBE ON REASONS AND CAUSES

In §2 I suggested that contemporary philosophers agree in looking to an account of reason-giving or rational explanation for an answer to the following questions, even as they disagree about the form such an account should take:

MQ: What is it to act for a reason?

EQ: How do we come to have first-personal knowledge of our reasons for action?

We also saw that discussions of reasons and causes in the work of Wittgenstein and Anscombe provided an important background to contemporary interest in these questions. In fact, these discussions point to another important site of broad consensus (and specific disagreement): proponents of all three accounts agree that the way in which Wittgenstein and Anscombe drew a distinction between reasons and causes was unhelpful, and needs to be somehow clarified before we can make progress on our two questions (though they differ about what that clarification will involve).

Broadly speaking, there are two opposing views on the matter. For the first group (which includes causal-psychologists influenced by Davidson's work), Anscombe and Wittgenstein provided flawed arguments for the claim that reasons cannot be causes. Progress on our two main questions therefore involves showing that these arguments are flawed, thereby allowing us to endorse the claim that reasons are causes in precisely the sense that Wittgenstein and Anscombe denied. This is understood to be a key step in providing an adequate response to our two main questions.

For the second group, the problem with Wittgenstein and Anscombe's treatment of the distinction lies not in their arguments, but in the unduly narrow conception of causation deployed in their discussions. This group agree that reasons are not causes, given the notion of ‘causality’ that Wittgenstein and Anscombe were working with. But they go on to argue that we can expand our notion of causality, and that once we do so it will once again be possible to endorse the claim that reasons are causes. As above, doing so is to be a key step to making progress with our two main questions.

In this chapter, I shall look more closely at some of the texts in which Wittgenstein and Anscombe appear to reject the claim that reasons are causes. My goal will be two-fold: first, to show in outline why Wittgenstein and Anscombe also think that the ways in which they draw a distinction between reasons and causes bear directly on our response to our two key questions; and second, to show that rather than being the origin of avoidable confusions, Wittgenstein and Anscombe’s discussions pose some fundamental questions that must be handled by contemporary accounts. In §5, I shall proceed to develop these arguments in more detail through a discussion of Wittgenstein’s treatment of *reading* in the *Philosophical Investigations.* As we shall see, *reading* is a concept that brings together many of the issues pertaining to the distinction between reasons and causes and their bearing on our main questions, and Wittgenstein’s treatment of the topic is ultimately more illuminating than his brief remarks in *The Blue and Brown Books.* This will then provide the basis for a more careful assessment of contemporary responses to (MQ) and (EQ) in subsequent chapters.

4.1 An Argument for the Claim: Reasons Cannot Be Causes?

4.1.1 Wittgenstein on Reasons and Causes in The Blue Book

I shall begin with the views of the first group, and return to those of the second in §4.3. This first group understand both Wittgenstein and Anscombe to be offering flawed arguments for the claim that reasons cannot be causes. I hope to show that, while this reading has some grounding in their texts—especially if one focuses in on particular passages such as the ones in *The Blue Book*—it in fact obscures what Anscombe and Wittgenstein are doing in drawing a contrast between reasons and causes. To this end, after showing why it might appear that

Anscombe and Wittgenstein are offering arguments for a strict distinction, I show that the conclusions they purportedly reach stand in direct tension with other claims they endorse. Interpretative charity demands that we at least consider an alternative reading, which I go on to outline in §4.2.

The passage most commonly-cited as providing Wittgensteinian arguments for the claim that reasons cannot be causes comes early in *The Blue Book.* Wittgenstein is discussing a case where someone teaches the pupil what they mean by 'red' with a sample, and then gives the order "Now paint me a red patch". The confusion between reason and cause sets in when we imagine a particular kind of response on the part of the pupil:

> One is led into this confusion by the ambiguous use of the word "why". Thus when the chain of reasons has come to an end and still the question "why?" is asked, one is inclined to give a cause instead of a reason. If, e.g., to the question, "why did you paint just this colour when I told you to paint a red patch?" you give the answer: "I have been shown a sample of this colour and the word 'red' was pronounced to me at the same time; and therefore this colour now always comes to my mind when I hear the word 'red' ", then you have given a cause for your action and not a reason. (BB14)

Wittgenstein continues:

> The proposition that your action has such and such a cause, is a hypothesis. The hypothesis is well-founded if one has had a number of experiences which, roughly speaking, agree in showing that your action is the regular sequel of certain conditions which we then call causes of the action. In order to know the reason which you had for making a certain statement, for acting in a particular way, etc., no number of agreeing experiences is necessary, and the statement of your reason is not a hypothesis. The difference between the grammars of "reason" and "cause" is quite similar to that between the grammars of "motive" and "cause".

> Of the cause one can say that one can't know it but can only conjecture it. On the other hand one often says: "Surely I must know why I did it" talking of the motive. When I say: "we can only conjecture the cause but we know the motive" this statement will be seen later on to be a grammatical one. The "can" refers to a logical possibility. (BB14)

Finally, Wittgenstein concludes:

> The double use of the word "why", asking for the cause and asking for the motive, together with the idea that we can know, and not only conjecture, our motives, gives rise to the confusion that a motive is a cause of which we are immediately aware, a cause 'seen from the inside', or a cause experienced. (BB14)

These comments appear to commit Wittgenstein to the following claims:

1. In responding to the question 'Why did you do such-and-such?', one can *either* give a reason *or* a cause for one's action. This reflects an ambiguity in the question 'Why?' which can be asked in either sense (i.e. in a sense that calls for one kind of response or the other).

2. One mark of the difference between a reason and a cause is that one can conjecture about causes—such conjectures being founded on and confirmed by repeat experiences—whereas one 'must know' the reasons for one's actions just insofar as one acts.

3. These points reflect the 'grammar' of our talk about *causes* and *reasons*. Statements that mark this difference, such as "we can only conjecture the cause but we know the motive [or reason]", are 'grammatical', and they describe 'logical possibilities'.

4. Lack of clarity about these grammatical differences produces philosophical confusions: specifically, the claim that a motive (or reason) is a cause of which we are 'immediately aware'.

Deferring questions about how exactly we are to understand the claims about grammar in points (3) and (4), these statements could seem to commit Wittgenstein to (a) drawing some a sharp distinction between reasons and causes, implying that reasons cannot be causes of our actions; and (b) endorsing three specific claims about our knowledge of reasons and causes, i.e. that we can *only* conjecture about the latter whereas we *must* know the former, but that this knowledge should not be understood as a matter of 'immediate awareness'.

Although I shall argue that Wittgenstein remains committed to versions of all four of these claims, they are overstated in this portion of *The Blue Book*, either due to the compressed character of these dictated notes or because Wittgenstein himself was unclear about some of these points. The main problem is that Wittgenstein appears to overstate his grammatical point, i.e. that "we can *only* conjecture about the cause".[1] I shall later suggest that this claim contradicts Wittgenstein's treatment of causal concepts in the *Philosophical Investigations*, and that there are independent grounds for doubting its truth (and its status as a grammatical remark). Indeed, the passage fits best with the rest of Wittgenstein's work if we see this claim as one he ultimately wants to criticise, rather than being something he means to endorse.

However, if we do take these claims at face value, we can represent Wittgenstein as offering an argument of the following form:

1. Knowledge of causes must be based on conjecture grounded by repeated experience

2. One must have knowledge of one's reasons for a particular action just insofar as one acts

3. Such knowledge cannot be based on conjecture grounded by repeated experience.

4. Therefore, what one has knowledge of cannot be a cause for one's action.

[1] The original text is ambiguous – Wittgenstein could be referring back to the case he has described, rather than making a general claim about our knowledge of causes. In that case, it would only be a grammatical point about the kind of cause described in that example.

This would be a straight-forward argument against the claim that reasons were causes, based on the truth of the claim made in the first premise.[2] If one understood the passage in this way, then to reject his conclusion one would argue against this first premise: it is not the case that knowledge of causes must be based on repeat experience.[3]

Indeed, the problem with this argument is that its first premise is false, or at least questionable—and in fact, an example similar to Wittgenstein's can be used to show why. It will help to begin with a grammatical remark about claims based on conjecture: one mark of such claims is that one can respond to them by asking 'How do you know that?' or 'Why do you think that?', where what one is asking for is the grounds for the conjecture. Wittgenstein imagines his interlocutor appealing to an associative habit and past training to explain why a pupil paints a particular shade in response to an instruction—and taken as a straight-forward causal claim, this *is* something that would be based on conjecture.[4] To see why, imagine the following scenario:

> You are trying to teach yourself to identify quite specific shades of colour, and you find that you learn them more quickly if you associate a visual image of an object of that colour with the relevant name. For instance, to remember 'chartreuse' you picture a bottle of the green liqueur.[5]

[2] This is the standard way of reading these paragraphs. See for instance [23, 117-118] . Some argument along these lines was part of the Wittgensteinian orthodoxy that I described Davidson as reacting against in §2.2.2.
Note also that this argument has a similar form to Dancy's argument that reasons are not causes as I presented it in §2.3.1, insofar as it argues that such-and-such is a mark of causation, and reasons-explanations do not have that feature.

[3] This is precisely what Davidson does on pp.17-18 of [15]. [23, 118] represents Anscombe as also providing grounds for rejecting this premise, though as will become apparent I understand her claims to be a development of Wittgenstein's more considered position.

[4] Note that its role in the actual passage is not that of a straight-forward claim, but rather a response that is given "when the chain of reasons has come to an end and still the question "why?" is asked" (BB14). To understand what Wittgenstein is doing here, it is important to ask why a person might feel a need to insist on asking the question 'why?' when the 'chain of reasons has come to an end', and what this particular response is supposed to provide in such a situation. This is explored at length in the rule-following sections of the *Philosophical Investigations*. However, this point does not prevent us from imagining such claims being given in a different context, where they are straight-forwardly causal. See the example in the text.

[5] As Wittgenstein points out, the fact that you rely on a visual image is of now great importance

Here one can imagine noticing that the formation of such habits helped one paint shades of the correct colour. If one asserted as much, one could do so as a conjecture based on such observations—asked ‘How do you know that?’, you might say that you’ve noticed that you tend to be better at painting correct shades of colours for which you’ve formed such habits, and anyway you find them helpful in particular cases.

But now contrast this with some more specific causal claims, e.g. ‘I pictured a green liqueur bottle because you said ‘chartreuse”, or ‘I painted that shade because I recalled a chartreuse-coloured liqueur’. These are both, I take it, plausible candidates for causal claims: the first describes one’s response to hearing a particular word, and the second describes what prompted one to paint a particular shade of green. But in either case, it would be odd to ask ‘How do you know that?’—if the enquirer appeared to understand the sense in which one meant the original assertion, one would not know what else to say, besides re-describing what happened.

This is because *what one is responding to* and *what prompted one to act* can be among the things that, as Anscombe put it, one ‘knows without observation’. In the cases we have described, hearing the word ‘chartreuse’ and imagining a bottle of green liqueur are what Anscombe calls ‘mental causes’:

> A mental cause is what someone would describe if he were asked the specific question: what produced this action or thought or feeling on your part: what did you see or hear or feel, or what ideas or images cropped up in your mind, and led to it? (I§11)

Now if mental causes are (a) genuine instances of causation, and (b) *can* be among things known without observation, then it follows that the first premise of our argument is false: for things known without observation are not based on conjecture, still less on repeat observation. Thus if Wittgenstein means to be offering an argument for the claim that reasons are not causes, it does not succeed.

here—you could also use a deck of picture cards or colour samples to the same end.

4.1.2 Anscombe on Reasons and Causes

Analogous problems arise when we look more closely at Anscombe's objection to Davidson's account, and her broader treatment of the distinction between reasons and causes. Her key argument against Davidson's claim seems to hinge around the idea that if a belief or intention were the cause of the agent's action, the agent would have a felt experience of that causal relation. Thus, one mistake she identifies is "to think that the relation of *being done in execution of a certain intention* or *being done intentionally*, is a causal relation between act and intention" [5, 95]. But in arguing against this claim, she repeatedly points out that the agent needn't have had any *felt or occurent experience* related to this purported cause. The example she discusses is of someone applying extra-force to unjam a telephone dial. Arguing against the claim that the intention to unjam it is a cause of the action, she points out that:

> All that introspection or observation can tell [the agent], we may suppose, is that *it seemed jammed* and then he acted. He doesn't find e.g. that the *thought* of the need to unjam it 'went through his head.' [5, 96]

She continues:

> But didn't he *want* it to move? Well, what was that supposed to be like? Did he *feel* something which he could call a desire that it should move, as when he is showing someone an experiment in which something is supposed to move and he watches it with anxiety? No. This is what he was doing – dialing a number. Of course he wanted it to move! But saying so does not add a new event to the record. [5, 96]

These passages might seem puzzling: why does Anscombe repeatedly suggest that a cause must be felt or experienced, something occurent that 'went through the agent's head'? Why should anyone need to assume *that* if they are to claim that reasons are causes?

This puzzle is only compounded by a consideration of Anscombe's discussion of causation in *Intention.* One of her primary goals, in the first half of the book, is to outline the sense of the question 'why?' that she is concerned with: "that in which an answer, if positive, gives a reason for acting" [4, §5]. Part of the way in which she does this is by a contrast with cases that ask for a 'mental cause' of action, in the sense characterized above.[6]

Anscombe is quite clear that she does not regard this notion of a mental cause as "itself of any great importance" [4, §11]. But one way in which one could make sense of the purported argument against the claim that intentions or beliefs are causes of action is by supposing that she meant to generalize her specific description of 'mental causes' into a general account of 'mental causation'.[7] This would involve the claim that if a psychological state is to be the cause of an action, it must be a 'mental cause' in the sense we have outlined, i.e. something occurent in the mind of the agent. If we did take Anscombe to be committed to such a claim, we could reconstruct her argument as follows:

1. If a belief or intention is the cause of an action, there must be a 'mental cause' – something occurent in the agent's consciousness.

2. But in fact there need be no such 'mental cause' when an agent acts from a particular intention or belief.

3. Therefore belief and intention are not causes of action.

Here, as with Wittgenstein, we have an argument that is based upon a positive assertion of what causality *must* be like – and, as before, the argument is flawed because that

[6] I.e.,

> A mental cause is what someone would describe if he were asked the specific question: what produced this action or thought or feeling on your part: what did you see or hear or feel, or what ideas or images cropped up in your mind, and led up to it? [4, §11]

[7] Causal-psychologists like Setiya appear to understand Anscombe's argument as working in this way—perhaps because their own accounts deploy a uniform notion of 'mental causation' (i.e. Setiya claims that the causation he is concerned with is common to any case in which a mental state causes some behaviour). If Anscombe's description of 'mental causes' were intended in this sense, then it would make sense to think of her as arguing that *all* cases of mental causation as involving felt experience.

premise is clearly false. This is precisely how Anscombe has been read by some contemporary philosophers. Thus Kieran Setiya states that it is, among other things, “a bad philosophy of mind—the illicit focus on what we *feel*—that leads Anscombe to miss the importance of mental causes”:

> Anscombe . . . wants to *contrast* mental causes with motives and reasons for acting. But her arguments for the contrast are bad. So, for instance, in asking whether an agent’s intention in acting is a mental cause, she is led to say “no” because she assumes that mental causes must be *felt* or *experienced* by their agent. [32, 47]

Even if we attribute a more plausible positive claim to Anscombe—e.g. that “one’s intention in doing something need not be a state that precedes one’s doing it, and therefore cannot be its efficient cause” [33, 130]—her argument must still seem flawed insofar as she does not attempt to provide a positive account of causality to support her claim. Indeed, in the absence of such an account, Setiya quotes Anscombe’s own words against her: “[t]he ‘topic of causality is in a state of too great confusion’ for us to assume that causes must precede their effects” [33, 130].

4.1.3 The Broader Context

Although this way of reading Wittgenstein and Anscombe has the advantage of identifying clear arguments in their texts, it ultimately leaves their approach to the distinction between reasons and causes rather mysterious. In both cases, we represent Anscombe and Wittgenstein as making an argument that reasons (or the psychological states that specify them) *cannot* be causes based upon some positive claim about what *must* be the case for something to count as a cause. But it then turns out that this positive claim is deeply puzzling: first, because it seems to be fairly implausible; and second, because the author appears to offer no positive account of causation to back it up.

A broader look at each philosopher's work suggests that something has gone wrong here. For in each case, it is easy to find instances where they themselves discuss cases that seem to provide clear counter-examples to their supposed positive claims. For instance, Anscombe explicitly notes that she is not denying that both beliefs and intentions can play causal role in bringing about particular actions. One example she gives is of a standing intention not to talk to the press—this, she says, might explain her refusal to do so on any particular occasion:

> [S]uppose I have a standing intention of never talking to the Press. Why, someone asks, did I refuse to see the representative of *Time* magazine? – and he is told of that long-standing resolution. 'It makes her reject such approaches without thinking about the particular case.' This is 'causal' because it says 'It makes her. . . ': it derives the action from a previous state. [5, 95]

Nothing in this causal story requires that the intention be a 'mental cause' in the sense we have outlined, e.g. that the agent *recall* or *think of* her intention when asked to give an interview. Her habit might be so ingrained that, on being told that the representative from *Time* magazine was here, she simply responded 'Send him away' without giving it further thought. The standing intention could still be cited as a cause of her refusal, this needn't involve any event—in her consciousness or otherwise—that we would call a manifestation of the intention, besides the instruction to send the reporter away: she simply hears how things are, and responds without thinking about it.

It is also not difficult to find Wittgenstein discussing cases which seem to involve mental causes that are 'known without observation' – for instance, in his notebooks in the passages collected together under the heading *Cause and Effect: Intuitive Awareness*, e.g. " If someone says: "I am frightened, because he looks so threatening"—this looks as if it were a case of recognizing a cause immediately without repeated experiments" [39, 371].[8] In the next

[8] The dialectic that shapes that discussion is rather complicated, but it is related to the points I make

chapter I shall also suggest that his discussion of reading in the *Philosophical Investigations* provides further such examples. Given these points, combined with the implausibility of the claims attributed to Wittgenstein and Anscombe, and the lack of supporting argument for them in the rest of their work, it makes sense to look for an alternative way to understand these passages.

4.2 Confusion About Grammar

We can take our clue from the further claims that Wittgenstein makes in *The Blue Book* passage about reasons and causes. Earlier I suggested that that passage commits him to the following four claims:

1. In responding to the question 'Why did you do such-and-such?', one can *either* give a reason *or* a cause for one's action. This reflects an ambiguity in the question 'Why?' which can be asked in either sense (i.e. in a sense that calls for one kind of response or the other).

2. One mark of the difference between a reason and a cause is that one can conjecture about causes—such conjectures being founded on and confirmed by repeat experiences—whereas one 'must know' the reasons for one's actions just insofar as one acts.

3. These points reflect the 'grammar' of our talk about *causes* and *reasons*. Statement that marking this difference, such as "we can only conjecture the cause but we know the motive [or reason]", are grammatical ones, and they describe 'logical possibilities'.

4. Unclarity about these grammatical differences produces philosophical confusions: specifically, the claim that a motive (or reason) is a cause of which we are 'immediately aware'.

in the next section about the role that appeals to conscious or felt experience play in both Wittgenstein and Anscombe's discussions.

To understand what Wittgenstein and Anscombe mean to be doing in their discussions of reasons and causes, and how these bear on the philosophical questions (MQ) and (EQ) that shape contemporary accounts, we need to better understand claims (3) and (4) and their appeal to 'grammar'. I suggested in §1 that ultimately Wittgenstein and Anscombe hold that clarity about 'grammar' will undermine our sense of philosophical questions the (MQ) and (EQ), and the urgent pressure we feel to formulate a philosophical account in response to them. Here I shall argue for three more restricted claims related to (3) and (4) above: (a) that both Wittgenstein and Anscombe see certain kinds of responses to these philosophical questions as rooted in grammatical confusions; (b) the positive claims that these arguments attribute to Wittgenstein and Anscombe are actually an attempt to give voice to confusions they find in the accounts they are responding to; and (c) that they take it that these confusions mean that these accounts cannot provide their proponents with what they think they need to answer philosophical questions such as (MQ) and (EQ).

4.2.1 Grammatical Confusions

Wittgenstein's description of the claim "we can only conjecture the cause but we know the motive" as a 'grammatical statement' should be connected to our earlier discussion of grammar and philosophy in §3.2. At this point our most immediate concern is with how specific responses to philosophical questions might reflect a confusion about grammar, and as such fail to provide what their proponents are looking for. One way in which a claim could reflect a confusion between grammars is if it tries to force one case to fit the grammar of another: if, for instance, I responded to the claim "The tree struck the window" by asking "On purpose or by accident?". This, I take it, is one of the reasons Wittgenstein compared such confusions to grammatical jokes.[9] Here is an example of one such joke:

[9] *Philosophical Investigations* §111:

> The problems arising through a misinterpretation of our forms of language have the character of depth. They are deep disquietudes; their roots are as deep in us as the forms of our language and their significance is as great as the importance of our language.—Let us ask ourselves: why do we feel a grammatical joke to be deep? (And that is what the depth of philosophy is.)

> There is a man on the shore of a lake, and he can't get to the other side of the lake: the mountains come steeply down into the lake and he has no boat and the water is too cold to swim in. But someone comes into view on the far side, and our man shouts: Yoo hoo, how can I get to the other side of the lake? and the other person shouts back You are on the other side of the lake![10]

The surface form of the phrase 'the other side of the lake' is the same as phrases such as 'the eastern shore of the lake', and it makes perfect sense to imagine someone responding to the question 'how can I get to the eastern side of the lake?' with the response 'you are on the eastern shore!'. The joke depends on the speaker acting as though this grammar applied to the original phrase.

We can extract two points from this that we can look for in the philosophical case:

1. A surface form of particular expressions that is apparently similar.

2. A feature of the broader grammar one expression that does not belong to the other.

My suggestion is that this is how Anscombe and Wittgenstein understand some general tendencies in the discussion of reasons and causes. They both note that the question 'Why?' is ambiguous between different senses: among them some that ask for a cause for one's action, and some that ask for a reason. One would be guilty of a grammatical confusion if one took features that applied to one kind of response to the question 'Why?', and insisted that they apply to the other as well.

It is precisely this kind of confusion that Anscombe and Wittgenstein are concerned to diagnose in the passages cited in §4.1. The form of their argument is therefore not the one we identified earlier:

1. ⟨Positive claim about what causation must be like⟩

[10] This joke was provided as an example by Cora Diamond at the 5th ALWS Wittgenstein Summer School in Kirchberg. My commentary above is directly informed by her remarks.

2. ⟨Claim that explanation by reasons lacks this feature⟩

3. ⟨Conclusion that reasons cannot be causes⟩

The mistake lies in seeing the claims they put forward as positive assertions about what causation or acting for a reason *must* involve, assertions that they mean to fully endorse as premises in an argument. My suggestion is that we should instead see them as identifying specific features of causality that do have some general application, and that one *could* wrongly take to apply to any explanation by reasons, if one confused the relevant grammar of causality with the grammar of reasons and sought to provide a general and substantive account of either.

The arguments quoted above are supposed to be directed against attempts to answer our philosophical questions (MQ) and (EQ) that take precisely this form, and the arguments themselves are therefore primarily diagnostic. The key point is not to make and defend a positive claim about causality or reasons in their own voice, but to diagnose the role that a confusion about some such claim plays in their interlocutor's arguments. Beyond this diagnosis, they also involve the further claim—which I shall only defend in outline in this chapter—that an attempt to answer these questions that is beset by such confusion *cannot* provide what the proponent thinks she needs for an adequate response to those questions.

The form of their argument is therefore closer to the following:

1. Your view makes you insist that explanation by reasons *must* have this feature.

2. But here is an explanation by reasons that does not have this feature.

Their diagnostic move is to claim that (1) is the result of taking a particular picture of how explanation by causes works, and trying to apply it to all cases of explanation by reasons.

In the rest of this section, I will show how this template can help us make sense of the passages we considered earlier, in part by showing that this helps us make sense of how they relate to other parts of each philosopher's work. I begin by relating Anscombe's discussion of causation to attempts to answer (MQ), showing why Anscombe takes responses that appeal to 'causal efficacy' to relate reasons to action to involve a grammatical confusion, and why this confusion means that any such account must fail. I then go on to more briefly show how aspects of both Anscombe' and Wittgenstein's discussion relate to attempts to answer (EQ) and attempts to account for our knowledge of our reasons for acting..[11] Then, in §4.3, we will look at what it would take for a philosophical account to fully absorb Wittgenstein' and Anscombe's arguments.

4.2.2 Anscombe: Causation and Grammar

Let's begin by looking more closely at (MQ) and its relation to some of Anscombe's explicit discussions of causation. As we saw in §2, philosophers often appeal to causation as part of an account of *what it is to act for a reason.* I suggested in the previous chapter that Davidson's Question ('what is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing it') could be understood as a request for such an account, and Davidson's response as a proposal that to act for a reason *is* for one's actions to be caused by a pair of psychological states. Thus, on Davidson's account, when we ask the question 'why did so-and-so do that', and what we are looking for is the person's reasons for acting, there is a sense in which the response to our question describes a form of causality: it presupposes that psychological states with that content were involved in the causation of the act in question.

We have also seen that Anscombe's response to Davidson is rather abrupt, and that the argumentative support she provides for it seems to depend upon some tenuous assumptions about mental causation. However, in light of the above, we can try to read Anscombe's response in a different way: not as arguing from a pair of positive claims about *what it is*

[11] This will only provide a preliminary sketch of the details of their arguments, which will be developed through the more detailed discussion of Wittgenstein's treatment of *reading* in the next chapter.

to act for a reason and *what it is to be a mental cause*, but rather diagnosing a tendency in Davidson to misapply the grammar of a particular causal case in his treatment of acting for a reason, together with a positive claim that such a confusion guarantees that his account can never provide him with what he wants from it.

To see how this might work, we need to first look at some of Anscombe's broader discussion of causation and action. It follows from the general observations above that thought and talk about causation will also have distinctive grammatical features: the fact that we are concerned with causality will be reflected by the kind of questions we ask, and the kind of answer that counts as an appropriate response. As we have seen, one kind of thing we can be asking for in posing a particular 'Why?'-question is what caused something to happen, and this will be reflected in what we count as an appropriate response to our question.

These ideas provide important background to understanding Anscombe's approach to the topic of causation in her paper *The Causation of Action*. The paper begins begins with a number of responses to the question 'What made the door shut?', accompanied by some "further questions and interests [that might be] naturally aroused by the different answers to the first, 'simple' question" [5, 90]. (Answers include "The wedge was propping the door open and got removed", "A blast of air blew it shut", and "Its own weight causes it to shut".)

It is important to note that one could rewrite Anscombe's question as 'Why did the door shut?', emphasizing the initial ambiguity in our use of 'Why?' that Anscombe means to partially clarify. Though this rephrasing needn't obscure the sense of the question, it does leave implicit something that was explicit in the original formulation, and that Anscombe takes to be an important feature of the grammar of causation: that an appropriate answer will indicate something *that made the door shut*, i.e. a cause from which that effect was derived.[12]

[12] This point is further developed through attention to the follow-up questions prompted by interest in the initial response. Each answer prompts a different set of questions, which despite their particular differences fit into two generic patterns: 'How did that come about?' and 'How does that work?'. As with the initial question, 'What made the door close?', these are both further possible disambiguations of the question 'Why?' that reflect the fact that we are concerned with causation. In asking the former question, "we are interested in picking out 'chains' of causality going back in time", whereas with the latter we are

It is, in fact, an important part of Anscombe's point in these papers that there might be a variety of appropriate answers to a question like 'What made the door close?', and that different answers might involve different kinds of causation. Indeed, part of what it is to have a 'highly general' concept such as 'cause' is to have an array of more specific causal concepts in one's vocabulary,and there needn't be a single specific feature that is shared by all the happenings that could be described by such concepts in virtue of which they count as instances of causation. [13] Nevertheless, despite the differences between each case, Anscombe takes it that there are also some essential similarities brought out by the form of her original question. As she puts it in "Causality and Determination", her inaugural lecture at the University of Cambridge:

> [C]ausality consists in the derivativeness of an effect from its causes. This is the core, the common feature, of causality in its various kinds. Effects derive from, arrive out of, their causes. [1, 136]

interested in understanding "the connexion between *established* links" in such a chain [5, 92-3]. Thus, to take Anscombe's first example, we can imagine the conversation:

> Why did the door shut? – That apparatus closed it? – Why?

continuing either with

> I pressed the button that activated it.

or

> It contains a spring-loaded bolt that pushes the door shut.

Both indicate that we are concerned with the causalities involved in the door's closing, but each question takes that inquiry in a different direction.

[13] Anscombe notes nine different possible answers to her question, and an range of follow up questions:

> All the answers are perfectly appropriate. We can pick out from among them those which name causes which act on the door: the artificial mechanism, the wind, the dog, the projectile, the magnet, the human. By contrast, the removal of the wedge was a 'causa removens prohibens' – a cause that removes a hindrance; the creation of a vacuum produced an imbalance of air pressure, as a result of which what moved the door – acted on it – was the air on the other side. And what would one say about the weight of the door, which caused it to shut 'of itself'? Neither that the weight moved the door nor that it did something that led to something else's moving it. [5, 91]

The responses thus involve an array of causal concepts: the wind blew it, the dog pushed it, the projectile struck it, the magnet attracted it, etc.

This, I take it, is Anscombe's way of bringing out a unity she sees in our various causal concepts.[14] It is precisely this feature that she highlights in her discussion of 'mental causes' (given in answer to the question 'what *produced* this action or thought or feeling on your part' [4, §11]), as well as her examples of when we would say that an intention or belief caused an action ("This is 'causal' because it says "It makes her. . . ": it derives the action from a previous state" [5, 85]).[15]

The important thing to note is that, in all these cases, an answer to the question 'why?' that is concerned with causality will point to *something* contemporaneous with, or prior to, the target of the question, which is represented as its cause. In some cases, we may be interested in tracing such causal interactions backwards; in others, we are concerned to characterize how a particular such interaction worked; but either way, in describing the causal interaction we are describing *something* occurent or actual at the time.

Misapplying the Grammar of Causation

Anscombe's general discussion of causation thus involves the following claims:

1. The unity in a (the?) notion of causation lies in the idea of the derivativeness of effect from cause.

2. The general idea of derivativeness of effect from cause encompasses *various* kinds of

[14] Note that it would be a mistake to see Anscombe as identifying some single and specific common feature—'derivativeness of cause from effect', say—that is shared by all instances of causation. This is reflected in the fact that the sense in which cause is derived from effect will vary from case to case: e.g. we needn't think of the way in which it is 'derived' from the removal of the wedge as the same as the way it is 'derived' from the movement of the wind. The "common feature" is only characterized in abstract terms, and what it specifically amounts to will vary from case to case.

[15] This might seem to go against my claim above that we should understand Anscombe as putting forward a positive claim about what causality *must* involve—after all, isn't she here claiming that causality *must* involve some idea of the derivativeness of an effect from its causes? It is unclear to me whether Anscombe would indeed endorse this claim, though I will discuss the point further in §4.2.2. Since Anscombe explicitly recognizes the four kinds of kinds of cause that Aristotle describes, and further notes that more are needed, her answer must partly depend on whether it makes sense to describe all of these as involving this idea of derivativeness. I suspect that she would find it inappropriate or misleading to characterize e.g. teleological causation in this way. This would suggest that her description here is meant to characterize a narrower notion of causation, e.g. notions of efficient causation that have been central to analytic philosophy.

causality, not a single type.

Though (1) strikes me as a plausible characterization of what unifies at least some of our causal concepts, and (2) strikes me as an important but neglected consequences of both Wittgenstein and Anscombe's work, I shall not be concerned here to defend either of these claims. What matters for our purposes is that her account suggests two different possible locations for philosophical confusions. First, one could take features of *one specific instance* of this generic kind of causality, and try to apply them to all others: this would involve taking one particular way in which an effect might be understood to be derived from its cause, and claiming that all other cases of derivativeness of effect from cause should be understood on this model.[16] Second, one could take some feature of this generic kind of causality, and try to apply it to other 'non-causal' cases to which it did not fit.

It is this second form of confusion that is most directly relevant to our concerns, since it is the one engendered by the ambiguous use of the question 'Why?' to ask for reasons and to ask for causes. If I am right, Anscombe's point is that the grammar associated with derivativeness of effect from cause does not apply to *a significant range* of answers to the question 'Why?' that give a person's reasons for acting. This is because she takes it that it is a feature of that grammar that the causes it describes are occurent or actual features of what happened at the time, whereas the grammar of some descriptions of reasons for action needn't have this character. Her point is therefore that some cases of explanation by reasons do not have this feature, and therefore:

3. Explanations by reasons—that is, explanation that represents its target as done *in pursuit of an end* or *on grounds of something believed*—does not have to have the same grammar as explanation by causes (though the common use of 'Why?' might lead us to assume that it does).

[16] Though I will not defend this claim, I suspect that Anscombe would see accounts that attempt to categorize causes into a single metaphysical category (e.g. events, states, etc) as making some such confusion.

Furthermore, any attempt to apply this grammar in the service of answering philosophical questions about acting for reasons cannot work. Specifically, if one claims that acting for a reason consists in some specific causal interaction, and one applies the grammar of causality described above, one will be compelled to find something occurent or actual, prior to or contemporaneous with the act, to count as the relevant cause. Anscombe's claim is that anything one could find that fit these criteria could be equally present in a case that was not an example of 'acting for a reason', and so cannot be what acting for a reason consists in. In trying to apply this notion of causation in service of answering our philosophical question, we are led to look for something characterized by a grammar that is not that of 'acting for a reason'. As a result, anything we can describe that fits the grammar we are trying to apply cannot be what is essential to acting for a reason.

Anscombe on Davidson

It is this error that Anscombe identifies in Davidson's proposal to explain what it is to act for a reason in terms of the causal efficacy of psychological states. Using the general terms above, we might describe Davidson's view of causation as one on which one event can be shown to be appropriately related to (i.e. derived from) another according to some physical law. The key point is that any instance of causation therefore requires actual and occurent *events* to play the role of causes. Thus, when Davidson suggests that to act for a reason is for one's action to be caused by particular psychological states, he is committed to the idea that there *must* be events that play the role of causes. Indeed, he suggests that "[i]n many cases it is not at all difficult to find events very closely associated with the primary reason" such as "the onslaught of a state or disposition" [15, 12]; and even when there is no clear candidate for such an event, we can nevertheless know that there must have been one, as when "we are ignorant of the event or sequence of events that led up to (caused) the collapse, but we are sure there was such an event or sequence of events" [15, 13].

Anscombe's response is two-fold: first, reporting one's reasons for acting does not necessarily involve reporting anything that happened at the time of the action, and to this extent lacks the grammar of causality; but more importantly, anything one cared to describe that did fit the grammar of 'events' could be equally present in a case that was not an example of acting for a reason. For of any event that we represent as causing the action (and, indeed, of any occurent or actual cause from which we represent the action as derived), we can always ask "But was the act done for the sake of the end and on the grounds of the belief?"—this because a cause describable in the same way can always be imagined in a scenario when this was not the case.

The most important claim here is therefore that the grammar of responses to the question 'Why?' that give a person's reasons for acting is not the same as the grammar of responses that give causes for action. The latter involves describing something that led up to or produced the action, and doing this is not the same as describing a person's reasons for acting.[17] The force of the question that Anscombe poses to Davidson is that it is supposed to bring out this difference in grammar. For giving someone's reasons involves, for instance, describing the end for the sake of which the action was undertaken, or showing the grounds the agent had for undertaking it. Answers to the question 'why?' that serve this role do not generally do so by representing some part of a causal chain that produced the action; they work in some other way.

4.2.3 Wittgenstein: Experience and Knowledge

Though the above sketch aims to explain the form of Anscombe's response to Davidson, it does not touch on the issue of why she seems to assume that mental causation must involve felt or conscious experience. This is because I think we need to distinguish two different parts in Anscombe's diagnosis:

1. If we are to understand belief or intention as causes of a person's action, applying the

[17] I shall later argue that even when we are describing reason-based responses, descriptions that emphasise the causal character of those responses will be precisely those that can also be applied to other kinds of response as well.

grammar of causality that we do in other cases, then we will be led to assume that there must be something (e.g. an event) either identical with, or somehow intrinsically related to, the belief or intention that led to the relevant action.

2. If we are to understand belief or intention as the causes of a person's actions, and if we are to claim that the agent has knowledge of that causal role just insofar as she acts, then we require some account of the immediacy of that knowledge that fits with the picture of causality we are applying. Sensations and felt experiences are one kind of event of which we might be said to have immediate knowledge.

Anscombe runs these points together because she assumes (with most contemporary accounts) that any purported response to (MQ) will also aim to provide resources for a response to (EQ) as well: it will try to explain how it is that we come to have knowledge of our reasons for acting. Nevertheless, it can be helpful to distinguish the two questions, since it will turn out that for both Anscombe and Wittgenstein, each brings with it its own confusions. In this section I shall briefly argue that Anscombe's reference to felt or conscious experience can be understood as an attempt to develop the line of thought that Wittgenstein expresses in the passage from *The Blue Book* discussed in §4.1.1. This argument will then be developed in more detail in §5. With this in mind, let's return to that passage and see how it might fit with the strategy I have characterized in this section.

Knowledge and Justification

Part of what is confusing about the original passage from *The Blue Book* is that Wittgenstein appears to endorse the following implausible claim: knowledge of causes must be based on conjecture from repeat experience. My suggestion for clarifying the passage is that, rather than reading this as a claim that Wittgenstein means to be putting forward in his own voice, we should hear it as something he thinks that his interlocutor wants to endorse that will compound the confusion between reasons and causes. This is because, given this picture of how one generally grounds one's knowledge of causal claims, it becomes difficult to see how one can explain the knowledge we have of our reasons for acting, especially if we also picture those reasons as the causes of our action. His point is thus not: knowledge of causes is always based on conjecture, whereas we must know our reasons for acting, so those reasons cannot be causes. Rather it could be put as: you take knowledge based on conjecture as your paradigm of causal knowledge; but knowledge of our reasons is not based on conjecture; so you will find this knowledge mysterious.[18]

To see how this works, it will be helpful to first imagine a broader epistemological project, before turning to particular attempts to answer (EQ).[19] Suppose, first, that one felt a need to provide a philosophical account of our knowledge of causes and causal relations that (a) showed that we were justified in that knowledge, or even (b) showed how such justified knowledge was so much as possible. Suppose further that, feeling this pressure, one began from examples of causal claims that are based on conjecture. As we noted above, it is a feature of such claims that they are subject to the question 'Why do you think that?' or 'How do you know that?', where that question asks for the *grounds* of such conjecture, and one possible response involves citing repeat experience. A philosophical account might then aim to explain (a) how grounds offered in response to this question could justify particular causal claims, or even (b) how it was so much as possible to take repeat experience as a ground for such a claim.

This project might involve its own difficulties and confusions. But suppose now that, as

[18] Admittedly, if this is the line of thought in these paragraphs, it is expressed in a highly compressed and misleading manner. My interpretation of Wittgenstein's discussion should provide a clearer example of him making an argument of this form, regardless of whether he saw things in exactly these terms at the time *The Blue Book* was dictated.

[19] Though I shall not base my argument on this claim, it seems likely that Wittgenstein has Russell's paper "The Limits of Empiricism" [31] in mind in these sections. See the editors' comments on [39]. Russell's paper (and Wittgenstein's engagement with it in the notes collected together as [39]), involve a much broader problematic concerning causation and knowledge. In his notes, Wittgenstein appears to be interested in a claim emerging from Russell's work to the effect that (a) empiricism cannot so much as make sense of our knowledge of causes, and (b) we need to appeal to the idea of 'intuitive awareness' of causes to ground such knowledge. The dialectic I sketch above is a simplified version of this broader problem, focusing in particular on knowledge of 'mental causes' and reasons.

part of this project, one also wanted to provide an account of our knowledge of the 'mental causes' of our action, and one assumed that this too must involve (a) finding grounds for our knowledge, and (b) showing how they could be grounds for our knowledge. Here we face a particular problem, because the basic form of the proposed general account does not fit our more specific range of cases, since it is a mark of our knowledge of 'mental causes' that it is not based on conjecture. For instance, if you prick me with a pin, my knowledge that the pinprick caused the pain I feel in my finger is not (usually) grounded in conjecture or repeat experience. This means that there will be no further grounds to which one could appeal by way of explaining this knowledge, beyond the fact of my feeling pain in response to the prick—that is, nothing I would cite as grounds for my taking the prick to be the cause of my pain. Given the general picture of what it is to explain how it is we come to have knowledge from which we started, this now looks like a troubling case: we need to find grounds for that knowledge, but no such grounds are available.

It is at this point that we can imagine our philosopher making a particular kind of appeal to the idea of 'felt experience' of 'mental causes': that is, to think of them as "a cause of which we are immediately aware, a cause 'seen from the inside', or a cause experienced", as Wittgenstein puts it in *The Blue Book*. For if our philosopher is to understand this knowledge on the picture she has of knowledge from conjecture, it seems that what she needs is to show that we have grounds that can explain our knowledge. Her problem is that no such grounds appear to be available. But our example provides her with another kind of case on which she can base her response: my knowledge than I felt a prick. This is also knowledge that seems to lack any further grounds beyond the particular sensation in question; but here our philosopher could understand the sensation as *itself* providing grounds for our knowledge of it: we are aware of the fact that we are in pain *in virtue of* our sensation of pain.[20] Applying the picture of knowledge by conjecture—i.e. knowledge

[20] This again involves applying the grammar of knowledge based on grounds to knowledge of sensation; ultimately philosophical clarity here will involve undermining the impulse to apply this picture of what knowledge *must* involve (i.e. grounds for one's claim) to every case of knowledge. See [11]

that is based on independent grounds—knowledge of sensation now looks like a special kind of case: for it looks as though the sensations themselves provide grounds for our knowledge of them. Knowledge of sensation is thus, in a sense, *self-grounding*, since we have grounds for such knowledge just insofar as we experience the relevant sensation.

This picture can then be extended to account for our knowledge of mental causes: we don't simply experience the pain, but also experience the cause ('experience the because' – cf. PI §177). To say that we have 'felt experience' of the causes of our acts could thus be taken to mean that this knowledge is also *self-grounding*. The grammar of the initial examples of causation has been made to apply to knowledge of sensation and then to knowledge of mental causes, albeit in a special way.[21]

[21] We will look more closely at why this proposal cannot give the philosopher what she wants in subsequent chapters: ultimately the grammar of knowledge by conjecture does not apply to *either* of these two cases. But it it worth noting now that this kind of proposal involves a further confusion of grammars. For the model our philosopher takes for her account of self-grounding knowledge is experience of sensation: if I have a particular sensation such as a pain, I can give expression to that pain in a self-ascription just insofar as I experience it. This gives us a partial parallel with the 'mental causes', which is what suggests 'experience of the cause' as a way to characterize self-grounding knowledge.

However, it is important to see that the parallel is only partial, and that the grammar of one case does not apply directly to the other. For instance, one feature of knowledge of sensation is that we do not apply the contrast between knowing and merely thinking that you know that you have some sensation: if I am suffering a pain in my side, I can 'just say' as much without my claim being based on observation. But whereas it *does* make sense to describe someone as 'merely thinking they knew the (mental) cause of their pain', it is less clear that it makes sense to describe someone as 'merely thinking the knew where they felt pain'. Wittgenstein provides an example of the former in one of his discussions of the claim that we have 'intuitive awareness' of causes:

> Don't we recognize immediately that the pain is produced by the blow we have received? Isn't this the cause and can there be any doubt about it?—But isn't it quite possible to suppose that in certain cases we are deceived about this? And later recognize the deception? It seems as though something hit us and at the same time we feel a pain. (Sometimes we think we are causing a sound by making a certain movement, but then realise it is independent of us.) [39, 373]

That such a scenario—however implausible—makes sense to us marks a contrast between self-ascriptions of pain, and self-ascriptions of the cause of one's pain. This can be seen in the difference between the scenario described by Wittgenstein, and the kind of case that Anscombe describes here:

> [It is not the case that] the place of pain (the feeling, not the damage) has to be accepted by someone I tell it to; for we can imagine circumstances in which it is not accepted. As e.g. if you say that your foot, not your hand, is very sore, but it is your hand you nurse, and you have no fear or objection to inconsiderate handling of your foot, and yet you point to your foot as the sore part: and so on. . . . [H]ere we should say that it is difficult to guess what you mean. (§8)

Knowledge of Our Reasons

With this background in place, we can now return to attempts to answer (EQ): that is, to explain how it is that we have knowledge of our reasons for acting. Here too we have an example of first-personal knowledge of which we would not normally enquire 'how do you know that?'—as Wittgenstein puts it in *The Blue Book* [38], one often says "Surely I must know why I did it" (BB15). If one tried to explain our having this knowledge on the model of knowledge from conjecture, one would face the same problem: there appear to be no grounds to which one can appeal to justify the knowledge, since one has it just insofar as one acts. Thus, here too one might seem forced to claim that the knowledge is *self*-grounding.

Suppose now that our philosopher also claimed that our reasons for acting were the causes of our action. They would then need to provide an account of the relevant cause that explained how our knowledge of it could be self-grounding. Since the paradigmatic model for such knowledge is supposed to be sensation, one can use the idea of 'felt experience of the cause' as a summation of its key features. Of course, a sophisticated philosopher could propose a more complicated account of how such knowledge could be self-grounding[22]—but even if it didn't involve an appeal to 'felt or conscious experience', its basic form would be the same insofar as it deployed the idea of a self-grounding source of knowledge. As we shall see, this means that a philosopher needn't endorse the 'bad philosophy of mind' that associates self-knowledge with sensation or felt-experience to end up with a version of the view Wittgenstein is criticising.

The dialectic I have sketched here is more complicated than the one described in §4.2.2, because it compounds two distinct confusions: first, a picture of what an account of knowledge must be like, and an attempt to fit all first-personal knowledge into that picture; and second, an attempt to understand reasons for acting on the model of causes. These two distinct sources are implicitly described in the original argument from *The Blue Book*:

[22] See the discussion of [33] in §6.

> The double use of the word "why", asking for the cause and asking for the motive, **together with** the idea that we can know, and not only conjecture, our motives, gives rise to the confusion that a motive is a cause of which we are immediately aware, a cause 'seen from the inside', or a cause experienced. (BB14 – my bold)

We can understand the original passage as a compressed description of **two** distinct locations for possible philosophical confusion: the double use of the word "why", *and* our application of 'know' to describe both knowledge based on conjecture and other kinds of knowledge. Each of these is a site of possible philosophical confusion: one could confuse the different grammars of 'why', i.e. the grammars of talk about reasons and causes; and/or one could confuse the grammars of various kinds of knowledge, i.e. knowledge of causes, knowledge of one's sensations, knowledge of one's intentions or reasons for acting, etc.[23] Wittgenstein then sees two such confusions in the kind of case he is concerned with: *if* one confuses the grammars of reasons and causes in general, and *if* one takes knowledge from grounds as one's picture of every kind of knowledge, *then* in coming to explain our knowledge of our reasons one will be tempted to treat them as causes of which we are immediately aware.

4.3 Clarity About Grammar

As stated in §4.1, there are two common ways of responding to these discussions in the contemporary literature. The first, which tends to go with the view sketched in 4.1 that these discussions put forward arguments based on implausible claims, attempts to show that we *can* in fact apply some generic notion of causality as part of our response to philosophical questions such as (MQ) and (EQ). If Wittgenstein and Anscombe are correct in their general diagnosis, then we would expect it to be the case that any proposal that such accounts put forward to explain *what it is to act for a reason*, or *how it is we come to have knowledge*

[23] Wittgenstein occasionally says things that suggest that he thinks it is a mistake to describe all of these as knowledge, e.g. at §246 he says "It can't be said of me at all (except perhaps as a joke) that I *know* I am in pain." A more careful discussion of this topic would need to accommodate these sections.

of our reasons will ultimately fail to provide what its proponents think they need from it. We can further expect that the reasons for its failure can be directly linked to the fact that they try to fit a particular instance of the grammar of causality to instances of acting for a reason. I shall argue that this is true of the causal-psychological accounts that will be our topic in §6.

The second kind of response suggests that, though Anscombe and Wittgenstein raise insurmountable problems for a specific conception of causality, an expanded notion can both accommodate their arguments *and* make room for the claim that rational explanation is a form of causal explanation. Philosophers that can be put together in this category understand Anscombe's arguments to be directed against attempts to explain what it is to act for a reason by an appeal to 'Humean' or 'efficient causation'. While they agree that this strategy cannot work, they claim that this is not grounds for an outright rejection of the claim that rational explanations are causal explanations, and therefore (perhaps) that what it is to act for a reason can be explained by an appeal to some form of causation. For (they argue) we need not restrict ourselves to the 'Humean' account, or even to 'efficient causation'—instead, we should challenge the claim that all causation fits this model, and look to provide accounts of other kinds of causality to which we can then appeal in answering questions like (MQ).

An early version of this line of thought can be seen in Peter Winch's preface to the second edition of his book *The Idea of A Social Science and Its Relation To Philosophy.* Reflecting on his earlier work—which had been part of the Wittgensteinian orthodoxy that Davidson had been reacting against—Winch states:

> I found myself denying that human behaviour can be understood in causal terms, when I should have been saying our understanding of human behaviour is not elucidated by anything like the account given of 'cause' by Hume (and Mill). [37, xii]

This point has been expanded in contemporary work by people like John Hyman and

Eric Marcus. Hyman suggests that "Wittgenstein and his followers faced a choice between challenging the Humean theory and excluding causation by desires, and they made the wrong choice" [23, 113] – rather than rejecting the idea that the Humean picture provided an adequate account of causality, they instead simply denied that desires could be understood as causes. Hyman suggests that Anscombe' and Wittgenstein's best insights can be preserved and extended by coming to see desires as a kind of teleological disposition that is manifested in our intentional actions.[24] Explanations that appeal to desires or intentions, or otherwise represent us as acting for the sake of some end, could then be understood as akin to other kinds of causal explanation that rely on an appeal to dispositions:

> [A]n explanation of an intentional act that refers to the desire the act expressed or to the intention with which it was done is both causal and teleological. It is causal because it refers to a disposition, and it is teleological because the kind of disposition it refers to is a disposition to pursue an aim, in other words, a disposition that is manifested in goal-directed behaviour. [23, 130]

Marcus also represents his project (which will be the topic of §7) as an attempt to combine the best insights of an Anscombean approach with an extended notion of causation:

> The real dispute between Davidson and the anti-causalists is whether rational explanations are made true by efficient causal connections or by some other kind of worldly relation. Framed thusly, I take the part of the anti-causalist, defending

[24] A particular advantage of an appeal to a teleological disposition is that it seems to provide the resources for a response to the problem of causal deviance. For once we allow ourselves to appeal to dispositions, we can characterize the dispositions that play a role in rational explanations in a way that rules out deviant cases. Since these philosophers understand the key mark of such cases (e.g. Davidson's climber example) to be that the deviant 'action' was not done *for the sake of* the end, they simply characterize the relevant disposition (and its acts) in a way that makes it clear that they are directed towards pursuit of an end. Thus Hyman argues that desires are dispositions that are "manifested in purposive or goal-directed behaviour . . . aimed at satisfying the desire", and that rational explanations that represent the agent as pursuing an end depend on an appeal to such dispositions . Marcus offers a more complicated account, claiming that such explanations depend on an appeal to an ability "to do what is to be done on the basis that something else is to be done" [24, 168]. In both cases, since the dispositions are only manifested in goal-directed behaviour, we have grounds for explaining the difference between paradigm and deviant cases: the former, but not the latter, are acts of the relevant disposition.

> what I take to be a broadly Anscombian point. However, as Winch ultimately realized, the best way to make this point is to argue that there are others kinds of causation besides efficient causation. [24, 163]

As we saw in §2.3.3, the heart of his account is a distinctive kind of 'rational causation', whose instances consist in acts of 'rational abilities' which (in the practical case) can be thought of as a kind of teleological disposition to do what is to be done because something else is to be done.

Thus, despite differences in the specifics of their accounts, the following provides a generic outline of the main claims that characterise this approach:

1. Explanations that appeal to dispositions and abilities are causal explanations.

2. Humean (or efficient) causation provides inadequate resources for an account of (all) such explanations.

3. An expanded notion of causation will allow us to include the notion of dispositions in our understanding of causality.

4. We can provide an account of a particular kind of disposition involved in rational action (a desire, a distinctive representational ability, etc.).

5. We can then see rational explanations as appealing to dispositions, and therefore as an instance of causal explanations.

4.3.1 Potential Problems For This Approach

Broadly speaking, then, this seems to be a promising approach—one that might preserve some of the insights Wittgenstein' and Anscombe's discussion of possible confusions involving notions of 'reasons' and 'causes', while also fitting with comments such as the following (originally made in a different context, but apposite here):

> Say what you choose, so long as it does not prevent you from seeing the facts. (And when you see them there is a good deal that you will not say.) (PI §79

The general spirit behind this approach might then be described as follows: if we can provide an account of acting for a reason that doesn't simply take instances of 'derivativeness of an effect from its causes' as its model, why shouldn't we call this a causal account? As long as we are clear about the differences between this kind of causation and the kind Anscombe was concerned with, we won't make the mistake of insisting that the one *must* have features that belong to the other.

Though I would not reject the spirit of these remarks, I think full consideration of Wittgenstein and Anscombe's arguments suggests that making good on this insight will involve its own difficulties. To do justice to these arguments, we need to take full measure of the ways in which explanation by reasons is both different from, and related to, explanation by other kinds of causality. An account could fail to do this is two ways: first, by failing to adequately bring out how reasons-explanations, and the dispositions involved in them, are different from other kinds of causal explanation; and second, by making this difference seem too great, so that it looks as though the acts of these dispositions are entirely outside the familiar causal order.[25] Here I shall briefly outline these two potential pitfalls. This will help prepare the ground for our examination of a more considered treatment of some of these issues in Wittgenstein's discussion of reading, and for a consideration of normativist accounts in §7.

Failure to Characterize the Difference

The general difficulty here lies in doing justice to the idea that the role of our reasons is importantly different from that of other kinds of causes, while nonetheless characterizing

[25] Indeed, though these may sound like contradictory failings, I shall suggest that they are in fact complementary: insisting on the differences between rational causation and other kinds of causality only serves to make the former seem somehow mysterious and unexplained, and an account that fails in one regard will tend to fail in the other as well.

rational explanations as appealing to a kind of causality. The first pitfall I will be concerned with arises from basing one's account entirely on an analogy with other kinds of disposition or causality. For example, an account might begin by pointing out that explanations like

[**D1:**] The vase broke because it was fragile

are genuinely explanatory, and intrinsically related to causal explanations of the kind Anscombe was concerned with, such as

[**D2:**] The vase broke because it was dropped.

The suggestion is that, just as causal explanation [D2] involves essential appeal to a particular kind of disposition, so too an explanation like

[**I2:**] James went to church to please his mother,

also involves appeal to some disposition, and that disposition provides the basis for seeing this as a causal explanation.[26]

The difficulty I want to highlight is that the analogy breaks down at a key point. The disposition described in [D1] is a causal disposition in the familiar sense, as is shown by the fact that [D2] represents what is described in its *explanandum* as derived from what is described in its *explanans*. [I2] differs from [D2] precisely insofar as it doesn't necessarily involve the idea of a trigger or stimulus, as reflected in the related explanation [D1].

Acknowledging this difference might not seem to pose any deep problem for such accounts. For they can simply respond that [D1/2] are examples of the efficient causality associated

[26] For Hyman, the relevant counterpart would be

[**I1**] James went to church because he wanted to please his mother,

where the desire described in the *explanans* describes a disposition to do things for the sake of pleasing his mother. For Marcus, the relevant counterpart might be the slightly more convoluted

[**I1**] James went to church because he was representing pleasing his mother as to-be-done,

but in this context the basic ideas is the same.

with one kind of disposition, whereas the differences we have pointed to show that [I2] is an example of *teleological* causality that is therefore associated with *teleological* dispositions. In other words, they can acknowledge a difference between the kinds of disposition involved in each case, but still allow that they play an analogous role.

One thing we have learnt from the discussion in §4.2 is that differences like this are possible locations for grammatical confusions. This should warn us to look carefully at attempts to characterize the new kind of disposition and its role in explanation: *wanting* to acknowledge a difference might not be sufficient for successfully clarifying this difference in one's account. But there is also a further problem: the idea of a teleological explanation, and a related kind of causality, still doesn't get us what we need for an account of rational explanations. For rational explanations—and therefore, on this account, the dispositions to which they appeal—have further features that mark them off from other kinds of teleological explanation. Compare the following examples:[27]

[P1:] The flower turned to the left because the plant was tracking the sun.

[O1:] Jones' right ventricle pumped blood through his pulmonary artery because it was pumping blood to his lungs.

[I1:] James is going to church to please his mother.

Each could be related to further judgements or explanations, but a relation to something like the following is a distinctive mark of [I1]:

[I2:] James is going to church because he believes that it will please his mother.

In other words, it is a mark of the kind of teleology that is represented in reason-giving explanations, and therefore of the kind of disposition to which they appeal, that it depends on

[27] The first two are taken from Eric Marcus' account. See [24, 51].

judgements or beliefs held by the agent.[28] An adequate account must clarify this difference from other kinds of teleological disposition or ability—and comparison with the other cases of dispositional explanation, even if it takes us beyond the Humean paradigm, will not be sufficient by itself. For it turns out that one thing that marks the 'teleological dispositions' that figure in these accounts off from other kinds of disposition is that their acts are subject to a distinctive kind of reason-giving explanation:

> *S* is doing A on the grounds that p.

This threatens to leave a mystery at the centre of these accounts: they aim to explain a distinctive form of causality, and do so by appeal to a distinctive kind of disposition. But it turns out that one thing that distinguishes this disposition from other teleological dispositions is that its acts depend on beliefs that specify reasons for acting. That is to say, the account appeals to a disposition whose distinctive feature is the very thing it was supposed to explain. Thus, simply saying that the dispositions involved are like other causal dispositions, except that they differ in some essential way, threatens to leave the character of that difference and the 'causality' associated with it mysterious.

Overstating the Difference

The second problem is almost the inverse of the first. Whereas the difficulties sketched above seemed to arise from treating the causality involved in rational explanations as too similar

[28] This is why, in characterizing the grammar of acting for a reason, Anscombe pointed out that the relevant acts were done both for the sake of the end *and* in view of the thing believed. There is a tendency in some recent literature to treat deviant cases as emerging *solely* from an inability to make sense of teleology on a Humean account. But it is possible to come up with cases where an action was done for the sake of the end, and can be explained by a belief, but not because it was done in view of the thing believed. Any case in which I am deceived about my grounds for acting would provide such an example. Or if self-deception is too controversial a topic, imagine the following scenario:

> I am going downtown, and can catch a train at 4.20 or 4.35. I mistakenly believe that the earlier train leaves from Platform 4, and miss it as a result, so I catch the latter train. Explaining my lateness to a friend, I saw "I'm on the 4:35 because I thought the 4.20 left from Platform 4".

My belief explains my intentional action, but not because I take that belief to specify grounds for the action. This is partially reflected in the psychological verb in the *explanans*, and the tense in which it is predicated ('I thought. . . ').

to other kinds of causality, the difficulty I want to highlight now stems from making it seem to be so different as to be utterly divorced from other kinds of causality.

While our explanation,

[I1:] James is going to church to please his mother,

is intrinsically related to,

[I2:] James is going to church because he believes that it will please his mother,

it might also be related to such explanations as

[I3:] James is going to church because his mother told him to,

or

[I4:] James is going to church because this morning he kept thinking about the importance his mother attached to it.

Both [I3] and [I4] represent James' going to church—the manifestation of the relevant teleological disposition—as a *response* to something, and therefore as involving a 'mental cause' in Anscombe's sense. They are therefore intrinsically related to explanations such as

[I5:] James is going to church because his mother said 'Go to church!',

and

[I6:] James is going to church because this morning he thought 'Mother would want me to!'.

That is to say, the same action described in our original explanations can also be subject to 'causal explanations' in our original sense.

Nothing in Anscombe's argument commits her to denying that there are various causalities involved in action, nor that, in a particular case, we might (as Davidson suggests) be quite certain that there must be a cause of a particular sort leading up to the action. However, she is quite clear that in many cases our investigations into the causalities associated with particular actions, and our sense of what kind of causes must have led up to an action, will *depend* on (rather than ground) our sense of the end for the sake of which the agent undertook the action, and the grounds on which she did it. Thus, speaking of the causalities involved in historical accounts of events, Anscombe states:

> The first thing to note is: these causalities are mostly to be understood derivatively. The derivation is from the understanding of action as intentional, calculated, voluntary, impulsive, involuntary, reluctant, concessive, passionate, etc. The first thing we know, upon the whole, is what proceedings are parleys, agreements, quarrels, struggles, embassies, wars, pressures, pursuits of given ends, routines, institutional practices of all sorts. That is to say: in our descriptions of their histories, we apply such conceptions of what people are engaged in. In the context of such application, then, the causalities to which we ascribe such events can get a foothold. [5, 107]

Indeed, in some cases, an understanding of the relevant causality might *essentially involve* an understanding of the related reasons-based explanations. This might be true wherever the action to be explained is represented as some kind of rational *response*—for it is implicit in the idea of a response to something that it does involve some form of the causality that Anscombe tried to characterize. If, for instance, I say that you stood up to greet your friend, we might ask about the moment you noticed her entering the room, etc – that is, if I represent your action as a certain kind of response, it makes sense to look for the the causal antecedents of that response.[29]

[29] Or to give Anscombe's more extended example:

> [T]his man was travelling from Aix to Ghent. What for? He was a messenger taking news. So in the situation in which the news was generated, and in which there was a requirement that he should take it, together with the instructions of whoever sent him off, and the exigencies or difficulties posed by his means of carrying out the purpose, together with accidental encounters

The second danger in introducing the dispositions involved in rational explanations purely by way of *contrast* to other cases is that in doing so one might obscure this point. For instance, one might correctly note that, in explanations such as [I1] and [I2], the 'cause' cited in the *explanans* is not a stimulus or trigger for the action described in the *explanandum*, and the explanation as a whole does not represent that kind of causality. But in saying this, one needs to leave open the possibility that the original explanation might be (perhaps intrinsically) related to explanations that do represent the action described in the *explanandum* as a response involving this kind of causality. Indeed, it might even turn out that, in some cases at least, an understanding of the *kind* of response in question, and therefore of this specific version of Anscombe's generic notion of causality, essentially involves seeing the disposition it manifests as 'rational', and thus seeing the act itself as subject to all three kinds of explanation.

An account would fail in this regard if it seems as though the manifestations of 'rational' dispositions it claimed were represented in reason-giving explanations *could not* be understood to involve causalities of this sort on pain of undermining the whole account. For while it is true that explanations such as [I1] and [I2] are not causal in the generic sense that Anscombe characterizes, the acts they explain (which are purportedly the manifestations of these 'rational dispositions') could nonetheless be essentially involved in various kinds of causalities *qua* the kinds of acts that the are. If that is right, it had better not be the case that acknowledging these causalities undermines the overall account of acting for a reason.

and concatenations of events with aspects of temperament and facts of people's excitements – all these will contribute causalities of various kinds to the event of his arrival or non-arrival at his destination. [5, 107]

CHAPTER 5

WITTGENSTEIN ON READING: PI §156-§178

My aim in this chapter is to begin to show how Wittgenstein's brief discussion of the concept *reading* in §156-178 of the *Philosophical Investigations* can contribute to a broader treatment of our topic 'acting for a reason'. The sections in question fall in the middle of the extended discussion of rule-following, which stretches back to at least §143 (though its themes are explicitly anticipated as early as §85), and continues through §242 and beyond. Topics covered in this longer sequence encompass the normative, explanatory, and psychological dimensions of rule-following, raising questions such as how the expression of a rule can determine behaviour as correct or incorrect, how it can explain such behaviour, and what role psychological states play in mediating these relationships. As such, the discussion as a whole has clear affinities with the contemporary discussions of 'acting for a reason' surveyed in §2.

But the discussion of *reading* has a more particular interest in relation to the topics discussed in §4, i.e. Wittgenstein's discussion of our ways of drawing a distinction between reasons and causes. As we have seen, discussions of Wittgenstein's treatment of this topic in his later work tend to focus on one or two passages from *The Blue Book*, since those are the places where he explicitly talks about reasons and causes, and appears to draw a sharp distinction between them. In this chapter, I shall argue that Wittgenstein continues to treat this topic—albeit in less explicit terms—throughout *The Brown Book* and the *Philosophical Investigations*, and does so in a way that further undermines the claim that he endorsed the argument outlined in §4.1.

Part of the interest of these sections lies in the fact that they repeat and develop the arguments I sketched in the previous chapter. As we shall see, Wittgenstein's interlocutor can be understood as being motivated by particular versions of our two philosophical questions (MQ) and (EQ):

$\mathbf{MQ}_R$**:** What is it to read?

$\mathbf{EQ}_R$**:** How do we come to have first-personal knowledge of our acts of reading?

The answers that the interlocutor suggests to these questions, and the character of Wittgenstein's response, are analogous to the more compressed arguments I attributed to the discussion from *The Blue Book.* But their application in a discussion of *reading* is of particular interest, since consideration of *what it is to read* complicates the idea of a sharp distinction between reasons and causes, or between reason-giving and causal explanations. This is because *reading* is an activity that intrinsically possesses both causal and normative aspects: what one says, when one reads, is a response to what is written on the page; and it can be evaluated as an act of reading insofar as it correctly renders what is written there. The first point makes *reading* a causal concept, since (in Anscombe's terms) it involves the idea of the derivativeness of an effect from its causes: when we describe someone as reading, we represent what they say as a response to, and therefore derived from, what is written on the page. The second point shows that reading involves a particular form of normativity: for in some minimal sense, what is written on the page provides a *reason* for the reader's utterances, since those utterances can be evaluated—*qua* acts of reading—as correct or incorrect renderings of that writing. Together, these points mean that our talk about acts of reading involves both the grammar of causation and of reasons. For a description:

> James read: "These words, it seems to me..."

entails an explanation of the form

> James said "These words, it seems to me..." because that was what was written
> on the page

while also being related to such judgements as

> James read what was written correctly.

On the other hand, a description such as,

> James misread: "Those words, it seems to me..."

might be related to an explanation of the form

> James said "Those words, it seems to me..." because he thought that was what was written on the page,

and a judgement,

> James read what was written incorrectly.

Reflection on Wittgenstein's discussion of *reading* will therefore lead to a more nuanced picture of his treatment of the relation between reasons and causes.

Consideration of acts of reading will also serve a broader role in my overall argument. For if such acts are instances of acting for a reason, then any general and substantive response to (MQ) ought to encompass such acts. Thus, even though the accounts surveyed in §2.3 were pitched in general and abstract terms, they ought to apply to acts of reading. If I can show that they do not apply to such acts, then I will have shown that they do not work as a general account of 'what it is to act for a reason. The discussion in this chapter will provide the background for our evaluation of contemporary accounts in chapters 6 and 7.

This chapter falls into four main sections. In §5.1, I provide an overview of Wittgenstein's discussion, and show how it repeats and develops the arguments laid out in the previous chapter. In §5.2, I begin to relate this discussion to the contemporary accounts of 'acting for a reason' discussed in §2. §5.2 shows how the interlocutor's proposals in §156-178 echo the basic form of reductive accounts, and suggests that Wittgenstein's discussion shows why such accounts must fail. Arguments from this section will be extended to contemporary reductive accounts in the next chapter. §5.3 develops an alternative approach to the question 'what is it to read?', and §7.2.1 shows how the results of this approach can be extended to cover other related cases. These discussions will provide a model for the overall approach to (MQ) proposed by this dissertation.

5.1 §156-§178: Wittgenstein on Reading

5.1.1 The Shape of the Discussion

The discussion of *reading* comprises a short sequence of passages from §156-§178, which fall in the middle of an extended investigation of *understanding* and related notions.[1] At the beginning of the excursus, Wittgenstein provides the following brief description of what he will mean by 'reading':[2]

> I am not counting the understanding of what is read as part of 'reading' for purposes of this investigation: reading is here the activity of rendering out loud what is written or printed; and also of writing from dictation, writing out something printed, playing from a score, and so on. (§156)

The investigation into this attenuated concept is guided by two kinds of question about acts of reading:

MQ$_R$: What does reading consist in?

EQ$_R$: How do we come to have first-personal knowledge of our acts of reading?

These questions are akin to those guiding the contemporary discussion of *acting for a reason* that we surveyed in §2. In the context of the *Philosophical Investigations*, (MQ$_R$)

[1] An earlier version of the discussion appears on pp.119-125 of *The Brown Book*

[2] Given this broader context, it comes as something of a surprise when, after insisting that an excursus on *reading* will make matters clearer, Wittgenstein states that he is "not counting the understanding of what is read as part of 'reading' for the purposes of this investigation". This attenuation of the concept *reading* might seem strange given the discussion that takes place in the surrounding sections of the book: if your topic is understanding, why exclude the understanding involved in reading from your investigation? And why extend the concept to include additional cases that seem to involve important differences from your central topic? By doing this, its scope seems to be narrowed in such a way that much of what gives reading a central place in our lives is completely excluded. In place of the richly evocative concept of reading, we get an attenuated and almost mechanical notion, something that could as well describe the activity of machines as of persons. One might think of what Wittgenstein is doing here as providing us with a description of a more "primitive" or simple activity that has various affinities with our everyday acts of reading, or perhaps as drawing our attention to certain aspects of our everyday activity, while simultaneously excluding others from view.

arises out of a comparison between the acts of an experienced reader and those of a beginner. The experienced reader reads fluently and effortlessly:

> His eye passes–as we say–along the printed words, he says them out loud—or only to himself; in particular he reads certain words by taking in their printed shapes as wholes; others when his eye has taken in the first syllables; others again he reads syllable by syllable, and an occasional one perhaps letter by letter. (§156)

The beginner, on the other hand, barely reads at all:

> [H]e reads the words by laboriously spelling them out.—Some however he guesses from the context, or perhaps he already partly knows the passage by heart. Then his teacher says that he is not really reading the words (and in certain cases that he is only pretending to read them). (§156)

Wittgenstein suggests that the stumbling and hesitant 'reading' of the beginner might bring us to ask ourselves a version of (MQ): what does *reading* consist in? After all, we do mark a difference between, on the one hand, the fluent and effortless reading of the experienced reader or the stumbling successes of the beginner, and on the other, occasions on which a person is merely guessing or shamming. In asking what reading consists in, (MQ_R) asks what it is that distinguishes those utterances that are genuine acts of reading from those that are not. And despite its mundane subject-matter, it has the same form as the *metaphysical question* we considered in the previous chapter, asking 'what is it to be X?', 'what is essential to being X?', or 'what does being X consist in?'.

The interlocutor's attempts to answer (MQ_R) bring attention to a number of further features of acts of reading. For in trying to say what *reading* consists in, the interlocutor notes that acts of reading involve certain kinds of knowledge on the part of the reader. First, there is the fact that "[a] man surely knows whether he is reading or only pretending to read!" (§159). Someone who is reading knows that they are doing so, as does someone who is guessing or pretending—and they know this just insofar as they are acting in this way (§159-161). Second, there is the fact that a person who has learnt to read "derives the sound of a word from the written pattern by the rule that we have given him" (§162). Someone who is reading acts in accordance with certain rules, and can be said to know what the rule requires here and now insofar as he is acting in accordance with it (§162-4). Third, there is the fact that in reading we are responding to particular written markings, and know that we are doing as much just insofar as we are reading them. The markings might then be said to 'cause' (§169) or 'guide' (§170) the reader's utterances, and so to be known by her as causing or guiding what she says—again, just insofar as she is reading them.

These moments in the discussion give further substance to (MQ_R), insofar as they suggest that acts of reading are as such characterized by these normative, explanatory, and psychological dimensions. But they also suggest a further range of questions, which we can gather together under (EQ_R). What becomes apparent is that the act of reading involves various kinds of first-personal knowledge on the part of the reader, which include knowledge that she is reading, knowledge of which signs she is responding to, and knowledge of how those signs are to be read. This means a reader has knowledge of her acts, and of the normative and explanatory dimensions of those acts, just insofar as she is their agent. (EQ_R) therefore has the form of an *epistemological question*, which it shares with questions about other kinds of first-personal knowledge, and about self-consciousness in general: how do we come to have first-personal knowledge of X?

As with the accounts we considered in §2, a thought that guides many of the interlocutor's responses is that an account of what it is to read ought to be able to explain how it is that the reader comes to have this knowledge just in virtue of her act of reading.[3] This is to treat (MQ_R) and (EQ_R) as essentially connected to each other, so that while the discussion as a whole is focused on the question 'what reading consists in?', it is typically coupled with

[3] The interlocutor arguably gives up this goal when he resorts to appeals to 'mechanism', which purport to answer (MQ_R) without addressing (EQ_R). See below.

questions about how we come to have knowledge of such acts.

5.1.2 The Form of the Interlocutor's Account

The discussion of 'reading' exemplifies a manner of engaging with philosophical positions that is characteristic of the *Philosophical Investigations* as a whole, with the positions taken by the interlocutor, and the ways in which Wittgenstein responds to them, providing an outline of patterns that will be found more broadly in philosophical discourse. Faced with what amount to metaphysical and epistemological questions, the interlocutor proposes a particular kind of account in response to them—and when that account is shown to be inadequate, insists that there *must* nevertheless be some account of that form, even if its details are unavailable to us.

As we shall see, the interlocutor's understanding of his questions, and the urgency with which he insists on the need for a response to them, both emerge from the perspective that he adopts on acts of reading: he 'zooms-in' on a particular utterance, and ignores the broader circumstances in which the utterance is made. But without these circumstances in view, it being an act of reading is obscured, and the utterance itself seems indistinguishable from that of someone pretending to read. (MQ_R) arises in direct response to this appearance: we need to be able to say in virtue of what this utterance counts as an act of reading, whereas other apparently indistinguishable utterances do not. If we cannot provide a basis for this distinction, we will lose our grip on what it is to be an act of reading, and with it our understanding of the normative, explanatory, and psychological dimensions of such acts.

Faced with this problem, the interlocutor seeks an account that can explain both what an act of reading consists in, and how it is that we come to have first-personal knowledge of those acts, in terms of some independently-specifiable, occurent, and self-identifying feature in virtue of which the act will count and be known as an act of reading. As such, his proposals have the form of a proto-reductive account. For the idea of an independently-specifiable basis promises to explain *what it is to read* in terms of features that are independent of our concept of *reading*, and thus to provide a response to our metaphysical question (MQ_R) whose content does not presuppose the concept it hopes to explain. Furthermore, the idea that this feature is occurent and self-identifying promises to explain *how we come to have first-personal knowledge of our acts of reading.* For if that feature is occurent with the act of reading, and self-identifying *qua* basis for a particular epistemic claim, it will show how we have grounds for our specific knowledge of our acts just insofar as we are reading.

Thus, while the interlocutor's specific proposals may seem primitive or implausible, they share essential characteristics with 'sophisticated' accounts of more familiar philosophical topics like the ones surveyed in the second chapter. If anything, the 'naive' character of the interlocutor's responses help make the form of such accounts, and the motivations behind them, particularly clear, so that Wittgenstein's responses in these sections can help us see why no account with these characteristics can meet the ambitions it sets for itself.

To bring out these parallels, let's look more closely at the perspective from which the interlocutor raises his metaphysical and epistemological questions. Here and elsewhere, he tends to raise his various worries from a perspective that 'zooms in' on something we describe in terms of the concept under discussion, in a way that isolates it from the circumstances and surroundings in which it belongs, and takes the object of concern as 'given' for investigation independently of these circumstances.[4] He then finds himself asking, of whatever he has isolated in this way, in virtue of what it counts (and is known) as something describable by the concept in question—and when he struggles to find anything specifiable from this perspective that will provide what he thinks is needed to explain this, he insists that there must nevertheless be something to play the relevant role, even if it is somehow 'hidden from view'. For if there is not, it seems to him that we will lose our grip on what it is to be an

[4] The same perspective, and with it the same kind of metaphysical and epistemological worries, can be found throughout the *Philosophical Investigations.* It is perhaps most succinctly expressed in §432: "Every sign *by itself* seems dead. *What* gives it life?". From the perspective adopted by the interlocutor, the sign itself is (e.g. an arrow) is just a "dead-line on paper" (§454), indistinguishable from a meaningless marking. Given this, the question how it is that some such marks can *mean* something, and be *understood* as meaning something, when they seem in themselves 'lifeless', becomes particularly pressing.

instance of the relevant concept, and of how we come to have knowledge of such instances.

In the sections on reading, this perspective emerges naturally out of the way the problem first strikes the interlocutor. We begin from "the use of [the word 'reading'] in the ordinary circumstances of our life":

> A person, let us say an Englishman, has received at school or at home one of the kinds of education that is usual among us, and in the course of it has learned to read his native language. Later, he reads books, letters, newspapers, and other things. (§156)

This describes something of the circumstances within which we normally describe people as readers. However, the interlocutor ignores all of this to focus in on a particular case: "what takes place when, say, he reads a newspaper?" (§156). It is from this perspective—i.e. zoomed in on the particular utterances made by an expert and a beginner, perhaps as they respond to the same signs—that he is brought to ask why the utterances of the expert count as *reading*, while those of the beginner are merely guesswork or pretence.[5] For once he has isolated the utterances from their circumstances, there seems to be nothing about them that shows why one, but not the other, counts as an act of reading: *qua* utterance—and indeed, *qua* utterance made in response to particular signs—there may be nothing to distinguish them. This forces the interlocutor to try to find something else to explain our distinction: a supplement to the expert's bare utterance in virtue of which it will count (and be known) as an act of reading.

As the discussion continues, the interlocutor goes on to note further features that characterize acts of reading. It is not simply that some utterances count as such acts, while others do not, and that the reader knows which is which. There is also the fact that an act of reading involves certain kind of normative or explanatory relations between the utterance and particular written letters, and that the reader has first-personal knowledge of these relations.

[5] In §151 he 'zooms in' on the moment at which a person understands the principle of the series, and in §138 the moment at which they grasp the meaning of the word.

These points all seem to belong together with the fact that this utterance counts as an act of reading, which suggests that his answer to (MQ$_R$) had better make it clear why any such utterance involves those normative or explanatory relations, and his answer to (EQ$_R$) had better make it clear how the reader has knowledge of them.

The perspective from which the puzzle emerges then shapes the kind of ‘account’ the interlocutor thinks he needs: he must find something in virtue of which one of the utterances—now considered indistinguishable in themselves—counts as an act of reading, whereas the other does not. This shows why the interlocutor comes to feel an urgent need for a response to our ‘philosophical questions’ (MQ$_R$) and (EQ$_R$), while also dictating the form of his response. What we thinks he needs is an *independently-specifiable* supplement to the act, i.e. one that can be described independently of the broader circumstances, in virtue of which an utterance will count as an act of reading. The most immediate place to look for such a supplement is alongside the utterance itself, and this suggests that the supplement should be something *occurent*, in the sense of taking place at the same time as or immediately prior to the utterance, so that it can play the role of making this specific utterance count as an act of reading. We could then say that the particular utterance counted as an act of reading in virtue of this independently-specifiable feature, answering our metaphysical question. Moreover, if this occurent and independently-specifiable supplement were also *self-identifying*, in the sense that its character was immediately apparent to the agent just insofar as it was present, it could also provide a basis for answering our epistemological questions as well. For if the reader is aware of this supplement just insofar as she is reading, then it can provide an immediate basis for her knowledge of those acts.

To bring out these parallels, and their relation to our discussion of passages from *The Blue Book* in §4, it is helpful to look at some passages that precede the earlier version of these sections presented in *The Brown Book*. There Wittgenstein is discussing what is involved in someone having an ‘ability’ that involves being ‘guided’ by some system of signs. When we try to work out what it is to be guided in this way, we have before our minds the picture of a particular kind of mechanism:

> Here, in the working of the pianola we have a clear case of certain actions, those of the hammers of the piano, being guided by the pattern of the holes in the pianola roll. We could use the expression "The pianola is reading off the record made by the perforations in the roll", and we might call patterns of such perforations complex signs or sentences. . . You see here the connection between the idea of being guided and the idea of being able to read new combinations of signs; for we should say that the pianola can read any pattern of perforations, of a particular kind, it is not built for one particular tune or set of tunes (like a musical box). . . - (118)

This gives us a particularly clear picture of *what it is to read*—for the pianola can be said to read the roll when the mechanism functions correctly. Moreover, that functioning consists in occurent events that can be described independently of the broader circumstances in which the 'acts of reading' occur. It thus provides us with a model for answering our version of (MQ): in this case, reading consists in the occurent, independently specifiable action of the mechanism that caused the pianola to play a particular note. The pianola is 'reading' if the notes it plays are caused in this way; and it is not 'reading' if it is caused to play notes by some other process.

Building on our discussion in the previous chapter, we can expect the interlocutor to go wrong if he takes this picture of one kind of causal transaction that he would call a case of 'reading', and insists that all other such acts should be understood in the same terms. For this will lead him to look for something like the occurent action of the mechanism, specifiable independently of the broader circumstances, in virtue of which a particular utterance will count as an act of reading. But when he considers the description of the fluent reader, no such thing seems to be involved. As Wittgenstein puts it in *The Brown Book*:

> It is clear that although we might use the ideas of such mechanisms as similes for

> describing the way in which [a person guided by symbols acts], no such mechanisms are actually involved. . . . We shall have to say that the use which we made of the expression "to be guided" in our examples of the pianola . . . is only one use within a family of usages, though these examples may serve as metaphors, ways of representation, for other usages. (119)

The interlocutor in the *Philosophical Investigations* is also struck by this feature of mechanisms. Faced with our version of (MQ), he insists:

> There are at all events two different mechanisms at work here. And what goes on in them must distinguish reading from not reading. (§156)

A 'mechanism' is precisely something describable in abstraction from the use or characteristic activity of that to which it belongs.[6] For example, think about the various ways in which we might describe an artefact such as a wrist-watch. One way of doing this would make clear that it was a device that played a role in a particular practice of time-keeping, and describing the watch in these terms would involve implicit appeal to the context provided by this practice. Here we would talk about the watch *as* a device for telling the time by e.g. referring to the hour-markings, the seconds-hand, etc., all of which would involve reference to the practice to which it belongs, and thus the circumstances of its use. But there are other ways of describing the watch that abstract from this practice, and describe it purely as an artefact that functions in a particular way. Here we might talk about the functioning of the gears, the way they move the hands, etc., without any explicit reference to the idea of time-keeping. This would be to describe the watch in terms of its mechanism, and not in terms of what it was for or how it was used.[7]

[6] See [12] on the role played by appeals to mechanism and sensation in these sections of the *Philosophical Investigations*.

[7] Here I have represented the description of the watch *qua* mechanism as an abstraction *from* the description of the watch *qua* time-piece. It is part of the interlocutor's account that the former kind of description is available not simply in abstraction from the latter, but completely independently from it. There is perhaps some plausibility to this claim as it is applied to artifacts: we can imagine someone unfamiliar

If acts of reading involved such a mechanism, then we would be able to say what makes one utterance an act of reading, and another e.g. a lucky guess, by describing the workings of that mechanism. These would consist of happenings that were contemporaneous with the act of reading, and that were describable independently of the broader circumstances in which the act occurred. As such, they seem to promise exactly the kind of account the interlocutor thinks he needs.

The idea of a 'mechanism' is picked up again in subsequent sections, e.g. in §157 where Wittgenstein points out that our application of the concept *reading* to ourselves and to other living creatures does not depend on any appeal to a mechanism (as it might if we were describing an artificial machine):

> [I]n the case of the living reading-machine, "reading" meant reacting to signs in such-and-such ways. This concept was therefore quite independent of that of a mental or other mechanism.

The interlocutor responds to this by suggesting that the only reason we do not appeal to such a mechanism (as we might when describing a machine) is "because of our too slight acquaintance with what goes on in the brain and the nervous system":

> If we had a more accurate knowledge of these things we should see what connexions were established by the training, and then we should be able to say when we looked into his brain: "Now he has *read* this word, now the reading connexion has been set up". (§158)

This indicates a further move on the interlocutor's part. Though he is unable to come up with an independently-specifiable basis that distinguishes acts of reading from acts that are not reading, he nonetheless insists that such a thing *must* be there: "it presumably

with watches finding such a device and seeking to understand its mechanism, without any conception of what the device is for. This would fit the criteria for mechanistic description mentioned in §149: "a knowledge of the construction of the apparatus, quite apart from what it does". However, as Wittgenstein indicates in that section, the same cannot be said of acts of reading and other psychological concepts.

must be like that—for otherwise how could we be so sure that there was such a connexion?" (§158). And if we are unable to describe such a basis now, that *must* be a result of a gap in our empirical knowledge, one that we know *must* (in principle at least) be fillable. The interlocutor's insistence at this point stems from the fact that, from the perspective he has adopted, it looks as though unless we find some such basis there will be *no* principled difference between the utterances of the beginner and those of the expert, meaning that we would lose our grip on the very idea of an *act of reading*.

The appeal to mechanism thus echoes the kinds of confusion I described in §4.2: the interlocutor takes a particular example of 'reading', and insists that all other cases be understood in the same terms. However, we now have a fuller picture of how a particular perspective on the acts he was concerned with led to this confusion: because he had focused in on each utterance, in isolation from its circumstances, he felt that his answer to (MQ$_R$) must be provided in the same terms.

We can see further parallels in the interlocutor's proposed responses to (EQ$_R$). As in *The Blue Book*, he is impressed by the fact that the reader can just say that he is reading, and in doing so point to that to which he is responding. This lends further support to the idea that there must be *something* that his act of reading consists in: for how else could he just know these things? However, that something cannot be the utterance itself, because that is indistinguishable from the utterance of the non-reader. There must be something else that grounds his knowledge.[8] This explains the naive mentalism evidenced by the way

[8] The situation is analogous to other cases that Wittgenstein describes in *The Brown Book* and returns to in the *Philosophical Investigations*:

> How does pointing to its colour differ from pointing to its shape?—We are inclined to say the difference is that we mean something different in the two cases. And 'meaning' here is to be some sort of process taking place while we point. What particularly tempts us to this view is that a man on being asked whether he pointed to the colour or the shape is, at least in most cases, able to answer this and to be certain that his answer is correct. If on the other hand, we look for two such characteristic mental acts as meaning the colour and meaning the shape, etc., we aren't able to find any, or at least none which must always accompany pointing to colour, pointing to shape, respectively. (80)

I.e. when a person is just able to answer, and to be certain that his answer is correct, we look for something to provide a basis for his knowledge: some process of meaning one rather than the other, contemporaneous with the act itself, in virtue of which it can be said to count (and be known) as an act of pointing to the he repeatedly comes back to the idea that *what it is to read* should be spelt out in terms of a *felt or conscious experience*: he suggests that reading is a 'special conscious activity of the mind' (§156), that it involves 'characteristic sensations' (§159), that "[t]he words that I utter *come* in a special way" that is characteristic of reading (§165), that we *feel* some kind of influence or causal connexion to the signs we are reading (§169), or that we "experience the because" (§177).

Felt or conscious experience strikes the interlocutor as a promising basis for a response because he takes it to meet all three of his criteria: he conceives of such sensations as something that could be both occurent with the act, describable in isolation from its circumstances, and immediately known by the agent *as* a particular kind of sensation.[9] If that were so, the occurrence of 'characteristic sensations' of reading (e.g. "sensations of hesitating, of looking closer, of misreading, of words following one another more or less smoothly" (§159)) could provide the required supplement to the utterance, in virtue of which it would be an act of reading (rather than, say, an act of pretending, where the speaker "will have none of the sensations that are characteristic of reading, and will perhaps have a set of sensations characteristic of cheating" (§159)).

Moreover, since the reader will be immediately aware of these sensations, just insofar as she is reading, they could also serve to explain how it is that she knows that she is reading rather than pretending. For if reading consists in the presence of such sensations, and the reader is immediately aware of these sensations as of a particular sort, then she will be aware that she is reading just insofar as she is reading. Indeed, this last point helps explain the more puzzling appeals to sensations that the interlocutor makes later in the discussion. For example, when seeking to explain the reader's knowledge that her utterance is a response to particular signs, the interlocutor says that he wants to say that he can

colour rather than the shape.

[9] This conception of sensations as independent from, and prior to, the language we use to express, describe, and refer to them, and to talk about our psychological life more generally, comes under sustained attack in §243 ff. Here Wittgenstein lays bare the picture of sensations as providing an independently given basis for our application of psychological concepts, and shows its incoherence.

"feel that ... the utterance was *connected* with seeing the signs" (§169), or that he had "experienced the '*because*"' (§176) Here too, the interlocutor is seeking to explain our first-personal knowledge of the explanatory or normative dimensions of our acts of reading by reference to a self-identifying sensation.

The interlocutor's naive mentalism thus promises to provide the basis for a proto-reductive response to both (MQ_R) and (EQ_R), since the felt sensations he appeals to would serve as a supplement to the bare utterance in virtue of which it would count, and be known as, an act of reading. Indeed, they make the basic features that the interlocutor requires from a supplement—that it be independently-specifiable, occurent, and self-identifying—particularly vivid.

To claim that there must be a mechanism that distinguishes acts of reading from other acts is to claim that there must be an independently specifiable basis in virtue of which they count as acts of reading, and to insist that we need such an account if we are to respond to (MQ_R), and thus keep our grip on what it is to be an act of reading.[10] Or again, to claim that there are 'characteristic sensations' that distinguish acts of reading from other acts is to claim that the independently specifiable basis in virtue of which they count as acts of reading also explains how we know they are such acts. The appeal to 'mechanism' or 'sensation', conceived as something describable in terms that are independent of the broader context in which it occurs, is thus a particular instance of the interlocutor's general approach. For our purposes, it is important that we focus primarily on the *form* of the interlocutor's account as much as his specific proposals. What matters is not the appeal to sensations, mechanisms, or neurological processes *per se*, but the fact that each proposed account involves the idea that we must be able to provide an occurent and independently-

[10] Since a mechanism is describable independently of broader circumstances, and its workings are occurent along with the acts it produces, the appeal to mechanism promises to meet at least two of the interlocutor's criteria for an ideal basis for his account of what it is to read. However the appeal to mechanism also potentially involves giving up on the idea that an account that provides an adequate response to (MQ_R) will also she light on (EQ_R), since there is no reason immediate reason to suppose that the reader is immediately aware of the workings of the mechanism, nor that those workings are self-identifying in the way that sensations seemed to be. At the very least, it would take a special kind of mechanism to meet these further criteria.

specifiable basis to explain why an utterance counts as an act of reading, and a self-identifying basis to explain how the reader comes to have certain kinds of first-personal knowledge of that act. If we view the interlocutor's proposals at this level of generality, we will be in a better position to understand what Wittgenstein is doing here, and to bring out the parallels with contemporary accounts—for, as we shall see, *any* account that proposes to explain 'what it is to be X' or 'how we come to have first-personal knowledge of X' in similar terms will have the same form, whatever its specific details. The interlocutor's proposals give us a vivid version of such an account, with the appeals to sensation or mechanism making the basic form particularly clear. But even accounts that have a more general focus than acts of reading, and draw on more sophisticated philosophical resources than the interlocutor's naive mentalism, will share this basic form so long as they insist that we need an account of this kind of basis to respond to our metaphysical and epistemological questions about our acts.

5.1.3 Wittgenstein's Response

Wittgenstein responds to such accounts by showing that they must fail on their own terms: they cannot identify an independently-specifiable feature that occurs in all and only cases of reading, since any feature that fits the interlocutor's criteria could also be present in a case that was not an instance of reading. The point is clearest in relation to the interlocutor's proposed appeal to sensations or 'special conscious mental activity'. Wittgenstein's initial response is that:

> [W]e have to admit that—as far as concerns uttering any *one* of the printed words—the same thing may take place in the consciousness of the pupil who is 'pretending' to read, as in that of the practised reader who is 'reading' it. (§156)

Since sensations, as the interlocutor envisages them, are specifiable independently of the broader circumstances, there is no reason we cannot describe a case in which the sensation in virtue of which something is supposed to count (or be known) as an act of reading is present in the consciousness of someone who is not reading.

This point is developed in subsequent sections. §159 begins with the proposal that "the one real criterion for anybody's *reading* is the conscious act of reading, the act of reading the sounds off from the letters". This is explicitly connected to the point about first-personal authority mentioned above: a person who is reading knows that they are reading, whereas a person who is merely pretending to read knows that they are pretending. The interlocutor proposes that the difference between the cases—and the basis for each person's knowledge—lies in the 'characteristic sensations' that accompany either genuine reading, or a sham like reciting a passage one has memorised in a language one cannot read. *Reading* (or pretending to read) consists in the presence of 'characteristic sensations' of the appropriate sort.

But after making this contrast, the interlocutor is brought to admit that we could describe cases in which a person, say, read a text while feeling all the characteristic sensations of reciting it from memory ("perhaps as the result of some kind of drug")—that is, the very sensations that were supposed to distinguish recitation from reading. The broader implication is that we can always imagine any such "characteristic sensation" to be felt in the wrong kind of case. Even if we did manage to identify a sensation that was present in all cases of reading, there is no reason to believe that—*qua* sensation—that it could not occur on occasions when we would not count the person as reading.

This same pattern of argument is applied to the interlocutor's attempts to characterize the normative and explanatory relations that belong to acts of reading. He begins from the thought that what makes this utterance an act of reading is that it stands in various such relations to particular letters or signs: if he can characterize what is distinctive about that connection, what makes it different from "any mere simultaneity of phenomena" (§176), he will be able to say that it is in virtue of *that* that the particular utterance counts as an act of reading. Thus in §162 he tries to give expression to the distinctive normative relations between sign and utterance, suggesting that the reading involves "deriving" the reproductions from the original; or again, in §165, that *reading* is a "quite particular process", and that when one reads the words come to one in a "special way". Later, he goes on to try to characterize the distinctive causal or explanatory aspects of reading, pointing out that the reader's utterances were 'influenced' or 'guided' by the letters in a particular way, and that this suggests a distinctive kind of causal connection is in play (§169-170).

The interlocutor hopes that by pointing to these relations, he will be able to provide a supplement to the bare utterance that will show why it counts as an act of reading. But here too, the same problem comes up: one could be deriving what one said from written words, or one could have the words come to one in a special way, and yet we would not count it as reading:

> Here I should like to say: "The words that I utter come in a special way." That is, they do not come as they would if I were for example making them up.—They come of themselves.—But even that is not enough; for the sounds of words may occur to me while I am looking at printed words, but that does not mean that I have read them.—In addition I might say here, neither do the spoken words occur to me as if, say, something reminded me of them. I should for example not wish to say: the printed word "nothing" always reminds me of the sound "nothing"—but the spoken words as it were slip in as one reads. §165

The difficulty is that the interlocutor needs to spell out a particular sense for *deriving* or *coming to one in a special way* that will apply to all and only cases of reading, and not include other phenomena such as e.g. reciting the words from memory while looking at the paper, or saying them spontaneously in response to symbols one cannot read (cf. §165-8). And it is not clear how this can be done by reference to an independently specifiable, occurent, and self-identifying feature present in all and only cases of reading. Anscombe provides a succinct summary of the difficulty in her essay *Wittgenstein: Whose Philosopher?*[10]:

> A 'special experience' or 'words coming in a special way' do not function as

> explanations of what reading is. A word might come to you in the 'special way', and any special way you care to describe otherwise than as 'the way the sounds come to you when you are reading the words' might be found in cases which are not cases of reading. As for that description, it is useless: one wants to know 'what way is that?' [10, 211]

This general dynamic suggests that the problem lies not in the details of the specific proposals, but in the form of the proposed account. The features that the interlocutor attributed to sensations (mechanisms, neurological processes, etc.)—the fact that they are describable independently of the circumstances of the act—also mean that we will always be able to imagine cases in which that feature is present in a case where circumstances were such that we could not say there was an act of reading. This means that, if utterances count as acts of reading *in virtue of* something that can be described independently of the concept 'reading', we will always be able to imagine 'deviant' cases, in which the independently-describable feature that is supposed to be essential to reading is present, but the speaker does not count as reading. This point applies as much to the interlocutor's attempts to characterize the normative or explanatory relations that belong to acts of reading as it does to his attempts to specify a psychological state or a neurological process.

The implication is that this same problem will face any account of this form: it will always be the case that the feature 'in virtue of which' something counts as such-and-such can be present in the wrong kind of case. As we saw in §2, this is precisely the form of the 'problem of causal deviance' faced by Davidson's causal-psychological account, and the other reductive accounts inspired by this approach, and we shall see in §6 that the way these accounts attempt to respond to this problem also echoes the interlocutor's responses. But what Wittgenstein aims to show us is that 'causal deviance' is not simply a problem in need of a philosophical solution, but rather an artifact of a particular perspective on our problem, and a form of account that emerges from that perspective. Any account that takes the acts it is concerned with as 'given' independently of their circumstances will ultimately begin from a conception of those acts that renders them indistinguishable from cases of the wrong sort. From this perspective, it will seem particularly urgent that we find some basis for distinguishing the cases, i.e. something in virtue of which certain acts count (and are known) as instances of the relevant concept.[11] But once the problem is seen in these terms—with the act itself 'given' in a way that renders it indistinguishable from a deviant case—there will be nothing that the interlocutor can appeal to that will provide him with a basis for the distinction he so urgently needs to make. For any independently-specifiable supplement he can provide, we can imagine a deviant case involving that supplement, and can do so precisely because the supplement purports to be independently-specifiable.[12]

Thus, if Wittgenstein succeeds in his aims here, we will come to see that *no* account could provide what the interlocutor desires *and* avoid the problem posed by deviant cases. There is no such thing as doing both at once. Moreover, the 'primitive' character of the interlocutor's account makes these contradictory features particularly vivid, since it is obvious that any sensation that accompanied an act of reading could be present in a deviant case. Beginning from this simple account helps teach us "to pass from a piece of disguised nonsense to something that is patent nonsense" (§464), so that we can see how the same problems will arise for *any* account with the same basic form, whatever its details.[13]

[11] It might also seem as though a non-reductive or primitivist account is not merely circular, but empty, since neither really purports to provide a basis for this distinction. Although the rhetoric used to present non-reductive accounts sometimes suggests that they *do* aim to provide such a basis, but further claim that it cannot be described in reductive terms, I shall suggest in §7 that it is a mistake to understand them in this way.

[12] A further move in this dialectic—which I do not consider here—is to insist that the supplement is such as to automatically rule out deviant cases. The problem with this proposal is that there is no way of making out what such a supplement could be that does not make it mysterious. See [20] on Platonism and rule-following.

[13] Wittgenstein's responses to (EQ_R) build on these arguments by showing that the interlocutor's attempt to specify a self-identifying ground for our knowledge doesn't fit our knowledge of our acts of reading. This is implicit in the discussion of mechanism, where Wittgenstein points out that the interlocutor's insistence on an unknown mechanism stands in tension with his epistemological claims. In §156, he suggests that "these mechanisms are only hypotheses, models designed to explain, to sum up, what you observe", and in §158, he advises the interlocutor "ask yourself: what do you know about these things?". Our knowledge of our own acts of reading, and those of other people, in no way depends on the ascertaining the presence of a particular mechanism, and the grammar that would fit talk about such a mechanism is out of place when it comes to a human being's ability to read. (On this point, see below.)

5.2 Reductive Accounts: Reading and Reading Explanation

This general overview of §156-§171 was intended to bring out some key moments in the discussion. It provides us with a model of a particular kind of philosophical account, and a way of responding to it, that will prove useful in the approaching the various accounts of 'rational explanation' surveyed in the second chapter.

It should already be clear that the 'primitive' moves of Wittgenstein's interlocutor, along with Wittgenstein's responses to them, are echoed in the contemporary debates on this topic. First, the perspective adopted by reductive accounts mirrors that of the interlocutor: they use the schemata for reason-giving explanation to 'zoom-in' on the relation between a particular action and a reason, and ask 'in virtue of what' does this count as a case of acting for a reason? Second, the form of reductive accounts of reason-giving explanation parallels that of Wittgenstein's interlocutor—and proponents of such accounts are often quite explicit about this fact. For instance, Kieran Setiya is careful to acknowledge the reductive form his account takes. Describing the shape of his approach, he states:

> [W]e can 'build' acting for reasons from materials present in deviant causality, along with others that are not themselves composed or defined in terms of such

The incongruity of the claim that we have knowledge of our acts of reading in virtue of some occurent feature of those acts is made particularly clear in the discussion of causation. The interlocutor attempts to ground his knowledge that his acts of reading are a response to particular signs by claiming that he *feels* a causal connection to the signs when he reads them (§169). Here too Wittgenstein points out that our claims to knowledge don't have the right grammar for the kind of causal connection the interlocutor has in mind:

> Causation is surely something established by experiments, by observing a regular concomitance of events for example. So how could I say that I *felt* something which is established by experiment? (§159)

This point is taken up again in the subsequent discussion of 'guidance':

> When I look back on the experience I have the feeling that what is essential about it is an 'experience of being influenced', of a connexion—as opposed to any mere simultaneity of phenomena: but at the same time I should not be willing to call any experienced phenomenon the "experience of being influenced". (This contains the germ of the idea that the will is not a phenomenon.) I should like to say that I had experienced the 'because', and yet I do not want to call any phenomenon the "experience of the because". (§176)

On Wittgenstein's response to the idea that we 'experience the because', see below.

> action. [33, 135]

This directly parallels the shape of the account I attributed to Wittgenstein's interlocutor in the sections on reading.[14] It is an account of what it is to be X in terms of ideas that apply equally well to things that are ¬X, and are as such independent of, and prior to, any account of X.

It might also seem that Wittgenstein's own responses suggest parallels with either primitivist or non-reductivist accounts. On the one hand, the general discussion of 'reading' helps us get in view the idea of a normatively-informed ability and the place it has in the explanation of certain acts, suggesting a clear parallel with the non-reductivist accounts outlined in the second chapter. On the other, Wittgenstein doesn't represent himself as providing any kind of *account* of 'what it is to read'—so perhaps his sympathies would actually lie with the primitivist's rejection of the need for a philosophical account of such concepts. Either way, consideration of his responses to the interlocutor promises to shed light on how either kind of account might respond to the push towards reductivism.

In §5.2, I will explore and develop these parallels, in a way that will lay the foundations for subsequent discussions of our broader topic of reason-giving explanation. To this end, I introduce an artificial class of explanations that I'll call 'reading explanations' that will these parallels explicit. Going forward, this will help us look at the ways in which the accounts of this form of explanation suggested by Wittgenstein's discussion echo the accounts of reason-giving explanation surveyed in the second chapter.

[14] Elsewhere in his work, again dealing with deviant causation, he makes an analagous move, claiming that to solve the problem we need a notion of 'sustained causation', but that this can be taken for granted in dealing with his specific topic:

> Sustained causation of a process towards its a goal is not unique to intentional action: it is present in purposive behaviour that is not intentional. So although it is something of which we lack an adequate theory, there is no circularity in taking it for granted here. [32, 32]

This is why I said earlier that Wittgenstein's interlocutor is particularly clear-sighted about the form taken by his account. Setiya ultimately needs an account of 'sustained causation', and so, once he sets aside this topic, cannot take that notion as given; the interlocutor, in his appeal to sensation, explicitly draws on something that he takes to be, in a radical sense, given.

5.2.1 Reading Explanations

The analogy depends on seeing acts of reading as simplistic case of acting for a reason: when a person reads what is written, their utterances can be explained as responses to those written markings, and can further be evaluated in terms of, or justified by reference to, those same signs. This is reflected in the fact that we might ask a person who was reading, 'Why did you say . . . ?', and she might point to what was written by way of response. This provides the basis for our introduction of the notion of a 'reading explanation':

> S said ". . . " because$_R$ "_____" is written on the page.

The subscript appended to the 'because' marks the fact that the explanations we are concerned with are those where what is written on the page explains the utterance *qua* act of reading. It is therefore supposed to differentiate 'reading-explanations' from those that explain other kinds of response to written signs.

As with the more general case, one might approach the metaphysical and epistemological questions that the interlocutor asks about reading by way of an investigation into 'reading explanations'. Here too the thought would be that if we can get clear on the explanatory character of these explanations, we would understand what reading consisted in, and why acts of reading involved certain kinds of knowledge on the part of the reader.

Reading Explanations and Rational Explanations

Reading explanation, and the acts it explains, have a number of interesting parallels with the broader class of reason-giving explanations. We can bring these out by considering the activity of a *reader*. Such a person makes the particular noises she does in response to the signs written on the page in front of her. She occasionally she misreads a sign, and when she notices this she corrects her utterance; other times she speaks too quickly or muddles her words; but, overall, she accurately renders most signs she is presented with, and competently corrects her mistakes.

As I noted in the introduction to this chapter, the relation between the sign and what the reader says can be plausibly described as a **causal** relation. In speaking as she does, the reader is reacting to the words she sees before her, such that it does not seem particularly forced to say that what caused her to say "..." was the fact that "_____" was written on the page. I further noted that *reading* also involves **normativity**, at least in the following minimal sense: what is written on the page offers a standard of correctness by which we can judge the person's performance, if she is indeed *reading* the signs before her. If a person is *reading*, she acts from an understanding of this normativity, and will recognize that the correctness of her utterances depends on what is written on the page.[15] If she says something other than what she takes to be written, she is no longer reading; and if she realizes that her rendering of what is written was incorrect, she must modify her performance accordingly.

Thus, while it might seem a little strained to say that what was written on the page *justified* or *provided grounds for* what the reader says, it is true that it both explains her utterances, and provides a standard or norm to which she holds herself accountable. In this sense, we might say that it gives her *reason* for what she says, both in the sense that it motivated or explained her utterance (i.e. she made it in response to this sign), and also in the sense that it provided a normative standard by which her act can be judged.

This gives us a very limited sense of "reason" (and with it, equally limited sense of "grounds" or "justification") that will be particular to cases of reading. We can mark this limited sense by appending our subscript to the word "reason$_R$". Broadly speaking, a particular sign is a reason$_R$ for a particular utterance just in case a competent reader of the relevant language would render the sign by an utterance of that type. In this sense, reading explanations can also reflects justificatory or normative dimensions of our talk about *reading.*

In paradigm cases of *reading*, it is essential that the *reader* take the sign she references in

[15] So, if the reader is making an honest effort of it, an explanation for any mistake she makes in reading will be e.g. that she thought that such-and-such was what was written. For instance, if she reads "CUT" where the written words have "CAT", the explanation must be that she thought what was written said "CUT". If she was otherwise competent, and admitted that what was written was "CAT", but nonetheless insisted that she was reading *that* when she said "CUT", we would be perplexed.

her explanation to be a reason$_R$ for what she said. This is reflected in the fact that we expect *readers* to be able to recognize and correct any mistakes they make in their performance. A person who, instead of reading, made the first sound who came to mind when confronted with the sign—and who happened in almost every case to make the appropriate sound—would have no notion that she was correcting her previous act if she were asked to reconsider her response. For suppose she did make the wrong noise, and was asked to try again. She might look at the sign more closely, clear her thoughts before coming out with any particular sound, etc, and perhaps come out with the correct response after doing all of this. But she would have no grounds for thinking that one was repsonse was better than the other, beyond the teacher's acceptance of her second utterance in place of her first.

The possibility of a *reader* making a mistake in her performance brings up a further interesting feature of reading explanations. We saw in §2 that some authors motivated their causal-psychological accounts of *acting for a reason* by focusing on defective cases, i.e. cases where an agent *took* such-and-such as their reason when in fact it was no reason at all. This raises the question of how we are to understand cases of *mis*reading, in relation to our explanatory schema.

In thinking about this, it will be important to differentiate what I will call DEVIANT and DEFECTIVE cases. A case is DEVIANT if the utterance is caused by the sign, but was not an act of reading. Such cases are significant because, as described, we can easily imagine an observer mistaking them for acts of reading. But despite appearances, in a deviant case the speaker does not take herself to be *reading*, and does not so much as intend to be responding to the written signs in that way. Using the jargon we introduced before, we might say that what differentiates paradigm cases of *reading* from deviant cases is that in the paradigm case, the written symbols are recognized by the speaker as a reason$_R$ for what she says, and the utterance is a particular kind of response to this recognition. In deviant cases, things are otherwise: either the speaker doesn't recognize the symbols as reasons$_R$ at all (as when a person merely appears to be reading), or the utterance is produced in some other way.[16]

A DEFECTIVE case, on the other hand, is one in which a competent *reader* makes some kind of error in her performance, i.e. cases in which a person *mis*reads a sign. We can divide such cases into two general groups: first, there will be those in which the person mis-perceives the written sign. As with paradigm cases, in a defective case the reader will take herself to have a reason$_R$ for her utterance; but in fact she does not have any such reason$_R$. Second, there are cases in which the reader bungles her performance and comes out with the wrong thing: here the reason$_R$ she recognizes does not justify what she says, as she herself would acknowledge.

Defective cases are different from deviant cases. We might characterize this difference by saying that the former are still, in an important sense, cases of *reading*, whereas the latter are not. Thus deviant cases are not subject to reading explanation, whereas defective cases are. Recognizing this might seem to pose a problem for us. For if defective cases are still cases of *reading*, and as such subject to reading explanation, then there will be instances of our schema in which the reason specified by way of explaining an utterance is not in fact a reason$_R$ for that utterance. To see this, consider a straightforward case of misreading: someone renders the sign "CUT" as "kæt" ("CAT") instead of "kət" ("CUT"). In such a case, we get the following instance of our schema:

> *S* said "kæt" because of the sign "CUT".

Here the reason given for the person's utterance is not a reason$_R$ for that utterance; it is a reason$_R$ to say "CUT", not "CAT".[17]

[16] Note that this does not mean that the written symbols can't be called "reasons" in some other sense. Take the case in which a person glances over symbols of an unknown language, and says whatever words come to mind (Wittgenstein asks his reader to do this in §169). Here the written symbols are also part of the explanation of what the person says, and in this sense might well be called "reasons for what she said"; but since she cannot read the language, she does not recognize them as reasons$_R$; and if she says what she says because of what is written, it is not because she read it. A person who did a convincing enough job of this task might appear to be reading from some unknown language; and a particularly skilled *reader* might, on being presented with a text in an unknown language, be able to make up some principles and read in accordance with them as she went. Wittgenstein describes some such case in §160

[17] Remember , this is compatible with the claim that the sign "CUT" was the reason our reader said

Since this is a defective case, we would nevertheless expect an explanation such as the following to be true:

> S said "kæt" because she thought sign was "CAT".[18]

Here we do have mention of something that *would* be a reason$_R$ for the utterance, and is understood as such by the reader. But now it seems that what is doing the explanatory work is not the actual sign, but some aspect of the reader's psychology, i.e. a particular belief about what was written on the page before her.

This might further suggest the following line of thought: we've admitted that both paradigm and defective cases are subject to reading explanation; but the only thing that is common to both kinds of case is a *belief* about a reason, not the reason itself. Given this, someone might claim that the schema for reading explanation that we've relied on so far is actually misleading; it should have read:

> S said "..." because$_R$ she thought the sign said "_____".

This line of reasoning is the exact parallel of that used to motivate certain kinds of causal-psychological accounts of reason-giving explanation more generally.

We can now make the parallels with accounts of reason-giving explanation explicit by reformulating Davidson's Question:

> What is the relation between a written word and a spoken one when the written word explains the spoken one by giving the agent's reason$_R$ for saying what she said?

This could be understood as a question about the difference between our various explanatory schemata. We adopted the following as our schema for reading explanation:

"CAT". The claim here is just that this cannot be the "reading" sense of reason: reason$_R$.

[18] Of course, this is not the *only* explanation we would accept here; but we need to be able to make some sense of the mistake, if only by recognizing that "CAT" looks a lot like "CUT", and a person (in a rush, who was tired, etc.) could easily confuse them.

S said "…" because$_R$ of the sign "_____".

Our question might be understood as asking how reading explanations differ from the wider group of explanations picked out by a schema like the following:

S said "…" because of the sign "_____"

After all, the subscript we added to our first schema is somewhat enigmatic. What exactly is it supposed to indicate? I said above that it marked the fact that we were only interested in explanations where the sign explained utterance in the way it does in an act of reading, but an obvious response would be an echo of Anscombe: "And what way is that?"

5.2.2 Reductive Accounts of Reading Explanation

With these preliminaries out of the way, we can formulate a reductive account of *what it is to read*. Its basic form will be as follows: in order to provide an explanation of what *reading* is or consists in, the account will appeal to some notion that is supposedly available independently of the concept *reading*, and then claim that an account of what *reading* consists in can be provided using this notion.

Following our earlier discussion of *acting for a reason*, the central kind of account we will consider will be a *causal-psychological account*. The notion that these accounts rely on will ultimately be one of *causation*, again supposedly graspable independently of and prior to any account of what it is to read. Taking this notion of causation as given, any such account will then aim to identify the right kind of causes for something to count as an instance of *reading*.

The possibility of misreading gives further shape to such accounts: for in looking for candidate causes in virtue of which something might count as a case of reading, they will have to either find causal antecedents that are present in both paradigm and defective cases, or deny that the latter are instances of reading. Supposing that they take the former route, the most plausible candidate for such a cause will be some psychological state or set of states present in both paradigm and defective cases. All and only utterances caused by the relevant state or states will count as reading, so the difference between instances of reading and other cases is to be understood in terms of the causes involved: the reader's utterances are caused by a particular kind of state; the non-reader's utterances are caused by different kinds of state. In both cases, we can apply an explanation of the following form:

S uttered "…" because"_____" was written on the page.

but with further specification we will be able to show that the causes of reading are significantly different from the causes of other utterances.

As I noted above, the perspective adopted by reductive accounts mirrors that of Wittgenstein's interlocutor insofar as they use the schema for 'reading-explanations' to isolate an act of reading from its circumstances. These accounts focus in on the utterance described in the *explanandum* of these explanations, and ask 'in virtue of what does this count as an act of reading?'. After all, we can easily imagine a counter-part utterance, also made in response to written signs, that did not count as an act of reading.

Once the problem has been posed in these terms, the reductive account goes on to suggest that the resources we need to make this distinction lie in the fact that our utterance counts as an act of reading in virtue of the way it is explained by the written sign. The thought is that we have what is essential to an act of reading fully in view just insofar as we begin from this schemata, and that we can therefore 'build up' to the idea that the utterance is an act of reading by providing an account of the kind of explanatory nexus the schemata represents.

Difficulties for Reductive Accounts

The problem faced by reductive accounts is that no cause they specify is sufficient to demarcate all and only cases of reading. Since the schema

S uttered "…" because "_____" was written on the page.

is common to both the beginner and the reader, a plausible initial move is to try to identify some causal antecedents that will be appealed to only in reading explanation. For instance, a psychologistic account might try to identify some particular mental state that would cause the reader's utterance but not the beginner's. But what would be a plausible candidate for such a state? Obviously the following must be true:

> *S* uttered "..." because she saw the sign "_____" written on the page.

This narrows down the range of explanations from those picked out by our original schema, since that included cases where the written sign was explanatory of what a person said, but not because she saw it. Nevertheless, it clearly won't do as an account of *reading explanation*, since both the beginner and the reader see the sign on the page.

What about adding an additional state to help differentiate the two cases?

> *S* uttered "..." because she saw the sign "_____" written on the page, and believed it was to be rendered as "...".

This seems more promising, and suggests a psychologistic account of what it is to read:

> To read is for one's utterance "..." to be caused by the perception of a sign "_____", and the belief that this sign is to be rendered as "...".

An analysis along these lines attempts to capture the reader's competence in terms of a particular belief: the belief that the written sign is to be rendered in a particular way. But once again, a problem of the same form arises: we can imagine a scenario in which this belief explains the utterances of a non-reader.[19] In fact, Wittgenstein describes such a case in §159:

> Suppose *A* wants to make *B* believe he can read Cyrillic script. He learns a Russian sentence by heart and says it while looking at the printed words as if he were reading them.

[19] A further problem lurks in the cheat-word 'rendered'. What can this mean besides 'read'? On the apparent circularity of accounts of this form, see below.

Person A certainly believes that the signs he pretends to *read* should be rendered in a particular way: that is the grounds for his pretence. Nevertheless, even though his action can be explained by his perception of the signs and his belief about how they ought to be rendered, he is not *reading*.

Another plausible line of thought would be to claim that there is something fundamentally different about the relevant psychological states of the beginner and the reader. Our interlocutor could develop this intuition by trying to provide a specification of a psychological state that could only be present in the reader, and not in the beginner. If such a state could be identified, the interlocutor could maintain that reading consists in utterances caused by that state. He might then claim that *reading explanations* could all be perspicuously rendered in the following schema:

> S uttered "…" because she saw the sign "_____" and was in state σ

and further provide an account of what it is to read in the following form:

> To read is for one's utterance "…" to be caused by a perception of the sign "_____" and state σ.

There might be various candidates for how to specify such a state—-I leave surveying them to the reader's imagination.[20] Whatever the specifics of the state σ, we can simply

[20] It is worth flagging two different shapes the account might take at this point. One version would be strictly reductive, in that is tried to specify state σ without any appeal to the concept *reading*, or at least claimed that some such specification was in principle possible. But another version might involve giving up the strictly reductive ambitions, and allowing that the concept *reading could* be used in specifying the *content* of the relevant state, but was not involved in specifying the relation between that state and the utterance in question. A hybrid reductivist might adopt something like the following schema:

> S uttered "…" because she saw the sign "______" written on the page and believed that in saying "…" she would be reading "______".

This would presumably not seem a happy suggestion to the strict reductivist, since it would yield the following—clearly circular—analysis of reading:

> To read is for one's utterance "…" to be caused by the perception of a sign "______" and the belief that in saying "…" one would be **reading** "______"

grant the reductivist his claim that such a state can in principle be specified without appeal to the concept *reading*, and ask again whether this yield an adequate account?

Based on the application of §156 of the *Philosophical Investigations* I suggested above, we would expect a response along the following lines: the same thing might take place, and yet it not be a case of reading. Examples of such cases are of necessity a little convoluted, but imagine the following scenario:

> A teacher is reading *Adventures of Huckleberry Finn* to her class, and has resolved in advance not to read the racial slurs that are part of the written text, but to substitute some other word in their place; upon seeing such a word in the text, she realizes that reading it would require that she utter some such slur; but she becomes so anxious in the face of this belief that she muddles her words and when it comes time to make a substitution, utters that very slur.[21]

The utterance of this person was caused by a perception of the sign "_____", and the belief that in saying "…" she would be reading this sign (or whatever the equivalent state is supposed to be). And yet one might claim that the connection between the sign and the utterance is not such that we would count this as reading. If that is correct, then it seems as though any state σ could not suffice—for that same state could be present, and cause an utterance—and yet the utterance not be an instance of reading![22]

[21] Nic Koziolek helped formulate this example.

[22] The case is complicated slightly by the fact that we are dealing with our everyday concept of reading, rather than Wittgenstein's more attenuated concept. This makes it natural to say: the person *read* what was written, and then decided not to say it outloud. Or even: the person *did* read it—by accident. I think it is nevertheless true that she did not *read* it in Wittgenstein's sense. But it might help to imagine another case, based on the scenario we described in §157. Imagine a *reading*-teacher who trains her best pupils in the following game: they are to render the signs as they have been taught, but whenever they come to an instance of one particular sign, they must not read it, but should instead produce some other sound of their choosing. Now, imagine someone playing this game: the special sign is "CAT", and they've decided in advance to render it some other way; but when they come to an instance of the sign, they are flustered by the thought that if they said "CAT" they would be reading the sign, and when they try to produce some other sound they end up saying "CAT"! We might say that this player did not read "CAT", even though that is what is written, and that is the sound he produced.

We can even imagine a simpler version of this case: for certain signs, the players are to chose a word to say rather than reading it. Some of them, some of the time, happen to chose the same word as was written

Faced with these counter-examples, the interlocutor might try a different tack: the problem is that the state didn't cause the act *in the right way*, and this is what is required for an utterance to count as an act of reading. So the task facing reductive accounts is not just that of identifying a particular *cause* that is active in all an only cases of reading, but further specifying the role this cause must play. But we now stand in need of a specification of what *the right way* might be. And of course, the same difficulty looms: any specification of *the right way* needs to ensure that causal connections of that sort are found only in instances of *reading*, and any candidate will be subject to the same sort of counter-example as was presented above.

Of course, the interlocutor could respond in an analagous way, suggesting that there *must be* a characteristic way in which utterances result from beliefs in cases of reading, even if there is a great deal we don't know about it.[23] If we knew more about the brain and the perceptual system, we might be able to formulate an adequate description—but for now, as a placeholder for such an account, we can say that an utterance "…" must be an *r-result* of the state σ for it to count a reading.[24] This is a version of the interlocutor's suggestion in §156 and §158 that our inability to specify the relevant state or causal connection is the result of a "too slight acquaintance with what goes on in the brain and the nervous system", a version of which is endorsed by Wayne Davis in his defence of causal-psychological accounts.[25]

(perhaps as a joke). But they do not *read* it (that's the joke!). And for our teacher, the criteria for reading are derived from the behaviour of the creatures she is training: she knows this player is competent at this game, and so wouldn't have READ "CAT"—and so, she concludes, he must have said it for some other reason. One final scenario: imagine a reading machine that can be programmed to read, rather than one that is trained to read. It has two distinct systems: one for rendering what is written, and one for picking a word at random and rendering that word. When confronted with the word "CAT", the machine is programmed to switch from the first system to the second, so that instead of rendering "CAT" it renders some other word randomly chosen by the system. On very rare occasions, the random system happens to pick the word "CAT" as the one to render, so that it might appear that the machine is malfunctioning, though everything is in fact in order.

[23] Cf. [18, 70]

[24] Note that this was also a possible response when we asked the interlocutor to specify a psychological state σ unique to readers—he might say that, though we are unable to specify such a state now, the progress of neuroscience will eventually allow us to do this.

[25] [18, 70]. The idea of an 'r-result' in the text echoes Davis' strategy:

The motivation behind interlocutor's insistence that there *must* be some such mechanism warrants further attention than I will give it here.[26] For our purposes, the important point about this move is that it simply defers the problem of specifying the relevant kind of connection, and in a way that seems at odds with our everyday use of the concept *reading.* For, as Wittgenstein points out, we are perfectly able to use this concept without this additional knowledge—so why should we think that *reading explanations* depend on it?[27]

This difficulty was already implicit in the bare specification of an "r-result", which amounts to nothing more than the idea of something caused in "the way utterances are caused when you are reading", along with an insistence that this "way" can in principle be specified in other terms. To make the point explicit: this sophisticated version of empiricism shares the same fundamental form as the empiricism suggested by Wittgenstein's interlocutor, except here it is science that is going to 'give' resources that are both independent and ultimately prior to the concept under investigation, rather than sentient experience.

> We know that there is a characteristic way in which intentional action results from desire, even though there is a great deal we do not know about it. I will say that ϕing is an *f-result* of the desire to ϕ when the action results from the desire in that way.

[26] Part of Wittgenstein's response consists in asking whether it is an *a priori* truth that there *must* be some neurological correlate for the process of reading, or whether the existence of such a correlate is "only probable". I take it that part of what Wittgenstein is trying to show here is that the question of whether there is a neural correlate to the process of reading is an empirical question, something we might discover through scientific investigation, but not something that belongs to the concept of *reading.* That is, there is nothing about the concept of *reading* per se, and the kind of intelligibility it reveals in a person's acts, that makes it the case that there *must* be such a correlate—our understanding and application of that concept does not depend on any understanding of, or indeed the truth or falsity of, such neurological claims. To insist otherwise is a philosophical dogma—or so Wittgenstein seems to suggest. For a further exploration of this thought, see e.g. *Zettel* §608-10, and *Blue Book* 117f:

> Note also how sure people are that to the ability to add or multiply or to say a poem by heart etc., there *must* correspond a peculiar state of the person's brain, although on the other hand they know next to nothing about such psycho-physiological correspondences.

[27] Obviously the discussion could continue here: the interlocutor might, for instance, maintain that this is a causal explanation, and that all causal explanations implicitly rely on the possibility of a theoretical account of the causation in question. We can traffic in such explanations without being able to provide such an account, but our practice implicitly relies on its possibility. One way of putting this point is to say that this form of empiricism fails to capture the self-conscious character of reading.

First, the need for such an account in response to questions like (MQ$_R$) and (EQ$_R$), and the form the account takes, both reflect the perspective it takes on its subject-matter. If we 'zoom-in' on particular acts, taking them as given independently of broader circumstances, then the question of what distinguishes those acts *qua* acts of reading (or *qua* instances of acting for a reason) becomes particularly pressing, because from the perspective we have adopted there seems to be nothing that distinguishes the utterance from the utterance of a non-reader. When we ask our version of (MQ), we are now asking in virtue of what one act counts as an act of reading, whereas another apparently similar utterance does not. By the same token, the question of how it is we come to have first-personal knowledge of our acts of reading becomes equally pressing. For since the utterance itself is indistinguishable from that of the non-reader, there seems to be nothing in it to ground the knowledge we have just insofar as we are reading. When we ask our version of (EQ), we are asking what can ground our knowledge of the utterance *qua* act of reading.

This perspective then shapes the proposed responses to these questions. For since the act itself is conceived as indistinguishable from other kinds of utterance, it looks like we need to specify some supplement to the act: ideally one in virtue of which it will count as an act of reading, *and* that will ground our first-personal knowledge of it as such. The pictures of causation and knowledge that the interlocutor tries to apply are then ones that fit these terms. For instance, sensations are occurent and describable independently from broader circumstances, and we can just say that we have them (and what they are like) insofar as we experience them. If acts of reading were always accompanied by such a supplement, we could appeal to it in our answers to (MQ$_R$) and (EQ$_R$). By the same token, the picture of causation provide by 'reading machines' such as the pianola is precisely one on which what makes sound count as an 'act of reading' is the occurent workings of a causal mechanism that can be described independently of the broader circumstances. This again provides a model for the kind of account the interlocutor thinks that he needs: there *must* be some analogue to this in the acts of reading of the beginner and the expert, in virtue of which they count (and, perhaps, are known) as acts of reading.

We can now see that despite its 'sophistication', the approach of our reductive account of reading-explanation shares these basic features. Our reading-explanations isolate a particular utterance, or a causal relation between an utterance and a sign, and ask: in virtue of what does this count as an act of reading? The focus on the particular utterance via the reading-explanation, together with the methodological insistence that the account not deploy the concept 'reading', together ensure that the perspective adopted by the account is analogous to that of Wittgenstein's interlocutor. The account then proposes to answer (MQ_R) by specifying some supplement to the utterance, e.g. a special mental state whose causal role makes this particular utterance count as an act of reading. But the same problem undermines both accounts: nothing they specify that meets the criteria of the account seems sufficient to distinguish cases of reading from some utterance by a non-reader. In chapter 6, I will show that this problem can be found in contemporary reductive accounts of 'acting for a reason' such as the one provided by Kieran Setiya.

5.3 Circumstances, Abilities, and Practices

In §5.1, I summarised §156-178 in a way that suggests that the difficulties the interlocutor finds in answering his questions, along with his understanding of the questions themselves and the urgency of his response, all emerge out of the perspective he adopts on acts of reading. Here, as elsewhere in the text, he zooms in on a particular moment of the act—an utterance made in response to a sign—in a way that cuts it off from the broader and familiar circumstances in which acts of this sort have their home. Wittgenstein's repeated references to these circumstances suggest that our understanding of the interlocutor's difficulties, and the questions that seem to prompt them, will all shift when we have these circumstances in view.

And yet what he means by 'circumstances'—and how attention to them is supposed to help with the metaphysical and epistemological concerns that drive the interlocutor—remains far from clear. For instance, it might seem as though point is that the interlocutor needs to widen his search for a response to his questions, and rather than focusing on the immediate context of the act, look to broader 'circumstances' to provide that in virtue of which the utterance will count as an act of reading. But if the line of thought I have sketched above is correct, this suggestion cannot be what Wittgenstein has in mind. The problem is not simply that the interlocutor is looking in the wrong place for what he wants, so that he could find something of the right sort if he looked elsewhere. Indeed, if I am right, then there is nothing that could provide what he thinks he needs. Rather, whatever this discussion is to teach us about acts of reading must come from our seeing how the interlocutor's understanding of his original questions, and the urgent need he feels for a particular kind of response, all emerge from his initial perspective. Somehow, keeping the 'circumstances' he obscures in view must change our conception of his questions, and in doing so help us achieve a certain kind of clarity about the nature of acts of reading.

It is helpful here to return to the details of Wittgenstein's discussion, and connect them to some of the themes that show up in the surrounding sections of the book. Let's begin with some of the circumstances that Wittgenstein describes, before the interlocutor zooms in on a particular utterance:

> A person, let us say an Englishman, has received at school or at home one of the kinds of education usual among us, and in the course of it has learned to read his native language. Later he reads books, letters, newspapers, and other things.

Speaking in general terms, we might say that the circumstances described here represent the reader as having acquired an ability to read through a familiar form of education. Having acquired this ability, he has been inculcated into a practice of reading, and exercises the ability in the various familiar ways in which reading figures in our lives. The particular acts of reading that the interlocutor 'zooms in' on make sense against this background.

Describing these circumstances in such terms brings out the importance of certain general concepts (ability, practice) whose significance is explored in the surrounding portion of the text. In the sections immediately prior to the discussion of reading, Wittgenstein makes the following observation:

> The grammar of the word “knows” is evidently closely related to that of “can”, “is able to”. But also closely related to that of “understands”. (‘Mastery’ of a technique,) (§150)

The discussion of the ability to ‘read’, abstracting as it does from questions of the role of ‘understanding’ in our everyday concept, can be seen as a partial development of this line of thought. Wittgenstein uses the concept ‘reading’—and, in particular, the comparison between the expert and the novice—as an example through which to explore the grammar involved in our saying that someone ‘is able to do such-and-such’.[28] In fact, the concept ‘reading’ involves a particular version of this grammar, since the ability it describes is one that is acquired through a certain form of training or education, one that makes the reader responsive to the normative relations that characterize a particular practice of reading and writing. The abilities in question are then acquired abilities—and ones that are acquired by inculcation into a particular kind of practice.

These ideas are picked up again in the section that provided the background for much of our discussion in the first chapter, §198. Wittgenstein suggests that we should understand the kind of connection that there is between a sign and my actions in following it by reference to the fact that “I have been trained to react to this sign in a particular way, and now I do so react to it” (§198). When the interlocutor complains that this is “only to give a causal connexion”, Wittgenstein responds:

[28] Baker and Hacker bracket the sections on reading and guidance together with earlier sections from §143ff., giving them the title ‘Understanding and Ability’, and treating our sections as clarifying the grammar of abilities by reference to the idea of reading. See also the discussion of abilities in *The Brown Book*, which transitions into an earlier version of the sections on reading.

> 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. (§198)

This is followed, in §199, by the observation that "[t]o obey a rule, to make a report, to give an order, to play a game of chess, are customs (uses, institutions)", and then again in §202, which begins "hence also 'obeying a rule' is a practice". The suggestion, I take it, is that if we want to attain clarity about the nature of any of these acts, we can begin by seeing them as manifestations of abilities or capacities that belong to us insofar as we are trained participants in particular kinds of practice.

In the rest of this section, I will build on what I take to be the key ideas expressed in these comments. I shall argue that 'reading-explanations' represent their targets as acts of the ability to read, and that this ability, and the acts that manifest it, cannot be characterized independently of certain features of the lives of its bearers that show them to be readers. These features belong together with the idea that reading is a particular kind of practice, and that the ability manifested in acts of reading is one that belongs to its bearers as members of that practice. This marks the ability to read as a distinctive kind of disposition, and reveals acts of reading as a distinctive kind of causal transaction manifesting that disposition. It will then turn out that reading-explanations can be understood to be a kind of causal-explanation, but only if we are clear about the ways in which the disposition they appeal to is essentially different from other kinds of causal disposition.

5.3.1 §157: Living Reading-Machines

Our aim, then, is to see how the idea of a practice could be essential to a characterization of the ability to read, and how this point might shape our understanding of reading-explanations and the acts they represent. To this end, it will be helpful to compare it with a range of similar abilities and dispositions that can be characterized in simpler terms.

To begin with, consider the way in which Wittgenstein introduces the idea of an 'ability' into the discussion of *reading* in §157ff. As we have seen, the interlocutor is looking for an independently-specifiable, occurent, and self-identifying feature in virtue of which an utterance will count as an act of reading. But if reading consisted in the occurrence of such a feature, then the acquisition of the ability to read would be marked by an abrupt transition from cases that lacked this feature, and so did not count as ‘reading’, to cases that had it, and so did. It would then make sense “to speak of the *first* word that he really read”, since it would be the first utterance that had this feature. But questions of this sort are not part of how we talk about acquired abilities like reading:

> When did he begin to read? Which was the first word that he *read*? This question makes no sense here.

The ‘grammar’ of a concept like reading—e.g. the kind of questions it makes sense to ask about acts of reading—is that of a particular kind of ability, and as such fundamentally different from the ‘grammar’ of the kind of independently-specifiable descriptions of mechanisms or occurent states that the interlocutor relies on. Wittgenstein notes that our description of the capacities of a machine might have the grammar the interlocutor deploys, and there it could make sense to say that the machine read only “after such-and-such parts had been connected by wires”, or that “the first word it read was . . . ”. But when we describe the abilities of a living being, *reading* means “reacting to written signs in such-and-such ways”, and the change that takes place is “a change in his *behaviour*”. The concept that we use to characterise this behaviour is “quite independent of that of a mental or other mechanism” (§157).

Building on the scenario described in §157, in which “human beings or creatures of some other kind are used by us as reading machines”, we can use this same strategy to introduce and compare further notions of ‘disposition’ and ‘ability’. So far we have a general point about the grammar of abilities as they figure in the description of acts of living beings, but subsequent sections develop further features that are more specific to abilities acquired through inculcation into a practice, and particularly into a practice like that of reading.

Let's begin by imagining creatures with a simpler version of such an ability. A widely reported study, published in the September 2016 issue of the *Proceedings of the National Academy of Sciences*, showed that pigeons can be trained to correctly differentiate between up to 60 'words' and 1000 'non-words'. The four-letter words were presented to the pigeons on a computer screen, and the birds were supposed to peck on the sign itself if presented with an English word, or peck on a different symbol if presented with a non-word. Popular reporting of the story described the results of the experiment in terms reminiscent of Wittgenstein's own thought-experiment: the pigeons can 'read', sort of, even if they can't understand what they read![29]

Case 1 To start with, imagine the experimenters in the early stages of their work, simply showing various pigeons four-letter combinations on a screen and observing their reactions. Suppose these scientists noticed that one particular pigeon reliably showed a differential response to some combinations: perhaps it almost always pecked at the screen when the sign shown was 'FCHD', and almost always pecked at the symbol when the sign shown was 'QWER'. After sufficient observation of this pattern of behaviour, our observers would be in a position to attribute a basic disposition to this bird, i.e. one that manifested itself in the simple patterns of behaviour I have just described.

This application of the concept 'disposition' has a number of notable features. First, the disposition is characterized solely by reference to observed patterns of behaviour of a particular bird, and these patterns are an essential part of the 'circumstances' in which it makes sense to speak of such a disposition. Second, attribution of the disposition brings with it the expectation that this pattern of behaviour will continue, and belongs together with e.g. looking for some further explanation in the cases where the pigeon does not behave in this

[29] The original paper states that the experiment yielded interesting hypotheses about precisely how pigeons learnt to make this kind of differentiation. According to the researchers, the pigeons did not simply 'memorize' the correct words, but rather "picked up on the orthographic properties that define words and used this knowledge to identify words they had never seen before". However, given my aims, the results of the experiment are not directly relevant

way. But the only normativity so far associated with this concept comes from the observed past behaviour, and the disposition we have described in terms of it—characterization of such a disposition needn't involve reference to anything besides the bearer and the fact of its previous patterns of response.[30] Given the attribution of such a disposition, certain behaviour is 'to-be-expected' of the bird (i.e. we make predictive judgments like 'This bird will peck the screen when shown this combination'), whereas other behaviour might stand in need of further explanation (i.e. if a bird that has such a disposition fails to behave in the expected way, it makes sense to ask why).[31]

This is the kind of disposition to which we shall see Setiya appeal in his causal-psychological account: a reliable transition from one act to another. We shall see in §6 that such a disposition provides a weak basis for an account of acting for a reason, since such dispositions come cheaply and provide fertile ground for formulating deviant cases. Setiya proposes that the solution lay in specifying the disposition in terms of the right causal input: the relevant disposition is distinguished by its dependence on his self-referential belief SR, and it will be this—rather than the form of the disposition itself—that makes it part of Setiya's account.

Case 2 We can contrast this case with a later scenario, in which the scientists have begun to train their pigeons by rewarding them if they differentiate certain four-letter combinations from others. Suppose the scientists picked the combinations that are 'to-be-rewarded' at random, and then drew up a list of these 'words' and trained the birds accordingly.[32] In this

[30] This ignores questions about how we should describe the 'signs' to which the pigeon responds. Describing them as composed of letters, and figuring out what counts as the same reaction to the *same sign*, arguably involves circumstances beyond the bird and its past behaviour. See below.

[31] A more complex concept of a disposition would involve reference to the kind of thing this creature is, i.e. to a disposition that could be characterized by saying 'Xs do A in circumstance C', or 'The X does A in circumstance C'. If we took ourselves to be providing the basis for natural historical judgments about creatures of this kind, then our concept of the disposition would involve reference to something beyond this particular creature and its pattern of response: it would involve reference to the kind of which this creature is an instance (see [35]). This contrasts with the case described above, where a description of the disposition only involves reference to the patterns of behaviour of a particular individual.

[32] In the actual experiment, the scientists rewarded the birds for differentiating between four-letter combinations that are English words, and four-letter combinations that are not. However, it will be helpful if we start by imagining the scenario as I describe it in the text above.

case, the ‘disposition’ in question has features that set it apart from the one described above. First, the disposition we are now concerned with is not characterized by reference to habits of past behaviour, but rather by the behaviour that accords with the list. Second, attribution of the disposition brings with it both the expectation that this pattern of behaviour will continue, and a standard against which any future behaviour can be judged as correct or incorrect.

One essential feature of the ‘circumstances’ in which it makes sense to speak of such a disposition is the role played by the list in training these birds, since here the criteria for possession of the disposition emerge not from a description of habits of past behaviour, but rather from the inculcation of patterns of response that accord with the ‘rules’ implicit in this list. Here we characterize the disposition by reference to those ‘rules’ (the screen is ‘to-be-pecked’ whenever a word on the list is shown), and such a characterization now involves the idea of *correct* behaviour, such that the disposition provides a standard by which we can evaluate, not simply predict, what the creature does. This brings with it further possibilities of judgment: the pigeon’s response to a particular combination is now not simply ‘to-be-expected’, but ‘to-be-done’ or correct. Moreover, if we ask ‘why did the pigeon peck the screen’, that act can be explained *qua* act of this kind of disposition by reference to the appearance of the relevant sign, the list, and the conditioning of the pigeon. Given knowledge of the list and the training, appeal to the sign is by itself sufficient to explain the act *qua* manifestation of the disposition.

The normativity involved in this kind of disposition—embodied by the role played by the list in training and evaluating the birds’ acts—already marks a difference between this and the previous case. For now an account of what it is to be an act of the relevant disposition essentially involves an appeal to the list, both insofar as it plays a particular kind of role in the training that inculcates the disposition, and insofar as we implicitly rely on it in describing and evaluating acts as manifestations of the disposition.

Case 3 The disposition inculcated in the actual experiment was similar to the one I have just described, except that rather than using an arbitrary list, the pigeons were trained to differentiate between four-letter combinations that are English words, and combinations that are not. Since the pigeons' disposition apparently involved a capacity to differentiate between novel cases based on orthographic properties of the signs, it was described (in the popular press at least) as a sort of primitive version to the human ability to read. But whatever the usefulness of this comparison for scientific purposes, the disposition I have described here involves a fundamental difference with even the simple abilities that Wittgenstein describes in introducing his concept of 'reading'. For unlike the examples Wittgenstein lists in §156, the version of the pigeon's disposition I have described is simply a capacity for differential response, and involves no idea that the bearer's responses have a 'contentful' relation to their stimuli, in the sense of e.g. rendering what they say.[33]

[33] In the case as we have imagined it, the only connection between stimulus and response are the arbitrary rules we set up as a basis for training. Of course, we can imagine further cases, in which the 'correctness' or normativity involved in the disposition has its source in something other than arbitrary rules, but still lacks this further dimension of e.g. 'rendering what is written'.

Jumping ahead for a moment to a fuller version of our concept of reading, the case that Wittgenstein describes in §166, in which we move from arbitrarily associating a certain sound with a particular symbol to habitually using it in this way, is one such transition:

> Imagine having to use this [arbitrary] mark regularly as a letter; so that you got used to uttering a particular sound at the sight of it, say the sound "sh". Can we say anything but that after a while this sound comes automatically when we look at the mark?

The fact that the mark changes from seeming unfamiliar to being something we respond to automatically is an important feature of the transition Wittgenstein describes here, just as familiarity marks an important difference between cases in which we arbitrarily respond to a series of signs, and cases in which we are reading them. Wittgenstein emphasises the importance of such experiences in comparing these cases in the next section:

> [T]here is certainly some uniformity in the experience of reading a page of print. For the process is a uniform one. And it is quite easy to understand that there is a difference between this process and one of, say, letting words occur to one at the sight of arbitrary marks.—For the mere look of a printed line is itself extremely characteristic—it presents, that is, a quite special appearance, the letters all roughly the same size, akin in shape too, and always recurring; most of the words constantly repeated and enormously familiar to us, like well-known faces.—Think of the uneasiness we feel when the spelling of a word is changed. (And of the still stronger feelings that questions about the spelling of words have aroused.) Of course, not all signs have impressed themselves on us so strongly. A sign in the algebra of logic for instance can be replaced by any other one without exciting a strong reaction in us.—
>
> Remember that the look of a word is familiar to us in the same kind of way as its sound. (§167)

This marks a point of transition to the kinds of abilities Wittgenstein describes in characterizing the concept of 'reading' he means to investigate. Each of the those abilities involves some kind of 'contentful' relationship between the 'stimulus' and the response: in rendering out loud what is written or printed, in copying from a page or writing out a dictation, in playing from a score, etc., our acts correspond to the 'stimuli' in a way that involves something more than mere differential response: we (for instance) say what is written, copy what is said, etc.[34] Rather than thinking of these as merely differential dispositions, we might think of them as abilities to "*derive* the reproduction from the original", as Wittgenstein puts it in §162:[35]

> Now suppose we have, for example, taught someone the Cyrillic alphabet, and told him how to pronounce each letter. Next we put a passage before him and he reads it, pronouncing every letter as we have taught him. In this case we shall very likely say that he derives the sound of a word from the written pattern by the rule that we have given him. And this is also a clear case of *reading*. (We might say that we had taught him the 'rule of the alphabet'.) §162

In criticizing the interlocutor's emphasis on felt or conscious experience, Wittgenstein does not mean to deny that the occurrence of such experiences, or the transitions between them, might play an important part in our life as readers. But the logical character of the transition is revealed, not in the various feelings that accompany it, but in the kind of questions we ask about the mark, or the kind of things it makes sense to us to say about it. Returning to the case in §166, Wittgenstein notes that to say that I now respond to the mark automatically "is to say: I no longer ask myself on seeing it "What sort of letter is that?"—nor, of course, do I tell myself "This mark makes me want to utter the sound 'sh"', nor yet "This mark somehow reminds me of the sound 'sh"'."

[34] An analogous experimental case might involve a creature who did not simply respond differentially to the signs, but whose response involved a 'structure' that could be mapped on to the signs. Such cases needn't involve 'language' in the familiar sense—we can imagine a creature that, say, produced a series of noises whose differential pitch corresponded to that of notes written on a stave. Of course, here too we can imagine a range of cases: a more basic one might involve simply copying lines or shapes from one piece of paper to another, i.e. 'stimuli' whose 'content' is fairly minimal. Nevertheless, even in this basic case we might say that the creature was *copying what was drawn*, rather than simply responding to it—the former notion being a more 'sophisticated' version of the latter.

[35] Cf:

> The difference between the meanings of "associate" and "copy" shows itself in the fact that it doesn't make sense to speak of a projection-method (rule of translation) for association. We say: "you haven't copied correctly", but not "you haven't associated correctly". PG p.92

The normativity involved in such abilities is more complex than the cases described above. First, the ability is characterized by reference to some system of norms for correct rendering—how it is that particular shapes are to-be-copied, written notes are to-be-played, signs are to-be-said. In the case we are primarily concerned with, the norms will be those belonging to the alphabetical representation of words. However, given such a 'system' of norms, it is the sign itself that provides the standard by which we predict or evaluate the acts of the disposition. A written sign "C-A-T" is to-be-read as "kæt", according to the alphabetical system, and while it is the system that determines how the sign is to-be-read, it is the sign itself that provides the standard by which the act is evaluated, since the disposition we are concerned with involves *reading what is written.* What counts as a correct response in this case is specifically determined by the letters on the page, since now a correct response to the sign *says what it says.* Compare this to the case described above, where the birds are to peck at the screen when shown a word on our list: if the word "C-A-T" is on the list, then a pigeon that pecks at the screen when shown that word is responding to the sign correctly, but the description of that response that represents it as an act of the disposition is available independently from a description of the sign.

Since we are concerned with the concept 'reading', we can think of this as the difference between *responding correctly to a sign*, where the sign calls for some (perhaps arbitrary) differential response, and *saying what the sign says*, where a description of the latter acts involves essential reference to the 'content' of the sign.[36] As we are still focused on the attenuated version of this concept, which abstracts from the idea of understanding what is written, this 'content' will simply involve some kind of isomorphism between the signs and the sounds produced in response to them: the sound "kæt" is to-be-said in response to the written markings "C-A-T", and the sound and the markings *'say' the same thing.* Even

[36] We can imagine a series of transitional cases here. For instance, in the original scenario the birds respond in the same way to any 'word' on the list, but we can imagine a more complex disposition that involved different responses to different words, but without any notion that there was a structure in the response that was derived from, or corresponded to, the 'word' on the screen.

though we are abstracting away from any understanding of this word, or the role it plays in a sentence, we can still describe this attenuated ability as an ability to ‘say what is written’ in the sense of producing sounds that correspond to the markings in a different way than a merely differential response could be said to ‘correspond’ to them.

Here the idea of a ‘rule of the alphabet’ is playing a role analagous to that of the list in our previous case, which means the ‘circumstances’ in which it makes sense to speak of such an ability must be such that it also makes sense to speak of an appropriate system of rules.[37] Given such circumstances, an account of what is it to be an act of the relevant disposition essentially involves an appeal to the idea of such a system, both insofar as the principles of the alphabet play a particular kind of role in the training that inculcates the disposition, and also insofar as we implicitly rely on those rules in describing and evaluating the acts that manifest the disposition.[38] It is a further feature of such a system of rules that they generate an indefinite number of instances, and that the characterization of the attendant ability involves a certain sense of ‘going on in the same way’. Whereas the ‘sameness of response’ of our earlier dispositions involved acts that were all tokens of the same type, a system of rules such as this brings with it the possibility that someone could have the relevant ability, and exercise it correctly in response to a sign she had never encountered before, so long as that sign could be rendered in terms of the relevant rules. Whereas in our previous case the disposition was understood by reference to a finite list of words used by the experimenters in training, this case is different, since we are supposing that the creatures master a system of rules that can be applied in an indefinite number of cases. Rather than simply responding differentially to a set number of signs, they learn to render those signs

[37] In speaking of a ‘rule of the alphabet’ or a system of rules, I am simplifying the way that alphabetical writing renders spoken language.

[38] In *The Brown Book*, Wittgenstein calls the kind of training involved in such cases ‘general training’:

> General trainings form a family whose members differ greatly from one another. The kind of thing I’m thinking of now mainly consists: a) of a training in a limited range of actions, b) of giving the pupil a lead to extend this range, and c) of random exercises and tests. (98)

according to a system of rules. This means that they can manifest their ability in response to any set of signs that conform to the rules of this system.

Case 4 The dispositions I described above needn't involve a capacity on the part of the 'living reading-machine' to either explain or correct it's utterances. To get something closer to *our* capacity to read, we could imagine that it is part of the training that inculcates the disposition that the creatures learn to respond to questions such as

> Why did you say . . . ?

by pointing to the mark they had rendered. Creatures who could do this much *might* also be able to go on to grasp the idea that they had made a mistake in their rendering. For instance, if the teacher responded,

> No! Try again,

they would respond to the sign again until they were told they had it right, and understand that this marked a difference in their response.[39] This could be further reflected in their capacity to check the responses of others as 'correct' or 'incorrect', and evaluate them accordingly:

> *S* read what is written correctly.

Recognizing a disposition as having this character involves attending to the more complicated form of its training, and the further features that belong together as manifestations of the disposition. Now it is not just the utterances that render the words that count as its manifestations, but also the 'explanations' and 'evaluations' that the bearer can produce insofar as she has the disposition.

[39] In contrast, we can imagine creatures that also responded to 'No! Try again' by rendering the same sign, and sometimes rendered it differently on a second try, but showed no understanding of the idea that one rendering was incorrect and the other correct.

Case 5 We might, if we wanted, go further and try to characterize the ability to read in its more familiar guise. Here, in the paradigmatic case at least, an act of the ability involves not simply rendering what is on the page, but understanding it as well. The normativity that belongs to this ability is still more complicated: in the earlier case, the ability involved knowing that "C-A-T" is to-be-read "kæt", whereas when *we* read we typically recognize it as part of a sentence that says e.g. that the cat is on the mat. Thus, in characterizing a fully-fledged ability to read, we should not simply describe its acts as involving knowledge that a particular sign is 'to-be-read' in a particular way; rather, it involves knowledge that the sign *says* such-and-such. The easiest way to get this difference in view is to imagine the difference between reading a sentence in a language one does not understand, and reading a sentence in a language one speaks: in the former case, there is a sense in which one 'knows what the sentence says', in that one can render it correctly according to the rules of the alphabet. But this is clearly different from the case in which one understands the language, and thus recognizes the sentence as e.g. saying that such-and-such. If one does not understand the language, one's act will still be internally related to the sign in the way described above, since it is an act of rendering what is written. But it will fall short of being a paradigmatic act of our familiar ability to read, since it does not involve an understanding of what is written, and thus no recognition of 'what it says' in the fuller sense.

This suggests another fundamental difference with the previous disposition and the other abilities that Wittgenstein focuses on in his discussion. An ability to read, in the attenuated sense, involves correctly rendering what is written according to the rules governing, say, an alphabetical system of writing, and an act of such an ability is evaluated according to whether it applies those rules correctly to the particular sign. In contrast, an ability to read, in the fuller sense, involves recognizing or understanding *what is written*, and we might judge an act of reading by how clearly it articulates that sense. Indeed, given the place that reading has in our lives, we can go on to imagine further complicated criteria by which we might judge acts of this ability, e.g. we might judge how well a person reads a poem according to our understanding of 'what the poem says', where such understanding involves more than merely knowing what the words mean.[40] The background against which we must describe the acts is no longer simply that of an alphabetical system, but rather that of a language in which things can be said.

This last cases emphasizes something important about Wittgenstein's discussion. His attenuated concept of *reading* draws our attention to particular aspects of our everyday concept, and in doing so leads us to temporarily set aside some of its other features. However, even our understanding of the attenuated concept, and the cases we describe in terms of it (i.e. the 'living reading-machines') *depend* on our having the concept of reading-with-understanding, since it is that that gives us the idea that what-is-read belongs to a language, and so says something. This fact has shaped our descriptions of every case. For instance, it makes sense to us to call the four-letter combinations that the pigeons respond to 'words', or to describe an isomorphic response as 'saying what is written', because we set it alongside our practice, and see one in terms of the other. This means that we see the marks as the kind of thing that belong to a *language*, one that can be used to do things like *say that such-and-such is the case*. Even the idea of an alphabetical system implicitly depended on this notion, since when we think of such a system we think of it as a means for representing not any old sound, but what can be said in a particular language. [41] As Rush Rhees put it, "[w]e cannot

[40] Lest this seem like a purely linguistic phenomena, note that we can make similar points about an ability to produces notes of a particular pitch. Indeed, we could re-describe this series of experiments in these terms, gradually building up to the idea of a creature that produces particular melodies in response to combinations of notes written on a score. Here too we can distinguish between the case in which a creature can reproduce notes 'mechanically' or 'without understanding', and a case in which we might say they were really e.g. *playing music*, and doing so with understanding. Getting clear about what it means to make 'music' that can be 'understood' would involve reflecting on the place that music has in our lives. On these points, cf. PI §527, but also §22.

[41] Given this, we can see how if we came across a creature that showed some such capacity for differential or even isomorphic response, but in circumstances quite different from the familiar surroundings of human life, we might not want to describe its ability in these terms. Wittgenstein describes a related case in §207:

> Let us imagine that the people in that country carried on the usual human activities and in the course of them employed, apparently, an articulate language. If we watch their behaviour we find it intelligible, it seems 'logical'. But when we try to learn their language we find it impossible to do so. For there is no regular connexion between what they say, the sounds they make, and their actions; but still these sounds are not superfluous, for if we gag one of the

say the man is reading except in connexion with certain ways of living: where people inscribe monuments, post public proclamations, keep records, write reports, write letters, etc., etc." [28, 49]. This indicates the tremendous complexity of the circumstances surrounding our own life with the concept 'reading'—something which, as Wittgenstein points out, "would be very difficult to describe even in rough outline" (§156). And it further suggests that, ultimately, it is only against this background that we see acts of reading as the kind of acts that they are.

Summary of Cases

A characterization of the acts of each disposition involved some appeal to 'circumstances' beyond those acts. In the first case, where the birds simply showed some apparently spontaneous differential pattern of response, seeing the particular act as a manifestation of the disposition involved seeing it in terms of that broader pattern. In this case, though a particular response was 'to-be-expected' insofar as it maintained the pattern, or in need of further explanation if it broke with it, there was no stronger idea that the birds were getting things right or wrong in responding in a particular way. This disposition was therefore comparable to 'physical dispositions' like solubility or fragility, i.e. dispositions that involve a predictable pattern of response.

In the second case, where the birds had been trained to respond differentially to particular words, seeing the particular acts as manifestations of the disposition involved seeing them

> people, it has the same consequences as with us; without the sounds their actions fall into confusion—as I feel like putting it.
>
> Are we to say that these people have a language: orders, reports, and the rest?
>
> There is not enough regularity for us to call it "language".

Even in our penultimate case, which most closely approximated the attenuated concept of 'reading' that Wittgenstein focuses on, it made sense to us to say that the creature 'said what was written' because we imagined that case as though the creature were reading something written in a language it did not understand. But if we re-imagine it in the circumstances described in §207, where despite a regularity of response we had no sense that the markings had any connection with the language of living creatures, we might be less inclined to describe it as 'saying what was written', no matter how complex the isomorphic relations between markings and sounds.

in terms of the list of words, and the training it was used in. Now a particular act was not simply 'to-be-expected' given the previous pattern of response, but could also be evaluated as correct or incorrect relative to the list of words. The disposition differed from the previous case because it involved this more robust notion of normativity. But this difference was not constituted any distinctive occurrence at the time of the act. Considered by themselves, there need be nothing to differentiate the responses of birds with this disposition from those in the previous case. It is only when we consider those acts along with the word-list and the training that the relevant patterns become apparent.

In the third case, we added a further aspect to the acts of the disposition: that they *rendered* the signs that prompted them. By itself, this involved the idea of a system of rules, and with them more complex patterns of differential response, that could be systematically related to differences in the signs. Unlike the previous case, this does seem to involve a difference that could be apparent in the act itself, since one will see the act as rendering what is written. But seeing a particular act in these terms involves seeing it together with the idea of a system of rules that give a method of projection between sign and utterance. Moreover, even this response could be indistinguishable in itself from one that *seemed* to render what was written, made by accident by a creature who lacked the disposition. It is the broader circumstances—the rules for rendering, the training that inculcates them, and the previous pattern of response—that give us the idea that one of these acts manifests a distinctive kind of disposition.

Of course, ultimately we are concerned with the more profound transformation involved in the idea that the signs do not simply *render what is marked down*, but rather *say what it says*. This latter involves the idea that the signs belong to a language. The 'circumstances' involved in making sense of this idea are far more complicated than any we have considered so far. But the basic point is the same: understanding a particular act as manifesting an ability to read involves seeing it together with the broader circumstances that show that the marks in question were signs belonging to a language, and that the utterance manifested a disposition to read that language.

5.3.2 Reading as a Practice

If representing something as an act of reading involves seeing it as an act of the kind of disposition described in case 5, then our grasp on the unity that characterizes particular acts depends on our seeing them as part of a practice of reading and writing. For our grasp on the explanatory and normative relations that characterize the act as a whole, and by extension our grasp on the kind of thing that constitutes its parts, both depend on our being readers.

These claims will become clearer we if think about how these acts look when we do not recognize them as part of a practice of reading: seen from outside such a practice, the nature of the acts—and with it, the normative or explanatory relationships that characterize them—are all obscured. The following two examples help illustrate this. First, imagine a case in which we, who are able to read, came across someone staring at some strange marks on a wall and making a series of unintelligible noises; after observing them for a few moments, we cannot work out what they are up to; but on being told that the markings are a cuneiform script, and the person was reading them, everything falls into place. It would only be once we understood the situation in these terms that it would be helpful to be told that e.g. they made *this* sound because of *this* marking. Noting a 'merely causal' connection between the markings and the noises would not help us understand what was going on. For in our puzzlement we could already see that there was *some* connection between the markings and their noises: our problem was that we did not understand its nature.

Compare this with a second case in which the observer is someone who does not know what it is to read. Such a person, coming across someone reading in a foreign language, might ask "why did he make that noise?". How might we, who can read, explain what was going on? Simply saying "He made it because of that marking on the wall", while true, would only give the inquirer a partial understanding of what was going on. She might, for instance, decide to join in the fun, pointing to different markings and making her own noises.

The only way we could get her to understand would be to teach her what *reading* is—and, arguably at least, a full grasp of that concept requires learning *how to read.*

Of course, if one knows something is an act of *reading*, one can begin to trace the causal-explanatory relations that belong to it *qua* such an act: the cause of the reader's utterances are the written words she is reading.[42] Here our ability to trace these causalities depends on our understanding that these are acts of reading.

This might seem like a purely epistemic point: after all, we said above that an observer who could not read might notice that the reader's acts were somehow responses to particular markings, and thus come to understand that there was some sort of causal relation between marking and utterance. Being able to *read* the language would certainly make tracing the causal relation between marking and utterance easier, but couldn't our observer, with enough patient observation, trace the same causal transactions by noting regularities of response?

One way of seeing the problem with this story is by returning to the distinction between paradigm, defective, and deviant cases. Our observer must be able to differentiate the first two from the latter; but because the observer cannot read the language, she does not know, of any particular transaction, which category it belongs to. Did the reader render the sign correctly? Did she make a mistake, and if so did she correct herself and move on, or did she fail to notice it? Did she interject a comment into her reading? Or perhaps some other kind of exclamation? Did she respond to *this* sign, or is it just an accidental marking? For these questions to so much as make sense, the observer needs to have *some* conception of what it is to read (or, at a bare minimum, that there is such a thing as *getting things right or wrong* in response to the signs). Of course, once she had some such conception, the patient observer might find grounds for answering some of these questions. But ultimately, to complete her task with confidence she must *learn the language herself.*

[42] We do not need to imagine the more fantastic cases described above to see this. I might proof-read this text by having someone else read it back to me: if they said something that sounded wrong, I would look to the words on the page to see whether I had made a mistake in my writing, or if they had made a mistake in reading.

If the goal is simply to trace causal relations between marking and utterance, regardless of their character, then noting regularities of response might be sufficient, though those regularities would almost inevitably include deviant cases. But the claim we are considering is that reading involves a distinctive *kind* of causal transaction that can only take place between a reader and a certain kind of marking. The point is not simply that reliably identifying this kind of transaction is improved by knowledge of the normativity that belongs to *reading*. It is rather that one cannot so much as make sense of the *kind* of transaction that it is without that knowledge. An outside observer who somehow managed to pick out all and only the causal transactions involved in acts of reading would still be in the dark as to something essential about the character of those transactions.[43]

What we see in these cases is that the intelligibility of the act is obscured from view by the fact that our observers are unable to locate it within a practice of reading. This point encompasses not merely the explanatory connection between a marking and an utterance, but also something essential about the marking and the utterance themselves. For instance, if one were unfamiliar with any practice of reading, one would not understand the sense in which the particular markings to which the person was responding constituted a *sign* of a particular sort. Someone who lacked any concept of reading and writing, or who was perhaps only familiar with more primitive uses of written marks to track goods and the like, might have no idea that such markings together amounted to a *word*, nor that they were individually e.g. letters of an *alphabet*. Nor would they understand that the utterance was an act of a particular kind, i.e. an act of saying what was written. Though they might be able to discern that there was some kind of causal connection between the markings and the utterances, the nature of the whole—that is, the act, and with it the character of both cause and causal-explanatory relation—would be obscured from view.

[43] Of course, this does not rule out the possibility of other kinds of causal transactions between sign and reader, nor other causal transactions between the signs and people who cannot read—and insofar as the causal transactions in all these cases involve a person responding to the signs, there will be certain resemblances and commonalities between them. But for all such commonalities, reading is in some sense *sui generis*, because of the kind of normativity involved in it.

5.3.3 The Variety Encompassed by Reading Explanations

These points help make sense of a further feature of Wittgenstein's discussion of reading that relates to its causal character: the difficulty the interlocutor has in making out the idea that acts of reading are 'guided by', 'derived from', or 'caused by' written signs. The interlocutor's difficulties emerge from a familiar source: he focuses in on one instance to which we would apply this term, and tries to understand all other cases on the same model.[44] The problem is that the features he finds in one case are lacking in another: the interlocutor looks for some feature in virtue of which the relevant description can be said to apply, but any such feature he finds is only present in particular cases. For instance, in §162 the interlocutor proposes that "You are reading when you *derive* the reproduction from the original", and goes on to try to specify what he means by 'derivation'. But the features he finds in any one case (e.g. reading off from a table according to a particular rule) seem to be missing in another. Wittgenstein summarises the situation in §164:

> In case (162) the meaning of the word "to derive" stood out clearly. But we told ourselves that this was only a quite special case of deriving; deriving in a quite special garb, which had to be stripped from it if we wanted to see the essence of deriving. So we stripped those particular coverings off; but then deriving itself disappeared.—In order to find the real artichoke, we divested it of its leaves. For certainly (162) was a special case of deriving; what is essential to deriving, however, was not hidden beneath the surface of this case, but this 'surface' was one case out of the family of cases of deriving.

When Anscombe discusses these sections, she states that:

> We repeatedly have as an argument against explaining *reading*, or *drawing*, or *influence*, or *being guided* in some way that is supposed to apply quite generally,

[44] Cf. our discussion of this kind of mistake in §4.2.2.

> that our cases are particular and that cases vary according to circumstances, and our ‘explanation’ is not borne out in a different sort of case. ...
>
> In short, the whole enquiry of these pages consists largely in rather convincing arguments against generalizing particular expressions that we are inclined to use in highly particular situations and cases. [10, 212]

These claims provide important background to my claim that reading is the act of a distinctive kind of capacity. For they suggest that what makes it distinctive needn’t be any particular feature of a given instance of reading. In fact, the ‘acts’ in question will encompass a great deal of variety. For instance, “the beginner reads the words laboriously by spelling them out”, whereas the expert “reads certain words by taking in their printed shapes as wholes; others when his eye has taken in the first syllables; other he reads syllable by syllable, and an occasional one perhaps letter by letter” (§156). The particular ‘processes’ involved in each case might then look very different: we can imagine the beginner tracing each syllable out with his finger, so that we can clearly see each utterance as a response to the particular grouping of letters on the page. By contrast, the expert might not ‘take in’ the whole word before uttering it—indeed, if he is familiar with the text, or understands what he is reading, he may skim over particular passages, so that we might say that he bases his response on the overall sense of the sentence, rather than close attention to each particular word.

If we imagine such a beginner and an expert reading the same passage, then their particular processes (i.e. what happens as they read) could be quite different. We can imagine connecting each of the beginner’s utterances to a particular written word on the page, so that we would be inclined to explain each utterance individually:

> *S* said ““kæt” because “C-A-T” is written on the page.

In contrast, it might make more sense to us to think of the expert as responding to particular phrases or even whole sentences:

S said “The cat is on the mat” because that is what is written on the page.

Both are cases of reading, and therefore a particular kind of response to what is written—but we do not count each of these as acts of reading in virtue of some particular, independently-describable feature of each utterance or the process that led up to it. Unlike the workings of a mechanism, the ‘process’ leading up to a particular utterance, and aspects of the character of the utterance itself, can be quite different from case to case.

Thus if we speak of reading as a distinctive kind of act, we must be careful not to think that we can say what is distinctive about that act by simply describing what happened independently of our counting it as an act of reading. Indeed, if our attention to particular utterances is not guided by that concept, we might describe and categorise those same acts in some other way, so that (from this perspective) the utterances of the beginner look like a *different kind* of act from those of the expert.[45]

5.3.4 (Mental) Causation and Reading Explanation

These points help clarify the relationship between reading explanations and the causal-psychological explanations that the reductivist focused on in her account. A reading-explanation,

RE: ‘*S* said “. . . ” because$_R$ “_____” was written on the page,

or better,

[45] For instance, the beginner’s response might involve quite specific ‘mental causes’, which he can recall and describe as part of his process when he reads, whereas the expert might be unable to remember anything particular about reading a given word. The beginner’s reading explanation might then be thought of as intrinsically-related to a further causal explanations, e.g.

S said “. . . ” because she mentally sounded out each syllable

which describe aspects of the same act described by a reading-explanation,

S said “. . . ” because$_R$ “_____” was written on the page

In contrast, the expert’s response might happen automatically, so that there was nothing that his act consisted in besides his utterance in response to the signs.

RE: *S* read, "...".

intrinsically involves a 'mental-cause' explanation,

PRE: '*S* said "..." because she saw the words on the page,

along with a straight-forward causal explanation,

CE: '*S* said "..." because of the words on the page.

But while (RE) specifies a 'reason$_R$' for the agent's action, (PRE) and (CE) do not: e.g. *that such-and-such was written on the page* is a 'reason$_R$', in contrast to *that she saw such-and-such written on the page.* Nevertheless, to fully understand the significance of both (PRE) and (CE), we need to see them as internally-related to (RE): the sight of the words on the page caused our agent to say something because she was reading what was written.

Of course, a causal-psychological explanation like (PRE) (or a causal explanation like (CE)) could be offered for any of the deviant cases that came up in our discussion of §§156-171. For each of those cases depended on the idea of an agent seeing signs written on the page, and responding to them in a particular way, without *reading* what was written. The suggestion that we are now considering is that what distinguishes (PRE) from these other cases is that it describes an act of the ability to read, and that this marks what it describes as a distinctive kind of causal transaction—one which needs to be understood by reference to the normativity that characterizes a practice of reading. (RE), (PRE), and (CE) would then all describe the same causal transaction, with only the first explicitly representing it as the kind of causal transaction that it is, i.e. as an act of the ability to read.

This helps us situate the account of reading explanation relative to the arguments made by Jonathan Dancy about the distinctive character of 'rational explanations', and their relationship to what look like causal-psychological explanations of the same acts. Of course, Dancy is not providing an account of 'reading', but if we transpose his arguments into this discussion we get the following two claims:

- (a) Reading-Explanations are a distinctive kind of explanation, since the description in the *explanans* needn't be true (i.e. defective cases are possible).
- (b) Causal-psychological 'explanations' like (PRE) in fact represent 'enabling conditions'; only (RE) is a genuine explanation of the act.

Let's begin with the second of these points. Acts of the ability to read intrinsically involve the causal transactions described in explanations like (PRE): to read is to see what is written on the page and to utter words in response. But (PRE) obscures something essential about the causal transaction it describes, i.e. the fact that it is an act of the ability to read. And the suggestion we are considering is that this is an essential characterization of this transaction, meaning that in order to know it *as the kind of thing it is* one needs to see it as an instance of that ability.[46] In other words, since (PRE) is internally-related to (RE), one does not understand the kind of causal transaction it represents unless one grasps this connection. Thus, if I understand (RE), then I know that an explanation like (PRE) is necessarily available, since what it is for a person to read is for their utterance to be caused by the sight of particular written markings. Furthermore, if I understand that what is represented in (PRE) is an act of reading, then I know that that explanation is necessarily related to an explanation like (RE), since the availability of such explanations belongs together with the fact that the causal transaction was an instance of the agent's ability to read.

The problem with the Dancy-esque claim that *seeing the words on the page* is merely an 'enabling condition' is that it appears to make the connection between the two explanations too weak. We can bring this out by comparing this Dancy-esque enabling condition to others that are characteristic of reading. For instance, in order for me to read there must

[46] Once again, this does not mean that one cannot attend to this causal transaction without knowing that it is an act of the ability to read: a person could note a causal connection between certain markings and the noises a person made, study its physical and physiological basis, etc. But insofar as they did not know that the transaction was an act of the agent's ability to read, they would not know the causal transaction for what it was.

be sufficient light for me to make out the words on the page, which would mean that an explanation-like description that made this dependence clear,

EN: *S* said "…" because there was sufficient lighting in the room,

should be closely analagous to our (PRE).[47] But whereas (EN) does simply describe a *condition* that is *required for* the act of reading, (PRE) describes the causal transaction that *is* that very act. Treating (PRE) as analogous to (EN) obscures this identity. This shows why it is important to see the act of reading as a causal transaction between the signs and the reader. For if one thinks of the act as consisting simply of the reader's utterance, one will think that both (EN), (RE), and (PRE) work in the same way: the act of reading is described in the *explanandum*, and a cause of that act is described in the *explanans*. But the claim that acts of reading are causal transactions is supposed to bring out the fact that it is reading explanations *as a whole* that represent acts of reading, i.e. *explanans* and *explanandum* together. This is why we appended the subscript to the explanation: to mark that the whole was describing an act of a particular sort. When we describe an utterance as an act of reading, we *already* purport to provide some explanation for it.[48] This brings out the fact that (RE), (EN), and (PRE) all function slightly differently.

The objections to Dancy's actual account that we saw in §2.3 hinge on this same point. As we saw, both reductivists (Setiya, Davis) and non-reductivists (Marcus) complain that Dancy's treatment of psychological states as enabling conditions does not shed any light on the necessary connection between psychological and reason-giving forms of explanation. The appeal to abilities suggests a solution, at least in the present case: there is a necessary

[47] This is actually an interesting case, since the enabling condition includes essential reference to reading, marked by the idea of 'sufficient lighting'. To see this, one must simply ask: sufficient *for what*? To make out the letters on the page well enough to read them. Sufficient light to make out that there is *something* written on the page is not enough, since in order to read one needs to be able to make out what the words say. (An analagous case might be saying that it was an enabling condition of reading that the words be written clearly. What do we mean by clearly? Clear enough that one can discern what each of them says—or, in other words, clear enough to read.)
It is thus a mark of abilities that our grasp of some of their enabling conditions involves essential reference to the ability.

[48] [36, 192] makes an analogous point about describing an act as the keeping of a promise).

connection between explanations like (RE) and (PRE) because they both describe the same act of the ability: a causal transaction between written marks and the speaker who reads them. (RE) describes this transaction in a way that captures what is essential to it, i.e. it represents it *as* an act of reading. By itself, (PRE) obscures this identity—but fully understood, (PRE) is seen to be essentially related to (RE), because they describe the same act.

Deviant and Defective Cases

An appeal to the ability to *read* is also essential to responding to the first Dancy-esque claim, (a), which hinges on the possibility of defective cases of *reading*. Earlier we saw a related problem, which was that the causal explanations of defective cases explained a person's acts of reading by citing something that was not a reason$_R$ for that act. Thus, in a defective case, we might be confronted with a pair of explanations such as:

RE$_1$: *S* said "kæt" because [as she thought] "C-A-T" was written on the page

PRE$_1$: *S* said "kæt" because she saw "C-U-T" written on the page

Transposing Dancy's terms, we could say that (RE$_1$) provides a reading explanation by citing a consideration that was not in fact true, but that the reader took to be true, and add that this marks reading explanations as non-factive and thus fundamentally different from genuine causal explanations. (PRE$_1$) would then describe a condition that enabled our reader's mistake (i.e. she wouldn't have made the mistake had she not been responding to that sign).

Defective cases analogous to these also provided the motivation for certain kinds of reductive account. If we reject Dancy's claim that (RE$_1$) is a genuine explanation, we have to replace it with the more familiar psychological form of explanation:

RE$'_1$: *S* said "kæt" because she believed that "C-A-T" was written on the page

The reductivist motivated his account by claiming that we had to treat paradigm and defective cases on a par with each other, and find resources that were equally present in both on which to base our account of the relevant kind of explanation. Since explanation by psychological state applies to both paradigm and defective cases, the reductivist would claim that explanations like RE'_1 reveal the true basis for reading explanations: explanatory appeal to a psychological state that caused the utterance.

However, once we introduce the idea that reading explanations appeal to the ability to *read*, as it has been described here, the grounds for the disparity between the paradigm and defective cases becomes clear. The general idea of a disposition or ability allows for the possibility of both paradigm and defective actualizations: if I say that a substance is fragile, I mean that it will break if struck, unless there are mitigating circumstances that interfere with the actualization of the disposition. Our understanding of the disposition is primarily in terms of the paradigmatic case, instances of which are intelligible in their own right: if I know that a vase is fragile, we are not surprised if it breaks when it is struck. In contrast, defective cases need further explanation: if I strike the vase hard enough, and it doesn't break, it makes sense to ask why.

The ability to *read* is the ability to render signs into utterances according to the principles of the relevant language. This involves both *recognizing the signs*, and *responding to them correctly*. Here too paradigmatic exercises are intelligible in their own right: if we know that S is *reading*, and correctly said "..." because she saw the sign "_____", then *qua* act of reading her utterance is comprehended. In contrast, if someone who knows how to read makes a mistake, it makes sense to ask why. Possible reasons will vary: the person might have been tired and mispronounced her words; or perhaps the font was unclear, such that she misread the sign and took it for another; or it might even be that her ability itself is defective insofar as she habitually misreads particular words. But it is the character of the ability that explains the general contours of its defective instances: if the paradigm act involves *recognizing* and *correctly responding* to signs, then deviant cases will be those in which a person *misrecognizes* and/or *responds incorrectly.* The psychological attribution of (RE_1') marks some such defect, while showing that the act still needs to be understood in terms of the normativity of the practice.[49]

Thus what gives reading explanation the distinctive character noted in our two Dancy-esque points is, in a sense, the fact that they are a distinctive kind of explanation. But we can now see that what is distinctive about this explanation is not that it is non-causal. It is rather that it describes a distinctive kind of causal transaction—one that must be understood as the act of a distinctive kind of ability.

5.4 Normativity and Explanation

5.4.1 Other Abilities and Practices

We have seen that the intelligibility of the acts described in reading explanations—and with it, the intelligibility of the causal-explanatory relations that characterize them *qua* the kinds of acts they are—depends directly on the fact that they are acts of *reading.* The appeal to the ability to read that characterizes our reading explanations is essential insofar as understanding the nature of the act, its cause, and the causal-explanatory relationship between them all hinge on seeing the whole as an act of reading. This point generalizes to other practices as well. Consider the following three explanations:

S1: *S* turned left because she the sign pointed that way,

[49] Note that it will often be enough that we see that one sign could easily be mistaken for another—given that background understanding, we might not need any special explanation of *why* the reader mistook them in this case (though perhaps we would if e.g. the act of reading was particularly important, or if she seemed to be giving it her full and undivided attention, spelling out each word carefully, etc.) This brings out the fact that our understanding of the source of error often depends on our familiarity with *reading.* For instance, let's return again to the person who said "CAT" in response to a sign saying "CUT". This is a simple enough mistake to make: the sign "CAT" looks a lot like the sign "CUT", and given fairly ordinary circumstances it would be easy to mistake one for the other. A mistake of this sort needn't be particularly perplexing, and that is because of this similarity between the signs; but our understanding of the relevant dimension of similarity might *depend on* our familiarity with reading. After all, the similarities and differences between signs that strike a reader needn't be the same as those that strike a non-reader. And if we were told that a person "misread" e.g. "CUT" as "ZEBRA", we would be more deeply perplexed.

RU1: *S* wrote '2, 4, 6, 8' because the instructions said to follow the rule '+2',

P1: *S* returned to the house because she promised T that she would.

In each case, the explanation explicitly locates the act it describes within some kind of practice, i.e. following signs, applying rules, keeping promises. Seen from outside the practice, the normative, and thus explanatory, relationship represented in our explanations is obscured. With reading, this amounted to the claim that we would not be able to see exactly what the relationship was between the markings and the utterance made by the agent. A similar point could be made in the other three cases.[50] Someone unfamiliar with any practice of sign-posting would be puzzled about what relationship there might be between e.g. the tapered end of the plank of wood and the direction a person moved in. And someone unfamiliar with any practice of promise-keeping might be puzzled about what relationship there might be between the fact that a person said they were going to do something, and their subsequently doing it.[51] In these and other cases, we can imagine someone unfamiliar with the practice exclaiming: well I don't see why *that's* a reason to do that! They cannot see the rationality of the act, and because of that they do not recognize the explanatory force of what is cited to explain it.

Indeed, from such a perspective certain fundamental descriptions of both the act and its cause are unavailable, meaning that the relevant explanation simply could not be understood. Take the case of sign-following: the outsider may not understand the claim that the plank of wood is *pointing* in a particular direction. And with the act of fidelity, what the outsider fails to understand is that the person's utterance was a *promise*, and that doing what she

[50] We can imagine a version of each explanation that did not make this point explicit:

S1': *S* turned left because the wood tapered in that direction.

RU1': *S* wrote '2, 4, 6, 8' because the phrase '+2' was written on the sheet,

[**P1':**] *S* returned to the house because she had said to T that she would do this.

[51] To make this last case more vivid, imagine that everything else speaks against the person going back to the house—how can merely *having said that you would* provide a reason in such circumstances?

said was keeping that promise. Just as someone outside the practice of reading could not recognize written marks as the kind of thing that provides 'reasons$_R$' for an utterance (i.e. as written words), here our outsiders can't recognize how a plank of wood, or something said in the past, can be the kind of thing that provides 'reasons' for behaving in a particular way. Furthermore, each case involves a certain kind of causal transaction whose nature depends on the practice: the sign, the expression of the rule, and the making of the promise are all essentially involved in the causality of each respective act by a kind of transaction that characterizes such acts. And so in each case, understanding the nexus of act, reason, and the causal-explanatory relation between them depends on seeing the whole as belonging to the relevant practice.[52]

This is the basis for the claim that what we have in view here is a distinctive form (or, perhaps, distinctive forms) of explanation. In each case, the explanatory power of our descriptions depends on appeal to an ability that marks the acts described as belonging to a particular practice, and the causal-explanatory relations represented by our descriptions are characteristic of the act *qua* act of that practice. This means that it is impossible to understand these explanations (or indeed to make sense of the act at all) without seeing the whole thing in terms of the normative relations that characterize the practice. It is the role played by these various forms of normativity that mark each explanation out as distinctive.[53]

[52] This isn't to deny that there are also fundamental differences between the cases. For instance, (S1) represents S as *responding* to the sign, whereas (P1) does not represent S as (in this sense at least) responding to what she said, though her act does depend in some other way on her earlier utterance.

[53] This point can be made at various levels of abstraction. In one sense, any written language might be thought of as providing the basis for a distinct class of reading explanations, i.e. those that appeal to the particular normative relations characteristic of that language. At a higher level of abstraction, we can talk of reading explanations in general, where we mean to encompass all explanations that are indexed to some language in this way. Finally, the comments above suggest a still broader category of practice-based explanation, i.e. explanations that appeal to some kind of practice as the source of a form of normativity essential to what is described in the explanation. Based on what we've said so far, this would be a hugely various category, encompassing everything from sign-following to promise-keeping, with much else in between. I have merely sketched this category here, and nothing I have said is meant to rule out the possibility that there are further differences to be described between various kinds of practice, and the abilities involved in them. For instance, it might seem odd to call 'promise-keeping' an ability, rather than, say, a 'virtue'. This difference is reflected in what it takes to genuinely have the ability, i.e. not simply the capacity to say I am going to do something and then do it, but also qualities such as the moral authority to make promises, and the trust-worthiness that comes with that.

Focusing on simplified cases, we can see that the normativity involved in *sign-following* (i.e. following the direction of a tapered piece of wood) is distinct from the normativity involved in *reading* (i.e. making utterances on the basis of written signs). What constitutes a 'reason' in each case (a piece of wood, a written mark), and what it is a reason to do (change direction, make an utterance) differs between the practices.[54] But we can also see a certain sort of commonality between them, insofar as the intelligibility of each kind of act, and the explanatory dimension of our talk about those acts, depends on seeing them as actualizations of abilities belonging to bearers of each of the respective practices. Sign-following explanations like

S1: *S* turned left because she the sign pointed that way

and reading explanations like

R1: *S* said "kæt" because "C-A-T" is written on the page

belong together as examples of practice-based explanations. What marks these explanations as distinct is that they describe a kind of reason-based responsiveness that belongs to bearers of the relevant practice, and in doing so represent causal-explanatory relations that need to be understood in terms of the normativity that belongs to that practice.

[54] Though of course they can also overlap: the sign might name that to which it points.

CHAPTER 6

REDUCTIVE ACCOUNTS

If Wittgenstein's discussion of *reading* is to have relevance to contemporary reductive accounts, then such accounts should have the following three features:

1. They will be framed in terms of (MQ) and (EQ), where those questions are understood to ask *in virtue of what* some piece of behaviour counts (and is known) as an instance of acting for a reason.

2. They will answer (MQ) by appealing to some occurent or actual process of causation, which will have some feature in virtue of which the behaviour it causes counts as an instance of acting for a reason.

3. They will answer (EQ) by claiming that this causal process involves some self-identifying component, which is immediately known by the agent and explains how she comes to have first-personal knowledge of her reasons for acting.

If (1)-(3) hold true of an account, then we would further expect that it will ultimately be unable to provide a response to (MQ) or (EQ) that meets its own criteria: anything it can specify that purports to be that in virtue of which something counts (and is known) as an instance of acting for a reason could be present in the wrong kind of case, and so cannot provide what the account requires.

The main body of this chapter is taken up by a discussion of the views of a particular causal-psychological theorist, Kieran Setiya. I have chosen to focus on Setiya because he is particularly clear about the motivations behind his approach, as well as the form and commitments of his final account (though—as we shall see—less clear about the details of his proposals). Given this clarity about the nature of reductivism, I take it that Setiya's account will share certain key features with *any* reductive approach to this topic, and as such I would expect similar difficulties to attend other accounts of this form. To bring this out, I will also make occasional mention of other causal-psychological accounts, showing that they are committed to analogous claims and face the same problems as a result.

Setiya proposes a causal-psychological account that is reductive insofar as it proposes to explain *what it is to act for a reason* by providing an account of rational explanation in terms that are independent from the concept of *acting for a reason.* As we saw in §2, the key idea behind causal-psychological accounts is that we can treat rational explanations as a kind of causal explanation that depends on a causal relation between the agent's psychological states and their actions. The reductive character of these accounts stems from the claim that the elements of this causation—the psychological states, the behaviour they cause, and the causal relation between them—can all be characterized in a way that adequately articulates their explanatory role independently of the concept of *acting for a reason.* This stance commits such accounts to a rejection of the claim that an appeal to some form of normativity is essential to a characterization of rational explanations and the causal interactions they represent, and it is this commitment that leaves them vulnerable to a version of the problem that faced Wittgenstein's interlocutor.

My ultimate goal in this chapter is then to show that this commitment undermines causal-psychological accounts because it leaves them unable to provide an adequate characterization of the causal interactions they claim that rational explanations represent. This problem becomes apparent through an investigation into the difficulties raised by the 'problem of causal deviance' (§6.2 and §6.3), which in turn raises broader concerns about the implausibility of picture of 'acting for a reason' that emerges from these accounts (§6.4). By tracing both problems back to the reductive form taken by causal-psychologism, the discussion in this chapter prepares the ground for a closer consideration of the non-reductive normativist account in chapter 7.

6.1 Causal-Psychological Accounts

6.1.1 *The Basic Commitments of Causal-Psychological Accounts*

The basic form of causal-psychological accounts, and the motivations behind them, were both presented in §2.3.2. Broadly speaking, such accounts propose to treat rational explanations as kind of causal explanation, and further suggest that we can provide a philosophically-adequate account of the kind of causal interaction that is *acting for a reason* in reductive terms. The first claim entails that end-specifying explanations of the form

S is doing A because she is doing B,

along with grounds-giving explanations of the form

S is doing A because p,

must both be understood as representing causal relations that are more fully characterized in the psychological variants of these explanations, i.e.

S is doing A because she desires/intends to do B
S is doing A because she believes that p.

According to the causal-psychologist, these latter explanations represent a causal relation between the psychological state specified in the *explanans* and the action specified in the *explanandum.* The key claim of their account is that *what it is to act for a reason* is for one's action to be caused in this way by some such psychological state.

The second claim is that the character of the causal interaction represented by these psychological explanations can be adequately specified in terms that are independent from any conception of *acting for a reason*, and thus from any specifically normative relations that belong with such a conception. This entails that the *explanans*, *explanandum*, and the explanatory-relation between them—respectively, a psychological state, an action, and a causal relation—can all be specified in these terms.[1]

[1] Although it would be in principle possible to offer a non-reductive causal-psychological account, I do

The Motivation for Providing a Reductive Account

Reductive accounts of this general form can be situated within a broader trend within philosophy of action, developing in part from the problems posed by Davidson that we discussed in §2. Much of the early response to Davidson's work, along with other strands in the literature, was concerned to provide a reductive account of concepts like *action* and *agency.* A reductive theory of *what it is to act* would explain what it was for a particular behaviour to be an *action* performed by a particular agent in terms that were independent of concepts like *action* and *agency.* In some cases, this motivation was taken to push in a strongly reductive direction, the ultimate aim being to show that concepts describing action could be reduced to concepts from the physical sciences. But not all reductive accounts were quite so ambitious in their scope. Many philosophers who were sceptical about the possibility of providing a reductive account of psychological states involving intensional content—and who were as such prepared to take such states as primitive—nevertheless hoped that other psychological concepts could be explained purely in terms of these states and concepts available independently of them. A causal-psychological theory of *what it is to act* would explain what it is for a particular behaviour to be an *action* performed by a particular agent in terms of, for example, causal relations that held between certain psychological states of that agent and his movements.

Reductive accounts of *what it is to act for a reason* have obvious affinities with reductive accounts of *what it is to act.* But there should, *prima facie*, be space for someone to defend the possibility of the former kind of account, without committing themselves to possibility of the latter. In other words, one could accept that the concepts used to characterize *action* and *agency* are primitive, and not subject to reductive accounts—or at the very least, remain neutral on that question—while nevertheless maintaining that one could provide a reductive account for the more specific notion, *acting for a reason.* One of the authors we will focus

not know of any that are explicitly presented in these terms. I will return later to the question of whether normativism or my own account could aptly be called 'non-reductive causal-psychologism'.

on—Kieran Setiya—explicitly presents himself as adopting this position, arguing that *agency* and *action* are concepts that find application throughout the natural world, far beyond cases where there are agents who are acting for a reason:

> *[D]oing something* is a completely general topic in the metaphysics of agency whose generality is obscured by the restriction of 'agency' to rational agents. Call it 'agency' or not, there is such a thing as the exercise of a power or capacity by an object, inanimate or otherwise, about which we can ask: can this be explained in other terms? Someone who answers no, and therefore helps himself to the idea of an agent's doing ϕ, may nonetheless insist on a reductive account of what it is to ϕ *on the ground that p*. This would be a causal-psychological theory of acting for reasons without a causal theory of action.... [33, 137]

Given this background, accounts that deal specifically with the topic of *acting for a reason* (rather than the apparently broader topic of *intentional action*) tend to focus on providing a reductive account of the causality that they claim is central to acting for a reason.[2] Their key claim is that the kind of causality in question is not unique to instances *acting for a reason*, and can therefore be characterized in independent terms.

Scientism

There are two broad motivations for providing such an account. The first stems from a commitment to a form of scientism. The claim here is that, if rational explanations are indeed a form of causal explanation, a philosophically-adequate account of that causality must be articulated in terms of concepts from the physical sciences, since *any* causality is ultimately explainable in these terms. Since the concepts of rational- or folk-psychology,

[2] This is presumably because they take it that the other terms in their account—a psychological state and an action—are not specific to *acting for a reason*, and are thus available to a reductive account. Though I think there are ultimately grounds to contest this claim, I will follow such accounts in focusing on causality, since the general problems I am concerned with can be demonstrated through this part of the account.

which are central to the talk from which ‘rational explanations’ are abstracted, are not deployed in such sciences, it must in principle be possible to provide an account of the relevant causality in independent terms. Specifically, since (on this picture) the various forms of normativity involved in the justification and evaluation of action are generally articulated in terms of concepts that are not deployed in the physical sciences, it cannot be the case that a characterization of the relevant form of causality involves essential appeal to these forms of normativity. For if it did, this would mean that this causality could not be articulated in terms of concepts available from physical sciences, contrary to the commitments of scientism.[3]

Metaphysics

The second motivation for providing a reductive account stems from a more fundamental commitment to the role philosophical accounts play in explaining ‘metaphysical truths’, and the form such explanations must take. Here the claim is that a philosophically-adequate account of a metaphysical truth must be reductive if it is to provide a non-trivial (i.e. non-circular) explanation of its topic. Although some metaphysical truths may be primitive, and thus unamenable to philosophical accounts of this form, truths about *acting for a reason* are not among them (say the causal-psychologists), so a reductive account must be available. This approach is compatible with the scientism described above, but it also fits with other attempts to explain *what it is to act for a reason* in terms of independently available concepts.

Since this motivation is more directly related to (MQ), and issues pertaining it, it will be worthwhile to characterize it in more detail. In §2, we saw how Setiya motivates his general approach by way of a metaphysical edict:

[3] There are stronger and weaker versions of this position. A strongly reductive account would maintain that we ought, in principle, be able to show how to reduce rational explanation to some form of physical explanation. However, we are primarily concerned with weakly reductive accounts, which treat rational explanations as a legitimate and autonomous form of explanation, while also claiming that they rely on a form of causality that must be explained in physicalist terms. This position avoids any broad commitment to physicalist reductionism, while still claiming that any form of causality must ultimately be fully-articulable in terms available from the physical sciences.

Setiya's Metaphysical Edict: If it is metaphysically necessary that p, the fact that p must be explained by the nature of things; it must follow from what they are.[33, 139]

If we accept this edict, then any necessary truths about *acting for a reason* must be explained by the nature of *what it is to act for a reason.*

Setiya's account focuses on two 'metaphysical truths', each of which pertains to one of our orienting philosophical questions (MQ) and (EQ). The first is (MT1), which we discussed briefly in the second chapter:

MT1: If A ϕs because$_R$ p, then A ϕs because she believes that p.[4]

According to Setiya, this conditional expresses a necessary truth about *acting for reasons*: whenever a person acts for the reason that p, that person must (at the very least) believe that p and act because of that belief.[5] This is an essential feature of our talk about actions and reasons, one that (we might think) expresses something fundamental about *what it is* to do this. Given Setiya's metaphysical edict, we should expect an adequate response to (MQ) to explain why such conditionals are necessarily true.[6]

[4] See [33, 134]. The subscript appended to 'because' indicates that we are dealing with an instance of reason-giving explanation.

[5] In the best case, a person knows that p, and acts on the basis of that knowledge.

[6] Indeed, in §4.3 I also relied on a version of Setiya's point. There I was concerned with attempts to provide an account of 'rational explanations' by way of an appeal to some teleological ability or disposition. My suggestion was that, whatever the prospects for such accounts, it will be a key mark of such dispositions that their acts are subject to a kind of ground-giving explanation. Thus, it is insufficient to focus on teleological explanations of the form,

S is doing A because she is doing B,

since, whatever else is true of the 'ability' or 'disposition' that such accounts appeal to, one thing that marks it off from other kinds of teleological ability or disposition is that its acts depend on what the agent knows of believes. This is why such explanations are intrinsically related to explanations of the form,

S is doing A because p,

where p specifies the agent's grounds for the action, and thus (usually) something she knows or believes to be the case.
A non-reductive account that appealed to a teleological disposition or ability should also be concerned with Setiya's (MT1), since it characterizes a distinctive mark of the acts of that disposition or ability.

Setiya's second 'metaphysical truth' derives from Anscombe's investigation in her book *Intention*:

> MT2: When someone is acting intentionally, there must be something he is doing intentionally, not merely trying to do, in the belief that he is doing it.[7]

The puzzle that motivates Setiya is why MT2 should be necessarily true of intentional action: "What is it about being done for reasons—or being susceptible to the question 'why?'—that requires the presence of belief?". Once again, it is the need to provide an explanation of this truth that provides a criteria for an adequate account:

> In effect, the puzzle about *Intention* is to explain the necessary connection between acting for reasons and having a true belief about what one is doing. It may seem odd to demand an explanation for a necessary truth. But the demand is quite familiar, even if this way of putting it is not. We are faced with two marks of intentional action: its connection with reasons, and its connection with belief. The question is why something that satisfies the first mark of intentional action must satisfy the second. A necessary truth cannot be mere happenstance, so there must be something in the nature of intentional action, something in *what it is* to be done for reasons—or to be susceptible to a certain sense of the question "why?"—that requires the presence of belief. To give a plausible account of this connection is a condition of adequacy on a philosophical theory of acting for reasons. . . . Why is it that, when an agent acts for reasons, there must be something she does in the *belief* that she is doing it? [32, 27-8]

(MT2) pertains directly to our second philosophical question (EQ), since it asks why it is that a person who acts for a reason (i.e. in a way that is subject to Anscombe's question 'why?') is as such in a position to answer that question. Being in such a position entails

[7] See [32, 26].

that the person have a belief about what she is doing, and her reasons for doing it – and if we can explain why this is necessarily so, we will shed light on how it is that the agent comes to have first-personal knowledge of her actions.

According to Setiya, then, the need for response to (MQ) and (EQ) is partly generated by need to explain certain metaphysical truths that are directly related to those questions. As we saw in §2, one of his main claims in defence of his reductive account is that only an account of that form can provide an adequate explanation of those truths.

Causality

Whatever their motivations, a key feature of causal-psychological accounts is that reason-giving explanations are a kind of causal-explanation, and the causality in question is one that can be found in other kinds of case. This last point is essential to the reductive form of these accounts, since if the kind of causality were uniquely involved in the acts represented by reason-giving explanations (as it is on non-reductive accounts), then it could not be specified in reductive terms.

This approach might seem to put some pressure on causal psychologists to provide an account of the relevant kind of causality. But in fact, contemporary accounts typically have very little to say about the subject. This marks a significant difference earlier philosophers such as Davidson, who had a robust general account of causation that he hoped to apply in his account.

Focusing on our primary example, Kieran Setiya claims that his general reductive approach allows him to remain neutral on the form that an account of the relevant causation must take. This means that he does not provide a positive characterization of the kind of causation appealed to by his account, relying instead on an indirect specification by stating that it is a kind of causation that is common to both intentional action and deviant cases:

> If A is doing ϕ on the ground that p, A is dong ϕ because she believes that p, in a sense of 'because' that applies to deviant causation. [33, 135]

Deviant cases are to be understood as cases where some non-intentional behaviour is caused by a mental state, in such a way that it would be a mistake to describe the content of that state a providing the agent's reason for that behaviour. Davidson's climber-case is the classic reference point here: the climber's belief that he can get out of danger by letting go of his companion causes him to release his grip on the rope. In his book Setiya also describes cases in which a belief can be seen to *motivate* a person's actions even though they are not aware that it is playing this role, or even perhaps aware that they hold the relevant belief. The content of such a belief cannot be the person's reason for their acts, but since the belief motivates those acts Setiya understands it to be (part of) their cause. Setiya's thought here is that he can specify the kind of causation in question by saying that it is common to these cases and cases of intentional action: the kind of causal relation between belief and action is the same. Setiya treats this indirect specification of the causality involved as sufficient for his purposes in giving an account of acting for a reason, providing grounds for deferring the task of providing a positive characterization of the causation in question. [8]

A different approach can be seen in the causal-psychological account provided by Wayne Davis, which partly rests on the scientism described above. In providing his account, Davis states that "We know that there is a characteristic way in which intentional actions result from desire, even though there is a great deal we do not know about it" [18, 70]. This knowledge is reflected in our recognition that certain cases can be taken as paradigmatic instances of action based on desire, whereas other cannot. In light of this, Davis takes himself to be in a position to use placeholders to specify 'the right way' for desire and belief to cause action, using the terms 'f-result' and 's-result' to include all and only actions caused in the appropriate way.[9] His key move is to assume that the resources for filling out these

[8] An analogy would be to provide an account of reading where one specified the kind of causality involved as common to any case in which people responded to the sight of written signs by making noises. One simply states that the causality in question is common to all these cases, and then inquires what further distinguishes cases of reading from the others.

[9] See [18, 70]

place-holders will be provided by a empirical research into the physiology of action.[10] In other words, he delegates the task of providing an account of the relevant form of causality to the empirical investigations of science. What's more, since these placeholders appear in Davis' final account of what it is to act for a reason, he seems to be committed to the claim that the account itself must remain incomplete until we know how to fill them out: his account tells us that to act for a reason is to be caused to act by one's psychological states in the relevant way, but it leaves characterization of the relevant kind of causal relations to the physical sciences. This means that there is an important sense in which we are not yet in a position to fully articulate what it is to act for a reason.

As we shall see, this reticence about the details of the kind of causality involved in reason-giving explanations makes it hard to provide a thorough critique of their account. Whereas Anscombe could show that Davidson's account explicitly attempted to the 'grammar' of a particular kind of causation to cases of rational explanation, and failed because nothing that fit that 'grammar' could provide what he thought he needed, the contemporary accounts we are considering here do not give us as much to work with. However, as we shall see, a version of Anscombe's objection can still be applied to them.

6.1.2 Setiya's Account

With this background in place, its time to look at the details of Setiya's account. Let's begin by looking at Setiya's proposed explanations of (**MT1**) and (**MT2**). Here is (**MT1**) again:

MT1: If A is doing ϕ because$_R$ p, then A is doing ϕ because she believes that p.

[10] Davis does not explicitly state that this gap will be filled by future empirical research. Thus it is possible that he means this to be a placeholder for a strictly philosophical solution. But given his remarks elsewhere in the paper—see, for instance, his claim that we will one day be able to "deepen our understanding of the process of in virtue of which we act for reasons" by appeal to facts about "the neurophysiological correlates of beliefs and desires" [18, 85]—it seems reasonable to treat Davis' move here as representative of the approach available to certain kinds of empiricists. Besides, if Davis does in fact anticipate a philosophical solution, he is in the same position as Setiya, whose view is explored in the next section. For another philosopher adopting a similar approach to this topic, see Goldman [22], who argues eventually neurophysiological research can be expected to yield "detailed delineation of the causal process that is characteristic of intentional action".

Setiya's solution to the puzzle presented by (**MT1**) is quite straight-forward. His causal-psychological account maintains that one acts for the reason that p when one's action is caused by particular psychological states, among them the belief that p, with the relevant kind of causation fixed in the manner outlined above. Given such an account, the truth of Setiya's (**MT1**) is self-evident, since what it is to act for the reason that p *is* (among other things) for one's action to be caused by the belief that p. Hence it follows trivially that if one acts because$_R$ p, one acts because one believes that p. That's just what it is to act for a reason. The 'because' of the consequent denotes a causal relation that is found beyond cases of intentional action, and in particular in so-called 'deviant cases' like that of Davidson's climber.[11]

Note the general form of this solution. The idea is that we clarify the explanatory relations between reasons, actions, and psychological states by showing them to be instances of a more familiar class of causal explanation, and as such to involve a familiar and antecedently understood kind of causality. The topic of *acting for a reason* is made intelligible by showing how it can be made to fit this broader picture of explanation and causation. This is a strategy we saw in Davidson, and it is one that is common to all causal-psychological accounts.

Setiya's defence of his account—and the plausibility of causal-psychological accounts in general—hinges on two claims. The first, mentioned above, is that no other candidate account is in a position to explain the truth of (**MT1**). That claim clearly depends, in part, on the viability of primitivist and non-reductive accounts, and more importantly on whether we agree with Setiya about the need to explain 'metaphysical truths' like (MT1). The

[11] Another causal-psychological theorist whose work will feature in this chapter—Wayne Davis—motivates his account in a similar fashion. He asks why it should be that citing a person's reasons should explain their actions, and suggests that a causal-psychological account is particularly well-placed to answer this question (see [18]). His account maintains that to act for the reason that p is for one's action to be cause, in the right way, by a belief and a desire relating to p. This means that the reason for one's action is always that it is caused by the relevant beliefs and desires. If one combines this with the additional claim:

> (C) If A causes B, then B is explained by A.

we get an answer to Davis' question: our actions are explained by our beliefs because those beliefs cause our actions, and if an action is caused by something, it is also explained by it.

second, which is our focus in this chapter, is that a reductive causal-psychological account has the resources to solve the problem of causal deviance. This will be the source of the main difficulties facing reductive accounts.

Setiya's explanation of (**MT2**) is more involved. Here, as a reminder, is the necessary truth in need of explanation:

> MT2: When someone is acting intentionally, there must be something he is doing intentionally, not merely trying to do, in the belief that he is doing it.[12]

Setiya ties this truth to the language we use in describing intentional action: we characterize people as 'taking' such-and-such as their reason. So, for instance, in doing ϕ because p, I *take* p as my reason for action, where this implies both an epistemic dimension—I think of p as my reason to act—but also a practical one—as Setiya puts it, I am moved to act by this recognition. Thus it is an essential feature of *acting for a reason* that the agent know her reason for acting. This provides the seed for Setiya's ultimate explanation of (**MT2**): if, in ϕing because p, I must believe that I am ϕing because p, then *a fortiori* I must also believe that I am ϕing. So to explain (**MT2**), we need an account of what is involved in taking something as one's reason from which it follows that, when one ϕs because p, one believes that one is ϕing because p, and thus believes that one is ϕing.

After ironing out some nuances, Setiya's final proposal for such an account is as follows:

> To take p as one's reason for doing ϕ is to have the desire-like belief that one is *hereby* doing ϕ because of the belief that p.

This solution has various parts requiring explanation. First, the claim that 'one is doing ϕ because of the belief that p' is to be understood as a claim about the causation of one's action. The claim is that the action is 'motivated' by the belief that p, where this means non-deviantly caused by that belief, and it is this that one recognizes in having a belief

[12] See [32, 26].

about one's reasons: in ϕing because p, I recognize that I am motivated by my belief that p, and this recognition turns out to be recognition of a non-deviant causal relationship between that belief and my action. (We see here that a solution to the problem of causal deviance remains central to Setiya's account.)

But there is additional complexity to Setiya's solution. Setiya is troubled by the thought that it is possible for someone to be motivated by a particular belief or desire, and know that they are so inclined, without consciously endorsing that motivation or 'taking' the relevant content as their reason. Cases such as the one involving Freud's inkstand, described below, are taken to be examples of this kind of scenario.[13] This means that more is required than that the agent recognize a particular belief as the non-deviant cause of their action: they must also act 'take' the content of that belief as their reason to act:

> If the content of taking-as-my-reason is to depict me as acting for a reason, not just as being motivated by a belief, it must depict me as being motivated by the way I *take* the consideration that p. In other words, the attitude of taking p as my reason to act must present *itself* as part of what motivates my action. The content of taking-as-one's-reason is thus *self-referential*: in acting because p, I take p to be a consideration belief in which motivates me to ϕ *because I so take it*. This attitude *does* depict me as acting for a reason, since it depicts me as being motivated partly by itself, namely by the very fact that I take p as my reason. [32, 45]

The 'hereby' in Setiya's formulation is intended to capture this self-referential element: I both recognize—and, in doing so, make it the case that—my belief that p is the non-deviant cause of my action. The idea at the heart of Setiya's account is that, "in acting for reasons,

[13] This is why it will be important that Setiya does *not* treat the Freud case he discusses an example of deviant causation. Freud is 'motivated' by his belief, where this means that the belief is the non-deviant cause of his action. As such, there is some common element between a Freudian case, and a case of intentional action proper.

we have beliefs about the psychological explanation of our actions" [32, 47]. But our beliefs about these explanations are of a special sort: they are such as to make themselves true.

Once again, Setiya's proposed account does indeed have the implications he needs: if, in ϕing for the reason that p, I believe myself to be ϕing because I believe that p, then it follows trivially that I believe that I am ϕing. What's more, unlike his initial proposal—'to take p as one's reason for doing ϕ is to have the desire-like belief that one is doing ϕ **for the reason** that p'—his ultimate account does not involve appeal to the notion of taking something as one's reason, which means that as long as the relevant sense of 'doing ϕ because of the belief that p' can be made out without relying on this concept, his proposed explanation will also meet non-circularity criterion on reductive accounts.

It is important that we note the overall form of Setiya's solution. It works by assuming that we can find a solution to the problem of causal deviance (as Setiya understands it), which would give us an account of what it is for a psychological state to *motivate* an action. Given this notion of *motivation*, we can then capture the additional dimension of self-consciousness or first-personal knowledge by characterizing a special state as playing a motivational role when we have first-personal knowledge of our reasons. This special state is tacked onto the antecedently available idea of 'motivation' to provide an account that is supposed to capture the distinctive kind of motivation and understanding characteristic of self-conscious agents. This general strategy is again one that is characteristic of causal-psychological accounts: insofar as they pay attention to the self-conscious character of intentional action, they hope to be able to capture it by simply adding an additional element to their account of motivation.

Here everything hinges on whether this special state can indeed capture all that is involved in our self-consciousness in action. In particular, the content of the self-referential belief that Setiya identifies must be true if and only if the agent is indeed taking p as his reason to ϕ. As we'll see, this makes Setiya's need for a solution to the problem of causal deviance two-fold: he needs it to explain the sense in which, when I ϕ because p, I ϕ because of my belief that p, entailing the truth of his (**MT1**). But also—as an extension of this—he needs it to characterize the content of the self-referential belief by which I both recognize and make it the case that I ϕ because p. What I recognize and institute here cannot be an instance of deviant causation: my belief has to cause my action in the right way. So whatever the content of my self-referential belief, it must be such as to distinguish cases where I genuinely do take p as my reason, and act because of this, and cases that we would describe in other terms.

6.1.3 Wittgenstein and Setiya

The above outline shows that Setiya's account has precisely the features we identified in the interlocutor's account of *reading*. Setiya's overall account of acting for a reason can be summarised as follows:

> To take p as one's reason for doing ϕ is to have the desire-like belief that one is *hereby* doing ϕ because of the belief that p, where the desire-like character of the belief means that it causes the action by way of the same causality as is to be found in cases of causal deviance, and the self-referential character of the content of the belief guarantees that the agent is active in—and so has knowledge of—bringing it about that this belief causes her to act.

More succinctly, some piece of behaviour will count as an instance of acting for a reason in virtue of being caused by the special state that Setiya describes. As we shall see, Setiya remains vague on the details of this causality, but the proposal seems to imply that his state is both fully actual at the time of the action, since it causes the act in the same way that Davidson's climber's psychological states cause his behaviour, and (in some sense) occurent,[14] as implied by the 'hereby' in its content. Moreover, since this belief is self-referential, it has the same self-grounding character that Wittgenstein's interlocutor attributed to felt or conscious

[14] Setiya anticipates Anscombean objections to this claim by asserting that the state needn't be felt or conscious. But, as we should expect from the previous two chapters, this anticipation misses the force of her point.

experience. The agent knows what she is doing and why because her actions are caused by a belief that represents itself as the cause of what she does. Since the kind of causation involved is common to deviant cases, and the content of the relevant belief makes no reference to the idea of ‘acting for a reason’, Setiya has provided a response to (MQ) and (EQ) that fits his criteria for a reductive account, while also mirroring the key features of the interlocutor’s proposals in the sections on *reading*.

Given this, we should expect Setiya’s account to fail in an analagous way. In the previous chapter, we saw that the interlocutor’s proposals fell foul to a version of the problem of causal deviance: anything he proposed as that in virtue of which an act counted (and was known) as an act of reading could equally be present in the wrong kind of case, and so could not play the role he asked of it. In §6.2 and §6.3, I will argue that both stages of the account Setiya proposes are still vulnerable to the problem of causal deviance.

Of course, Setiya is quite clear that his account must ‘solve’ the problem of causal deviance. Indeed, the origin of the problem lies in the way the account begins from the idea that the relevant kind of causality can be found in cases that are not instances of acting for a reason, specifically (for Setiya) the kind of causally deviant case envisaged by Anscombe and Davidson. From the start, given the way he has identified the relevant notion of causality, he has put himself in a position where he must explain what differentiates instances of *acting for a reason* from the broader class of instances of this kind of causation. Cases of causal deviance make this problem acute, because they are cases that, according to the causal-psychological account, involve both the same kind of cause, and the same kind of causation, as genuine instances of *acting for a reason.* So in order to differentiate deviant and non-deviant cases, the account must be able to say what it is involved in psychological states causing behaviour *in the right way*—and, given their overall form, they must explain this in independent terms.[15]

[15] Note that non-reductive accounts are not faced with this problem. For if one maintains that there is a distinctive kind of causation involved in *acting for a reason*—for instance, by claiming an adequate characterization of this causality involved representing its results as done *in pursuit of an end* and *on the*

It is therefore no surprise that Setiya explicitly raises this as a problem for his own account:

> How is the problem of causal deviance to be solved? That is, how can we supplement the condition of acting because one believes that p, in a sense of 'because' that applies to deviant causation, in order to say what it is to act on the ground that p? In particular, what is the role of normative considerations or standards of practical rationality in reasons-explanations? [33, 152]

In the next two sections, we will see whether the solution he proposes is successful.

6.2 Motivation and Causal Deviance

Since Setiya claims that he can use his indirect method to specify the kind of causality involved in reason-giving explanations he remains vague on its details. What we do know is that it is common to cases of causal deviance where a psychological state causes a piece of behaviour without specifying the agent's reason for that behaviour, as with Davidson's climber case. This approach sets the terms in which Setiya approaches his account, since his most immediate task is to explain how genuine instances of acting for a reason differ from these deviant cases. For not just any case in which a psychological state causes us to act can count as an instance of acting for a reason.

Setiya (along with many other authors in the contemporary literature) understands the problem of causal deviance in terms of (a) whether the relevant action was 'purposive' or not, and (b) whether it was 'no accident'. Davidson's climber case can obviously be understood in just these terms: one thing that marks his behaviour is that it was not done *for the sake of* lightening his load, even though a desire for that end played a role in causing it. Had the climber acted for the sake of this end, his behaviour would have been purposive in

basis of belief—then deviant cases would already be ruled out: we would not need to ask "But was it done in pursuit of the end and in view of the thing believed?", since we would already be committed to this in representing it as an instance of this kind of causation.

the relevant sense, and would therefore have been 'no accident' insofar as it arose from a disposition to pursue this end.

This means that Setiya does not count cases in which behaviour was in some sense 'purposive', and the result of a stable disposition, as cases of causal deviance. For instance, Setiya has cause to discuss an example from Freud taken up in work by David Velleman. In Setiya's (and Velleman's) accounts, the example is described in the barest detail: Freud registers his dislike of an inkstand that does not match his new desk, and later unintentionally breaks it 'with peculiar and remarkable clumsiness'; Freud takes his act to show that he was moved to break the inkstand 'by the belief that breaking it will persuade his sister to buy him a new one, and the desire for a matching inkstand' [32, 33]. Setiya *does not* count this as an instance of deviant causation:

> [T]his causation is non-deviant, since the desire to break the inkstand guides the movements of Freud's arm in sweeping it to the floor. [32, 33][16]

[16] This is the only example of a psychological state 'guiding' something other than an intentional action in the first half of Setiya's book, and it is quite difficult to see from it what he means by 'guidance'. The idea of 'pursuit of an end' that Anscombe draws our attention to in her work in *Intention* (which is one of Setiya's points of reference here) involved the idea of ongoing action, such that a person would e.g. immediately do something else if their act failed to achieve their end, or else give up on the relevant intention. This is what we understood 'pursuit of an end' to be. 'Guidance', as it occurs in the Freud example, is not quite the same thing. After all, given that Freud's act in breaking the inkstand was supposed to be unintentional—he clumsily knocks the stand off the table—we can hardly imagine him following it up to make sure his end had been met (i.e. we don't imagine Freud checking the inkstand, and perhaps stamping on it if it wasn't damaged to his liking). The relevant desire doesn't 'guide' the immediate pursuit of the relevant end because there is no immediate pursuit!

We might imagine that Freud 'fails' to break the inkstand, and that the scenario repeats itself with some regularity—he treats it with general disregard, and is quite careless in his handling of it, without intending to behave in this way. Here perhaps we have something closer to sustained unconscious pursuit of an end: but that becomes apparent through the pattern of behaviour—it is not something we see in any particular case. This seems to be one feature that will set cases of unconscious motivation apart from more straight-forward instances of intentional action: the patterns of behaviour in which the motivation is discerned will be much broader in scope than those involved in completion of a particular intentional action, and their relation to the pursued end will perhaps be far less obvious (certainly to the agent, but also perhaps to someone taking an interpretative stance).

In this particular instance, if the desire 'guides' anything, it is only the clumsy movement of Freud's arm. And perhaps here we do get an analogue to the kind of 'guidance' Anscombe drew our attention to: Freud's interpretation, if true, does reveal his act of breaking the inkstand to be something other than an accident—i.e. to be, in an important and revealing sense, quite different from a case in which he was startled and dropped it, or was simply and straight-forwardly clumsy. It does so by revealing a purposiveness in the act that was not known by the agent, and that can be described in terms of a desire to break the inkstand.

Nevertheless, he does feel a pressure to show how his account can accommodate such cases, since unconscious motivation leaves open the possibility that a person might be motivated (i.e. caused in the right way) to act by a psychological state, without knowing her reasons for acting. Setiya thus takes it that he needs to explain *in virtue of what* paradigmatic cases of acting for a reason differ from these cases, i.e. why—in the paradigm case—we have knowledge of what motivates our actions.

Ultimately, I think both kinds of case should be understood as 'deviant' relative to our central topic, and I shall refer to them both in these terms in what follows.[17] This will mean that I see Setiya's proposed solution as having two parts (i.e. where he would see two solutions to two different problems). First, he introduces a notion of 'motivation' that is intended to characterize the general causal process involved when a psychological state explains an action by specifying a reason. Although, broadly speaking, it will be the same kind of causality involved here and in deviant cases, the idea is to introduce some further qualifications to explain the specific features of this causality as it applies to motivated action. Second, he introduces a particular psychological state that plays a role in this causal process in paradigm cases of acting for a reason. This involves showing how the content of the relevant psychological state can be specified in such a way as to rule out cases of unconscious motivation (i.e. cases where a person is motivated to act by a particular belief or desire without 'taking' the content of that state as their reason for acting).

Both stages of this proposed solution are vulnerable to versions of the problem of causal deviance. In this section, I will look at the first part of Setiya's solution, his treatment of

Furthermore, since this act took time, and involved movement, it can be broken into stages, each of which might be seen as being done for the sake of the unconsciously pursued end. So perhaps the desire can be seen as unifying the various moments of Freud's movement, and in this sense 'guiding' it to its conclusion. That, at any rate, seems to be Setiya's understanding of the case.

[17] A fully-fledged defence of this claim would involve arguing for the view that there is an essential difference between conscious and unconscious motivation, because there is a difference in the sense in which the action can be said to be done 'in pursuit of the end and on the grounds of the thing believed'. Such an argument is far beyond the scope of this dissertation. However, since Setiya agrees that his account needs to deal with both kind of case, classing them together as instances of 'deviant causation' is primarily an exegetical device, and not fundamental to my argument.

'motivation', which is most directly related to (MT1) and his response to (MQ). In §6.3, I turn to his account of the knowledge that is absent in cases of unconscious motivation, which is most directly related to (MT2) and (EQ).

6.2.1 Setiya's Solution: Guidance and Motivation

Since all of our authors focus on the explanation of intentional action, they are particularly concerned with action that is done for the sake of some finite end, and understand the problem of causal deviance in terms of such cases. As stated in the previous section, this means that there is a particular pressure to show why an action counts as 'purposive', in the sense of being done for the sake of some end, since this is understood to mark off genuine instances of causal explanation from deviant cases. This feature was lacking in the cases we considered in the previous chapter: an act of reading *might* be purposive, in the sense of being done for the sake of some further finite end, but we abstracted from that point in our discussion. For us, cases of deviant causation were simply ones in which a person reacted to some signs by making an utterance that could be taken to render those signs, even though they were not in fact reading them.

But despite this specific difference, the general analogy holds. As we saw in §6.1.1, Setiya relies on the idea that the causality involved in acting for a reason is whatever kind is found in deviant cases involving psychological states. Once the relevant kind of causation has been specified in this way, Setiya's remaining task is to "'build' acting for reasons from materials present in deviant causality, along with others that are not themselves composed or defined in terms of such action" [33, 135]. Since (a) intentional action is purposive, and (b) the behaviour described in deviant cases is not, this will involve showing how we can add to the general notion of causality that applies in both cases to capture these specific features of intentional action.

The first task in 'building' this notion is explaining what it is for a psychological state to genuinely 'guide' or 'motivate' an action, rather than merely cause it as it does in deviant cases.[18] Setiya makes several brief proposals for what such 'guidance' or'motivation' consists in. Though his remarks are, by his own admission, meant to be merely suggestive of the general form a solution might take, they make it possible to discern certain key characteristics that are required by the overall form of the causal-psychological account. As in the previous chapter, the problem with each proposed solution is that it does not in fact serve to differentiate genuine instances of acting for a reason from deviant cases.

Guidance

One proposal for 'building' *acting for reasons* from material present in cases of deviant causation is to introduce a notion of 'guidance' to capture the idea that the action was done in pursuit of an end. Setiya's suggestion here involves two parts. His first proposal is that we introduce an idea of 'guidance' to characterize the causal relation between the relevant psychological states and basic actions. Setiya characterizes this idea as follows:

> [W]hen an agent ϕs intentionally, he *wants* to ϕ, and this desire not only causes but continues to guide behaviour towards its object. [32, 32]

Guidance is a form of 'sustained causation'. The idea, then, is that the basic actions involved in the relevant behaviour are all 'guided' by the relevant desire, which specifies the end, by a process of 'sustained causation'. This feature was absent in Davidson's case, since the climber's release of the rope was a nervous reaction, and as such not 'guided' by his desire to relieve his burden and reach safety.

Non-basic action is then explained in terms of its relations to basic actions so characterized:

> If I do something with the end of doing ϕ, I must have a plan for doing ϕ by performing that action, and I count as doing ϕ intentionally just in case I do it in accordance with my plan. [32, 32]

[18] The second, discussed in §6.3, is specifying a specific psychological state to play the role of cause.

The notion of ‘guidance’ that Setiya introduces here raises some questions about the overall form of his account. If this notion of ‘guidance’ is fundamental to a characterization of the *kind* of causation involved in intentional action, then he *cannot* rely on a specification of this kind of causation as common to both paradigm and deviant cases. What makes the climber-case ‘deviant’ is precisely the absence of any purposiveness—and thus, presumably, ‘guidance’—in the climber’s behaviour. However, if it is not essential to a characterization of the kind of causation, we need a further story about why some instances of such causation count as cases of ‘guidance’, while others do not.

Setiya’s response here is of the same form as his earlier specification of the kind of causality involved in his account. That is, he supposes that ‘guidance’ can be understood as a matter of ‘sustained causation’, and that the idea of ‘sustained causation’ should be familiar from other kinds of case:

> Sustained causation of a process towards its a goal is not unique to intentional action: it is present in purposive behaviour that is not intentional. So although it is something of which we lack an adequate theory, there is no circularity in taking it for granted here. [32, 32]

It is a little unclear what Setiya has in mind here, but the bare idea of sustained causation—where that simply means a cause that is concurrent with what it causes, rather than preceding it—is not by itself sufficient for his purposes. For we can easily re-imagine the climber case in such a way that, rather than a single thought causing a startle response, we imagine the climber dwelling on a thought and gradually releasing the rope as a result. Suppose, for instance, that instead of being startled by the fact that the thought crossed his mind, the climber is instead upset by the fact that he is dwelling on it, and begins to tremble as a result, which causes him to gradually lose his grip on the rope. Here the trembling of his hand caused by the belief *is* his letting go of the rope, and is caused by the ongoing presence of the belief, in what seems to be a fine case of sustained causation. Moreover, even if we could make out sufficient conditions for sustained causation of a single act, this would still not be enough. What is needed is sustained causation of a series of actions all unified in pursuit of a specific end. The problem of characterizing these causal relations is not solved by shifting from a picture that involves causes preceding effects, to one in which they occur concurrently.[19]

Motivation

Elsewhere, Setiya tries to directly introduce an idea of 'motivation' to distinguish instances of acting for a reason from deviant cases. In characterizing this concept, Setiya says the following:

> The (non-deviant) motivation of desire is a matter of one's *dispositions*, and one can act for any reason that corresponds to a belief by which one is disposed to be moved. [32, 65]

For Setiya, to say that a psychological state 'motivates' an action is just to say that it is the non-deviant cause of that action. Here we learn that motivation is a matter of the dispositions by which one is liable to be moved, since insofar as a person is disposed to be caused to act in particular ways by the relevant psychological state(s), the causation in question counts as non-deviant:

> The causing of one mental state by another is non-deviant, in A, just in case A is *disposed* to make this transition, and this disposition was operative in its taking place. [32, 32]

This generic notion of a disposition articulated here depends on the idea that A has a disposition to move from C to E if this transition is 'part of the ordinary course of A's

[19] Another kind of case that could be described in terms of sustained causation is provided by Sebastian Rödl in *Self Consciousness*: we are asked to imagine someone who wishes to lose weight, and is made so anxious by this desire that they do lose weight, but not by acting on the desire. Losing weight is surely a sustained process, which suggests that the causation by the desire should also be understood as sustained. However, this is not a straight-forward counter-example, as Setiya could respond that losing weight is a non-basic action, and that the problem is that the agent does not achieve his goal by acting in accordance with his plan.

mental life' [32, 32], i.e. if A reliably transitions from C to E, and can be expected to do so in future.[20] This means that acts manifesting that disposition are 'no accident': they have their source in some enduring fact about the agent. It is this feature of dispositions that makes it plausible that they must play some role in distinguishing deviant from non-deviant cases: for one feature that characterizes many deviant cases is that the act in question is, in some sense, an accident.[21]

But once again, it seems as though any version of this proposal that Setiya can endorse has problems that are directly analagous to the problems with the concepts of 'guidance' and 'sustained causation' introduced above. For the bare notion of a reliable transition between states does not by itself seem sufficient to distinguish acting for a reason from deviant cases. After all, it doesn't change the status of Davidson's climber case to say that the climber reliably behaved in this way (though it would make one wonder why anyone partnered with him!). That is to say, we can easily re-imagine the case to be one in which the climber has a nervous disposition, and tended to be spooked by thoughts of this sort when climbing. Clearly the fact that this makes his behaviour an instance of a reliable transition from beliefs about climbing conditions to dangerous behaviour doesn't make this into a case in which the climber was 'motivated' by those beliefs.

Proportionality

In more recent work, Setiya has appealed to a notion of 'proportionality', partly in response to the kind of objection raised in the last paragraph. Building on work by Stephen Yablo, Neil McDonnell provides the following characterization of proportionality:

[20] Notice that this generic notion of a disposition could easily be extended to agent's other than human beings, including both animals and inanimate objects. An 'agent', in this sense, has a disposition just insofar as they are disposed to move from C to E in particular conditions. It is this broad applicability that suits this notion of a disposition to reductive accounts.

[21] So, for instance, Davidson's climber does not *mean* to let go of the rope; he did not do it with the end of e.g. killing his companion; and this is reflected in what he does afterwards. Some philosophers (e.g. Setiya) take this as *the* mark of deviant causation. Thus any case where action springs from a reliable disposition in the agent, no matter how bizarre, is classified as non-deviant. See e.g. [33], particularly his discussion of the case from Freud (pp.33-4) and the example from Nagel (pp.64-5), both discussed elsewhere in this chapter.

> The core idea is that a cause is proportional to its effect if and only if there is no more general specification of the event that would have sufficed for the effect (no more determinable specification), and there is no more precise specification of the cause (no more determinable specification) that is required. [25, 162][22]

Setiya then applies this notion to his particular notion of a disposition:

> Inspired by Yablo on proportionality, we might regard the transition from C to E as non-deviant if and only if it is the exercise of a disposition to go from C to E, and one does not have this disposition merely in virtue of being disposed to go from some more inclusive condition, C*, of which C is a determinate, to E. In that sense, or perhaps some strengthening of it, the specific content of the mental state is relevant to motivation. [34, 534]

This is then supposed to develop the idea of a 'reliable disposition' to a more specific notion covering the role that psychological states play in causing action. Since the dispositions involved in intentional action will involve transitions from one psychological state to another (e.g. from belief to desire, or from desire to intention), or from a psychological state to an action, the appeal to 'proportionality' is supposed to ensure that the *content* of the psychological states is relevant to this transition. The idea is that, for a psychological state to be the proportional cause of an action, the content of that state has to be essential to the motivation in question: if some more generic specification of the state (e.g. a 'nervous disposition') would serve equally well to characterize its causal role, then the specific state is not the proportional cause of the what follows it.

McDonnell proposes an absence of proportionality as the key characteristic of all cases of causal deviance, and suggests that accounts that rely on appeals to causal relations can

[22] A common example involves a pigeon that is trained to peck at red things: one could explain a particular peck by saying 'The pigeon pecked because of the red square'. However, the cause cited in this explanation is not proportional to the effect it is called on to explain, since the pigeon would have pecked at any red object. Thus the proportional cause of the pigeon's act was the redness of the square.

ensure that those relations are non-deviant by specifying that the cause be proportional to its effect. Here is his complete description of what is involved in Davidson's climber case:

> We can grant that the nervousness caused the rope to slip but ask: what caused the nervousness? In this instance, the thoughts of letting the rope slip may be the thoughts that caused the nervousness, but they are only the proportional cause if there is no more general specification of that event that would have caused the same state. What about such beliefs and wants are unnerving, we might wonder? It is surely nothing to do with ropes or weight or grip per se, which are the given content of the thought, but rather it is the fact that in this context that content belongs to a broader type of thought, of willfully harming someone else, and the consequences of such. It is that determinable type of thought, concerning harm and selfishness, and not the determinate sort, concerning letting a rope slip, that is unnerving. As such, it is these determinable thoughts that serve as the proportional cause of the nervousness, not the overly specific, determinate, thoughts regarding ropes and slippage. By invoking the proportionality constraint, we can insist that it was the thoughts of harming someone else at all that caused the slippage and, without the additional detail concerning ropes, etc., this thought will not appropriately match the outcome in the manner required to be a genuine counterexample. Once again, the initial cause has been overly specified to create the problem.

This is then meant to rule out the kind of objection I made above: that there are any number of dispositions that could led to a 'reliable transition' between the climber's belief and desire, and his letting go of the rope, without changing the fact that this was an instance of deviant causation. Proportionality purports to solve this problem by making the specific content of the climber's beliefs relevant to a characterization of the appropriate disposition: it is a disposition to act on this quite specific belief about the circumstances of his behaviour.

It is worth noting that the appeal to proportionality implicitly relies on a particular description of the case: what unnerved the climber was the thought of hurting someone, and his particular beliefs about ropes or weight or grip were irrelevant. That is certainly a fine way of expanding on Davidson's case, but it is not the only possibility: one could very easily describe a case in which the climber *was* reacting to more determinate features of his situation. For example, imagine that the climber had been involved in an accident on *this* rock-face before, with *this* partner, using *this* equipment. Suppose further that the belief that he could reach safety by loosening his grip on the rope unnerves him because of the echoes of this previous climb. It doesn't seem too difficult to expand on the case to a point where the thing that unnerves the climber, and causes him to loosen his grip, is a belief about acting in *this* way in *these* circumstances with *this* person on *this* occasion. Such a belief would be proportional to the action in question, but would nonetheless still count as a deviant cause.[23]

Normativity and Explanation

It is a key feature of all three of these proposals that they attempt to characterize what it is for a psychological state to properly cause an action without any appeal to normative relations between the content of that state and the agent's understanding of the action. This is because Setiya takes it to be an important commitment of reductive accounts that they be able to explain *what it is to act for a reason* in non-circular terms. Since our grasp of the relevant form of normativity (i.e. of what is a reason for what) essentially involves some idea of 'reasons', Setiya deliberately attempts to characterize the causal process involved in acting for a reason without any appeal to such normative relations.

In each case, this ultimately means that his proposed description of the relevant causal process fails to adequately distinguish cases in which a psychological state specifies a reason

[23] This shows a general problem with attempts to solve the problem of causal deviance by showing that the relevant behaviour was 'no accident'. For philosophers can always arrange things so that behaviour was 'no accident'—in the sense that it was caused by some reliable fact about the agent—and yet not intentional.

for an action from cases in which it causes it in some other way.[24] The appeal to 'proportionality' seems to me to show an inchoate awareness of this problem. For what that notion is supposed to do is precisely to sort cases into the appropriate groups. However, the appeal to a generic notion of 'proportionality' obscures how it does this in each particular case. For instance, our sense of what would be a proportional effect of the climber's belief if his action was intentional depends on our grasping what one might take that belief to give him reason to do.[25] Characterization of the action in these terms will depend essentially on representing it as an intentional action, and on representing more specific normative aspects of the situation to which the agent is responding. Given such an understanding, we can begin to identify and trace out relevant causal connections, imagine counter-factual situations, and perhaps work out how to apply ideas about 'proportionality' to the particular case. But our ability to do this depends on our already understanding that what we are describing an intentional action: once again, the relevant sense of proportionality depends on that notion.[26]

This is not to deny that some notion of 'proportionality' might turn out to be extensionally adequate in terms of sorting deviant from non-deviant cases. But its capacity to do this would come from its representing causes as proportional to effects *qua* intentional

[24] This might appear to be a result of preliminary character of Setiya's suggestion: perhaps with some clever work one could come up with some supplement that successfully sorted cases the right way. The arguments of subsequent sections, together with the discussion of Wittgenstein and Anscombe's work in the previous two chapters, should give on grounds to doubt this eventuality. At best such an account might turn out to be extensionally adequate – on which, see the further discussion of 'proportionality' below

[25] That is, unless we are to understand 'proportionality' a based entirely on empirical observations, as it might in our characterization of a disposition whose characterization relied entirely on observation of past patterns of behaviour, e.g. the disposition of pigeon described in a previous footnote. The problem with applying this to the climber case is that, without appeal to various normative relations that characterize the content of this belief, it is not clear *which* past patterns of behaviour are supposed to be relevant.

[26] Setiya rejects this claim. After introducing the notion of proportionality in relation to dispositions, he continues:

> Still, the content need not be relevant in the normative sense that the disposition exercised is even approximately rational. As I argue in the book, I can decide to drink coffee because I love Sophocles, or to put money in a pencil sharpener because I want a drink from a nearby vending machine – where this is the because of motivation or non-deviant causality – so long as I am appropriately disposed. Such dispositions are unusual, but their presence would ensure that the relevant transition is not . . . a fluke. [34, 534]

For discussion of the kind of case Setiya describes here, see §6.4.1

action under such-and-such a description. The further specification is essential to its role, and means that the relevant understanding of ‘proportionality’ can have no place in a reductive account: it depends fundamentally on an understanding of cause and effect in terms of intentional action and the specific normative relations involved in it, and so cannot be part of a reductive account of that concept without making the account circular.[27]

6.3 Setiya’s Self-Referential Belief

In §6.2, I tried to show a consistent problem in Setiya’s attempts to specify various features of the causal process involved when the content of a psychological state motivates an action. Given the cursory nature of Setiya’s proposals, these arguments might appear inconclusive. After all, the real substance of his causal-psychological account lies in his specification of the specific psychological state that causes our actions, and it will ultimately be in virtue of being caused by this state that an action counts (and is known) as an instance of acting for a reason.

Setiya’s specification of this state is part of his final proposal for what it is to act for a reason:

> To take p as one’s reason for doing ϕ is to have the desire-like belief that one is *hereby* doing ϕ because of the belief that p.

It is part of his claim that the desire-like self-referential belief (henceforth ’SR’), whose form is:

> I am hereby ϕing because of my belief that p,

[27] Another way of putting this point is to say that, without characterizing the ‘proportionality’ between cause and effect in terms belonging to the description of intentional action, any account that relied on proportionality would leave its grounds mysterious. The account would perhaps be extensionally adequate, in that it would sort deviant and non-deviant cases correctly, but it would leave us without any explanation of why the cases sorted in this way. ‘Proportionality’ amounts to a demand for an extensionally adequate account, without an attempt to explain

and the further belief it specifies (i.e. the belief that *p*). Together these two beliefs must 'motivate'—and thus be the non-deviant cause of—the agent's action. The 'hereby' in the content of SR serves to both institute and recognize this motivational connection, and further to explain the agent's knowledge of what she is doing and why. As Setiya puts it,

> To do something because *p* is to be motivated by an explanation of one's action that appeals to belief. [32, 51]

Behaviour will then count (and be known) as an instance of acting for a reason in virtue of being partially caused by SR.[28]

Given the concerns raised in §6.2, the exact role that SR is supposed to play in the causation of action is somewhat unclear. For instance, Setiya takes it that he can avoid Anscombe's objection by rejecting the idea that we must appeal to 'felt or conscious experience' in understanding the causal role played by SR. But this leaves the role played by the 'hereby' in its content somewhat mysterious. The obvious analogies to other cases (e.g. performatives such as 'I hereby declare you married') suggests the state must be occurent or actual at the time of the act to play its causal role. Setiya's suggestion is that we simply understand beliefs like SR on the model of other beliefs that give grounds for our action: clearly I can act on the grounds that *p*, without the thought that *p* 'running through my head' at the time I act. But other beliefs lack the self-referential feature that is integral to SR's causal role. It would seem that I must, in some sense, 'think' the 'hereby' in order for it to do its work, and that that 'act' must be occurent with the action it explains. But whatever our general story about the role of occurent belief's relation to consciousness and its role in action, it is unclear how any occurent belief could have the pseudo-performative role that Setiya describes without being in some sense conscious.[29]

[28] Thus Setiya's explanation of (**MT2**) depends on an appeal to, and then builds on, his idea of 'motivation', where that just means non-deviant causation by a psychological state. This should make it clear that Setiya does need an account of 'motivation'—specification of SR, combined with the causality involved in deviant cases, will not suffice for his purposes.

[29] One way of seeing this is by asking whether the 'hereby' can ultimately be understood as referring to

Since Setiya does not develop his account in much detail, it is difficult to grasp exactly what role he envisages SR as playing. But even supposing we could make out such a role, his formulation of SR leaves him vulnerable to another iteration of the problem of causal deviance. In this section, I shall show how.[30]

6.3.1 Delusion About One's Reasons

The self-referential belief SR is supposed to capture the distinctively self-conscious character of motivation in rational beings. Setiya's overall account depends on the claim that the notion of a disposition outlined above, together with the content of his self-referential belief SR, are together enough to capture the idea of an agent self-consciously acting for a reason. However, the overall tendency of our argument leads us to suspect that SR will fail to provide Setiya with what he wants from it.

One obvious test of Setiya's claim is whether his self-referential belief could be substituted—*salve veritate*—for the belief that one is taking p as one's reason to ϕ. However, without further specification, it seems that the belief that Setiya identifies could be true, and yet the belief that one is ϕing for the reason that p be false. As an example of such a case, consider the following scenario:

> Imagine an individual, A, who is the sort to take pleasure in knowing everyone's business and telling them what to do. A has an opportunity to take up a position

the content of the belief, or a specific act of the agent. Setiya suggests the former, which suggests that it is somehow the belief, rather than me, that brings it about that I act for a particular reason.

[30] Since he has failed to provide an account of 'motivation' that can differentiate the broader class of cases in which I may not have first-personal knowledge of my reasons for action and what he considers deviant cases, his account is also vulnerable to a version of the objections raised in the previous section. For it seems perfectly possible that I might think

> I am hereby ϕing because of my belief that p,

and thereby be caused to ϕ, without doing so on the basis of my self-referential belief (i.e. my intention). Suppose, for instance, I decided to lie down in the belief that it would soothe my headache, and as I was getting up I tripped and ended up lying on the floor. I would be lying on the floor because of my intention to do so, but not in fulfillment of that intention. Since Setiya lacks a plausible story about 'motivation', his account remains vulnerable to this kind of case. The case I consider in §6.3.1 is a further iteration of this kind of problem.

> of authority, and is motivated (in Setiya's sense) to do so by the knowledge that it will give him pleasure to exercise this authority over his peers—in other words, what in fact causes him to seek this position is the belief that attaining it will give him pleasure. Suppose further that A is not so clear-headed as to acknowledge, to himself or to others, that this is what motivates him to seek the position; instead he believes he is seeking it because his past experience has better prepared him for the position in question than any other potential candidate.

So far this looks like a perfectly familiar case of delusion: what motivates A to seek the job is the promise of a certain kind of pleasure, and *this* is what causes him to seek it: his stated 'reason' is so much hot air. But now suppose the following is also true:

> Like most of us, A likes to think he is acting for noble reasons; in fact, he would not be seeking the position if he could not find some fact that *would* justify his act: he is justified in his belief that he has more experience than any other candidate, and this would be a good reason for him to apply for the job, though it is not what in fact motivates him. The delusion that he is motivated by noble reasons is integral to his undertaking the relevant act.

It seems plausible that delusions often work in this way: that people are only able to act in a certain way because they can find reasons other than those that genuinely motivate them to justify and putatively explain their action. The difficulty for Setiya's account comes from the fact that his characterization of the role played by SR sounds like a case of delusion. Recall Setiya's own characterization of his view:

> To do something because p is to be motivated by an explanation of one's action that appeals to belief. [32, 51]

The self-referential belief represents a person as motivated (because caused to act) by the idea that they are motivated by a particular belief, and that fits cases of delusion: we bring ourselves to behave in a certain way because we like the idea that in doing so we are acting for good reasons.

Now notice that if, in such a case, the person would not have acted if they did not believe there was some acceptable reason for what they did, it seems right to say that the relevant facts, and the psychological states associated with them, must enter into the causal explanation of the action: had A not believed that he was well-suited to the job, and that he was pursuing it for this reason, he would never have applied for it. Given this, the following could, in the right circumstances, stand as an explanation of A's act:

> A applied for the position because he thought that he was the best possible candidate.

Of course, such an explanation would only be appropriate to someone who understood that it aimed to explain which particular delusion was driving *A* on this occasion. For, per our stipulation, A does not 'act for this reason', but from some other consideration: because exercising the authority that comes with the position would give him pleasure. Setiya's self-referential belief seems like an ideal vehicle for such a delusion: the person believes that she is ϕing for the reason R; and, in virtue of this very belief, does indeed ϕ; but, per our stipulation, not for the reason R. If such a delusion *is* an instance of the kind of self-referential belief that Setiya has specified, then its truth does *not* entail the truth of the belief that one is doing ϕ for the reason R, and the two contents cannot be exchanged *salve veritate.*

Setiya could, of course, embrace this kind of example: he could claim the fact that our agent is best-suited for the job *is* *A*'s reason for acting, since he makes it such by his delusional belief (which on this account is perhaps not so delusional), which as such motivates his action. But he is also motivated by a further belief, that he will be able to derive pleasure from the exercise of power that comes with this job. This is *a* reason for his action, in that it explains his action, without being *his* reason for his action, because he did not *take* it as such.[31]

Is this an acceptable description of the case? To me, it seems to get things backwards. What we want to say is that A *didn't really* act because he thought he was the best candidate. If this belief led to his action, it is surely a case of what we might call 'deviant motivation'. The real reason—the one that we would say, in the familiar sense, motivated the action—was the more sordid one, and we want our account (and our assessment of A) to reflect that fact. This means that A can't simply make it the case that the noble belief was his reason by believing that it was. We want to be able to hold onto the fact that, however the noble belief might have been involved in the aetiology of A's action, it does not explain it in the way it would were it really A's reason for action.

Causal Deviance and SR

What we have here is an essence another version of the problem of causal deviance, but one that does not easily fit with Setiya's understanding of deviant causality as involving 'accident'. The case is deviant, relative to Setiya's account, because it involves a self-referential belief (SR) about a further belief causing one's action, and along with that a real causal relation between that further belief and the action under consideration. The problem is that the further belief is not related to the action *in the right way.* Setiya needs to show us that his account has the resources to recognize and rule out this deviance.

The problem we are seeing here is that Setiya's account forces him to treat *any* self-conscious causal aetiology (i.e. aetiology that essentially involves his self-conscious state) as amounting to a rational explanation. Any belief plugged into the right slots in such an aetiology just is, as such, the person's reason for their action, and is recognized as such by

[31] Another option Setiya might instead insist that the deluded person *is* motivated by the self-referential belief, and with it the noble belief. To be motivated by a belief involves a disposition to be moved by it, and our deluded agent *is* disposed to certain actions because of their noble belief. So, in this sense, the belief that he is the best qualified candidate for the job *is* his reason for acting, even though, from another perspective, we could insist that it was not his real reason at all. As with the response I consider in the main text, the plausibility of this line of thought depends on a particular understanding of the 'dispositions' involved in intentional action, which I discuss in the next section.

them because of the role played by the self-referential belief. But in at least some cases, beliefs that are part of a self-conscious causal aetiology, while certainly being something we might count as a reason why the person acted (i.e. something we could include in an explanation of their action), are nonetheless not something they would recognize as their reason for acting, or indeed as a reason to act at all.

Thus cases of delusion bring out a further problem with Setiya's account: it only really captures the idea of 'the reason S ϕed', and not 'S's reason for ϕing'. His guiding idea is that we can transform the former into the latter by adding a self-conscious element to the aetiology of the action. But that still leaves us with the problem of differentiating cases where the agent is, as we might put it, self-consciously involved in the reason for her action, and cases where she genuinely takes something as her reason. In other words, the account seems to collapse together two different kinds of case: a person can recognize the casual role played by a particular belief, without thinking that the content of that belief provides any reason for the relevant action; and a person can recognize (or believe) that a belief does provide reasons for an action, and act on that knowledge. Setiya's account treats the former as though it were just an instance of the latter, with the result that he is forced to count such cases as examples of *acting for a reason*, and the explanations we might describe them with as reason-giving explanations.

6.4 Normativity and Explanation

In §6.2 and §6.3.1, I have shown that both stages of Setiya's account remain vulnerable to a version of the problem of causal deviance. Because he does not provide a detailed account of the causality involved in his account, it is difficult to directly apply the arguments specific from the previous chapter: we don't know enough about how Setiya conceives these causal relations to see where he has gone wrong. However, the general form of those arguments still applies: whatever the specifics of the problem, Setiya's difficulties seems to arise from the fact that he wants a reductive account of his topic. For it is this that prevents him from appealing to ideas related to ‘acting for a reason’ to specify what it is for a psychological state to cause an action ‘in the right way’.

This means that the source of Setiya’s difficulties is not so much the specifics of the kind of causality he pictures as his whole way of conceiving that causality. For whatever its details, Setiya’s reductivism requires that the causality in question be specifiable independent of any appeal to the idea of *acting for a reason*, and this leads Setiya to reject the idea that any appeal to ‘normative’ or ‘justifying’ reasons can be essential to its specification. It is his insistence on the non-circularity that characterizes reductive accounts that leaves him vulnerable to the various iterations of the problem of causal deviance.

In this final section of the chapter, I will directly address this move on the part of reductive accounts to separate the normative from the explanatory. As well as being the ultimate source of the vulnerability these accounts have to the problem of causal deviance, this distinction also commits them to a number of bizarre and perhaps ultimately unintelligible claims. Showing this will provide essential background for a further appraisal of non-reductive accounts in §7, since such accounts consciously reject both the overall form of reductive accounts, and their more specific distinction between the normative and the explanatory.

Dispositions and Reductive Accounts

It is helpful to begin this discussion by returning to Setiya’s notion of a disposition, since some such notion will also play a role in non-reductive accounts. Moreover, the generic notion of a disposition described by Setiya provides an illuminating reflection of the overall form of his own account: for it is a key point of Setiya’s account that he attempt to characterize this disposition without any appeal to normative relations that characterize its content.

Such dispositions come relatively cheaply: any reliable transition will count as a manifestation of one. So long as the agent reliably transitions from C to E, the disposition is in place, regardless of the character of C or E. If such dispositions involve normativity at all, it is only insofar as they lead us to expect certain transitions to take place, and perhaps make us look for further explanation if the transition does not happen when it ought to.

This means that Setiya is committed to the claim that there is no general constraint on which dispositions are *possible* sources for the causal relations involved in acting for a reason.[32] Someone could, in principle, be motivated to act by *any* psychological state by which they were disposed to be moved, no matter how bizarre the connection between the content of that state and the action it explained.[33] Setiya recognizes that the fact that his view entails this possibility 'may seem to be a serious objection'—and this is indeed what I shall argue—but he suggests that his theory actually gets things right here.

> [In the bizarre cases] we assume that these dispositions are missing, and thus that the causal stories are deviant. Once the dispositions are supplied, however, the possibilities in question seem bizarre, but real. Imagine, for instance, that I have been conditioned to put a dime in my pencil sharpener when I believe that I could get the drink I want from a nearby vending machine. This has become a stable tendency of mine. Strange though it is, I can now be *motivated* to put a dime in my pencil sharpener by the relevant belief. And it is not impossible for me to put a dime in my pencil sharpener on the ground that I could get the drink I want from a nearby vending machine. (An odd decision, but there it is.)[34]

This generic notion of a disposition—independent of any normative relations between cause and effect—is essential for Setiya's point about the intelligibility of such scenarios.

[32] Though, as we shall see, he does argue that there are constraints on the kind of dispositions that are likely to be established and persist in rational creatures like us

[33] Also, as we shall see, regardless of whether they themselves *take their to be* any intelligible connection between that content and their action do, besides the brute fact of their disposition to be moved (i.e. their 'motivation').

[34] In fact, Setiya's view entails not just that I can be 'motivated' to put the dime in the pencil sharpener, and so find myself acting in this way—and presumably be surprised by that fact—but even than I can quite self-consciously *take* the fact that I can get a drink *as* my reason for behaving in this way, even while denying that there is *any* normative connection between these two things. That is, I can consciously and deliberately take something as my reason (in the motivating sense), while admitting that it is no reason at all (in the normative sense), just so long as I am in fact disposed to act because of the relevant psychological states. On this further point, see §6.4.1

The idea of such dispositions is supposed to encompass the case in which I am disposed to put a dime in a pencil sharpener whenever I believe that I could get a drink from a nearby vending machine. There is, we are supposing, no normative relation between these facts—the fact that I could get a drink I want from the nearby vending machine provides no genuine 'normative reason' for putting a dime in the pencil sharpener, nor does the agent suppose that it does. The only connection between these two things comes from the fact that the agent happens to have a disposition to transition from one to the other, and the mere fact of this disposition is enough to make the belief that he could get a drink from the nearby machine count as the agent's reason for acting. The apparently bizarre, irrational nature of this disposition is irrelevant.

Of course, Setiya thinks that in most cases our dispositions will be more familiar, perhaps more 'rational', and as such reflective of genuine normative relations. This would be the case if the person we are considering had the more familiar disposition that transitions from believing 'I could get the drink I want from a nearby vending machine' to putting a dime in its coin-slot. Again, on Setiya's picture, the disposition itself is just a particular fact about this person: someone could have such a disposition, and act from it, regardless of any connection between putting dimes in the coin-slot of a vending machine and getting a drink. However, it so happens that in our world this is a particularly useful disposition to have: it will make one more successful in getting drinks than the alternative described above, and perhaps because of this is much more widespread. But though this disposition is useful because it reflects genuine normative relations between things, the possibility of such a disposition does not essentially depend on the reality of those relations.

Thus, while this particular disposition happens to reflect normative relations that characterize our world, it is, in itself, not fundamentally different from the irrational disposition. The characterization of a disposition as rational or irrational, or its characterization in other normative terms, is external to a fundamental characterization of the disposition itself. And while it is more likely that the dispositions involved in acting for a reason will reflect normative relations—dispositions of this sort are more useful, and a majority of them is essential to our counting as minded at all[35]—there is nothing about the very idea of *acting for a reason* that requires this of the dispositions involved. Someone who is motivated to act by an irrational disposition presents as good a model of *what it is to act for a reason* as someone who is acting from a rational one.

It is no accident that dispositions of this sort should be at the centre of Setiya's account. They are a reflection of the particular way of understanding the distinction between normative and motivating reasons that we considered in §2. Because causal-psychological accounts see motivating reasons as completely distinct from normative reasons, they count it as an advantage that they purport to explain the causal relations involved in intentional action in a way that is completely independent of any normativity. Thus the dispositions that ground such causal relations must also be at best externally related to normative facts, since if they were not, the causal relations they explain would seem to depend on the relevant normative facts, and would be as such evaluable in terms of them. But it is the central claim of causal-reductive accounts that there is no special kind of causation involved in acting for a reason, and no special kind of explanation either. Thus in Setiya's account, these causal relations must be seen to depend on dispositions that are independent of normative relations, and that play their explanatory role without appeal to them:

> In general, the dispositions that govern intentional action must approximate the standards of good practical thought. The pressure is for one's dispositions, taken together, to be at least moderately good. But there is no reason to suppose that, in each instance, practical thought or the motivation of action must be "made intelligible by being revealed to be, or to approximate to being, as [it] rationally ought to be" [26], or that **a special kind of explanation is involved**. In a given case, one's reason for acting may be simply and irredeemably bad, with no resemblance to anything that could count as good or adequate in practical

[35] See Setiya's comments on Davidson's constitutive ideal of rationality [32, 65]

> thought. And it may be explained by the corresponding disposition, **not in essentially normative terms**. [32, 66] [my bold]

6.4.1 Cases of Logical Deviance and Alienation

Davis on Intelligibility and Action

This sharp distinction between the normative and the explanatory—a distinction that, as we've seen, is common to all causal-psychological accounts, emerging as it does from their overall form—forces such accounts to embrace the possibility of various bizarre scenarios. For if one understands motivating reasons as *purely* explanatory—and, moreover, as explaining by reference to causal relations that are independent of normative relations—then there is nothing to rule out explanatory causal relations of this sort holding between beliefs and actions that do not stand in any genuine normative relations to each other.

There are more and less extreme versions of this view. For instance, on Wayne Davis' account of *what it is to act for a reason*, the following is a genuine 'logical possibility': a world in which most (or even all) people, when they believe that p, are generally caused to act by desires of the form ϕ if $\neg p$. In other words, it is perfectly possible for someone generally prompted to action by beliefs or desires that are *no* reason for the act in question. Such a person or group, unlike us, would not be called "rational animals in the descriptive sense", and in fact such behaviour "would seem to be a real possibility for those suffering from extreme psychosis".[36] But since the causal nexus that characterizes *acting for a reason* is empirically discovered, there is nothing incoherent about imagining scenarios in which a different causal process relates mental states to behaviour in a completely different pattern from the one to be found in the lives of 'rational animals in the descriptive sense'.

It is worth pausing to spell out more explicitly what Davis is committed to here. This is not simply the idea that a person might, on occasion, take themselves to have grounds for

[36] [18, 85]

an action when in fact they had no reason to act as they did. Nor is it the idea that, on occasion, a person might be prompted to behave in a certain way by beliefs that were, by their own lights, no reason to behave in that way. In both of these cases, we are imagining isolated or occasional disruptions within patterns of behaviour that are broadly rational. Davis is asking us to imagine a case in which no such broader pattern is discernible: we count people as having the relevant beliefs and desires, but they are prompted to action in a way that does not hang together 'rationally' with the psychological states we attribute to them.

Suppose for now that this scenario makes sense, and that we came across creatures who were wired in this way—let's call them Contrarians. Given the belief that p, and the desire to ϕ if $\neg p$, such a creature would generally form a desire to ϕ, and then ϕ because of this desire. Of course, given Davis' definitions, we could not say that it 'acted for the reason that p'; and presumably (though things get a little hazy here), the Contrarian's desire to ϕ, and subsequent ϕing, would not be an s- or f-result of the relevant states.[37] But its belief that p, together with its desire to ϕ if $\neg p$, would nevertheless be the reason why it ϕed, in the sense that it was the cause—and thus proper explanation of—its action.

That humans are not Contrarians, as a species or as individuals, is also—if I understand Davis's view—an empirical matter. Perhaps finding out facts about the species in general helps us to predict, of any given individual, that they are unlikely to be wired like the Contrarian. And perhaps those of us who are akin to Contrarians (like Davis' 'psychotic') will turn out to be a rarity. But it is an implication of Davis' account that things might

[37] Supposing for a moment that the idea of a Contrarian is coherent, the difficulty here can be made more explicit by focusing on an assumption guiding Davis' account. He is committed to the claim that there *must* be some difference between the causal processes involved in the Contrarians behaviour and the causal processes involved in our behaviour, and that this difference can be specified in terms that are independent of our understanding at least our own actions as such as to be done for reasons. But one might question the warrant for this claim: why think that we can know, prior to actual empirical investigation into e.g. the physiological and neurological underpinnings of action, that there *must* be a difference between the physiological and neurological processes of the Contrarian and the analagous processes in our own bodies? If empirical investigation revealed no such difference, would we have to give up on the idea that the Contrarian was fundamentally different from us in the way described?

have been otherwise: it might have turned out that most or all humans are Contrarians, and that nobody ever acted for a reason, despite having beliefs and desires that would enable them to if only they were wired the right way. Perhaps in such a scenario there would be the occasional rational agent, serving as a counterpart to Davis' 'psychotic'.

In fact, on Davis' account the only reason we have for saying that nobody 'acts for a reason' in this scenario seems to be a definitional fiat: we have defined *what it is to act for a reason* in such a way that it excludes the behaviour of Contrarians. But otherwise everything seems to be in logical or metaphysical order. At worst, what we've described is empirically unlikely or perhaps impossible, if only because our Contrarians probably wouldn't do very well in the world as we know it.

Setiya on Intelligibility

Davis' commitments are extreme because they emerge from the combination of his reductive ambitions and his empiricist tendencies. As we saw earlier, Davis is quite happy to defer the task of spelling out what it is for a psychological state to cause an action 'in the right way' to the sciences, and because of this is committed to the claim that there is a sense in which, until such a task is completed, we don't fully know *what it is to act for a reason.* Philosophers like Setiya—who take their task to be one of articulating the understanding we already have of ourselves, insofar as we traffic in rational explanations—cannot be satisfied by this approach. For it presents our understanding of the essential nature of such explanations as partial and in need of supplementation. And while this might be a quite sensible commitment when it comes to explanations of phenomena that are quite independent of us (such as the movements of the tides, or the behaviour of sub-atomic particles), it seems a poor fit for the kind of explanations we can provide of our own actions, just insofar as we act.

Nevertheless, Setiya's account of the causal underpinnings of rational explanations also commits him to the possibility of bizarre situations: in particular, cases in which an agent 'takes' some consideration as the motivating reason for his action, without seeing it as—in any way—a normative reason for his action. The possibility of such cases follows directly from Setiya's rejection of the 'guise of the good thesis', along with his whole-hearted embrace of the radical distinction between motivating and normative reasons, both of which are expressed in his account of motivation as resting on normatively-neutral dispositions characterized above. As we saw, Setiya sees that his account allows for the possibility of seemingly bizarre cases, but he claims that such cases are indeed possible, and that it is a virtue of his account to allow them.

One difficulty lies in understanding the character of the cases Setiya imagines, since they are not described in much detail. Since there are a variety of possible cases imaginable here, it will be helpful to quote Setiya's description in full, in order to work out exactly what he is committed to here:

> Imagine, for instance, that I have been conditioned to put a dime in my pencil sharpener when I believe I could get the drink I want from a nearby vending machine. This has become a stable tendency of mine. Strange though it is, I can now be *motivated* to put a dime in my pencil sharpener by the relevant belief. And it is not impossible for me [to] put a dime in my pencil sharpener on the ground that I could get the drink I want from a nearby vending machine. (An odd decision, but there it is.) What is true is that . . . such cases are necessarily rare, and that, when they occur, I will typically reject my own behaviour as irrational and consequently give it up. As a *local* phenomenon, however, it is possible to be moved in ways that do not even tend towards good practical thought. [32, 65]

Setiya's thought is that since the disposition to behave in the relevant way is present, this is a genuine case of reasons-explanation: the agent is caused to act by the relevant belief and desire, and not in a deviant manner.

What is unclear is how the person understands their actions: according to Setiya's view, they are in a position to provide an explanation for what they do (acting for a reason entails this). So, on being prompted to explain their behaviour, they would reply that they were putting a dime in the pencil sharpener because they wanted a drink from the nearby vending machine. However, Setiya's description does not allow us to go much further into the imagined agent's psychology. Do they also think that putting a dime in the pencil sharpener is a *means* to getting a drink from the nearby vending machine, or perhaps required if one is to do this? Such beliefs, however bizarre, *do* make their actions intelligible, because they involve the idea of some kind of normative connection between the two acts—the question then would be how they came to have such beliefs in the first place.[38] But once the case is understood in this way, it is no different from any other example of action on the basis of false belief. What's more, it would seem to force us to take the role played by these normative relations more seriously, in a way that pushes against Setiya's rejection of the guise of the good thesis.

Another possible way of understanding the case is as one in which the agent is somehow alienated from their action: they feel compelled to put the dime in the pencil sharpener, without understanding why. This doesn't quite capture the scenario Setiya describes, since there it is important that the agent would explain their action by reference to their desire to get a drink from the nearby vending machine (rather than by reference to a felt compulsion). We can, perhaps, imagine the agent *noticing* a connection between this compulsion and the desire to get a drink in this manner; but if this is something they could notice or discover about their behaviour, it seems that the self-reflexive character of the relevant psychological states emphasized in Seitya's account is missing.

[38] This brings out a further difficulty in understanding Setiya's imagined example. He says that the agent has the relevant disposition as a result of conditioning. But one might ask how exactly such conditioning is to work. If, for instance, the agent has been put in situations in which he cannot get a drink from a vending machine *unless* he puts a dime in a pencil sharpener, then his having the belief that one is a means to, or at least required for, the other becomes intelligible. This makes the case like any other instance of imperfect intentional action, and no longer genuinely irrational. If we are not to understand the case in this way, then how was the relevant disposition installed? After all, 'conditioning' takes time and repetition, and as Setiya notes, anyone who found themselves behaving in this way on a single occasion "will typically reject [their] own behaviour as irrational and give it up". How, then, is conditioning to continue, unless the person is put in the kind of situation just described in which their behaviour will not strike them as irrational?

What we need is a case in which the agent believes something of the following form:

> I am *hereby* putting a dime in the pencil sharpener because I want a drink from the nearby vending machine.

or

> I am *hereby* putting a dime in the pencil sharpener because I believe that I could get the drink I want from the nearby vending machine.

and *in virtue of thinking that thought* makes it the case that this is their reason for acting. In other words, the agent needs to have a self-referential belief that they are acting this way because of their further belief about the vending machine, and to act *because* of these beliefs (i.e. their action must be caused by them).[39] They cannot simply find themselves acting in this way: they must be prompted by a self-reflexive belief.

Such cases seem to involve a strange kind of alienation on the part of the agent: she knows what she is doing, and in some sense she also knows why she is doing it, but the explanation doesn't help her make sense of her action at all.[40] While she does know the cause of her action, she does not understand *why* she was acts that way. It seems to me that what we should say here is not that she knows the correct reason-giving explanation of her action, but rather that there is no such explanation to be known.

Perhaps the closest we can come to the case as Setiya imagines it is an agent with a bizarre psychological disorder, who acts from an understanding of this disorder. For instance, we can imagine someone offering the following explanation for their action:

[39] Perhaps this provides a clue to how Setiya understands the 'conditioning' involved in installing such dispositions. Presumably, at first, the agent needn't have a belief of this form. This could perhaps be how they only later come to notice the connection. However, once the disposition is stable and they are aware of it, actions prompted by it might perhaps involve a belief of the form Setiya requires.

[40] It is not like a case of expressive action, where I can e.g. know that I'm tapping my fingers because I am nervous, though I don't take the fact that I'm nervous to be a reason to tap my fingers. I have no reason for the action at all, but recognizing it as expressive helps make sense of it.

> I am *hereby* putting a dime in the pencil sharpener because I believe that I could get the drink I want from the nearby vending machine, since such beliefs make me unbearably anxious until I put a dime in a pencil sharpener.

Here a belief with the form of SR would be true:

> I am *hereby* putting a dime in the pencil sharpener because I believe that I could get the drink I want from the nearby vending machine.

But I would suggest that this is another example of deviant causation. While it is true that the agent acts because of his belief—and indeed makes it the case that this is so—he does not take his belief itself as a reason to act, but rather the attendant anxiety. What's more, such a case shows that the idea of deviant causation doesn't—as Setiya supposes—rest on the idea that the relevant behaviour is an accident. The anxiety in question here could be a regular and reliable fact about the poor agent's psychological life: as stable a disposition as any you care to name.

In such a case the agent recognizes the role played by this particular belief in the aetiology of his action, without acknowledging the belief as a reason for his action. Since the belief is involved in causing his action, we could count it as the (or at least a) reason why he acts—that is, we might include the belief in a causal explanation of the action. But it is not his reason *for* acting: if anything could be described as such, it would be the goal of relieving his anxiety.

Contrarians and Supra-Logical Aliens

We can also come at this issue at a more general level by comparing Setiya's position to the one defended by Davis. As we saw, Davis' empiricism led him to embrace the possibility of Contrarians, and Setiya's arguments here must also force him to the entertain this possibility. In Setiya's terms, a Contrarian would be a person who, instead of having a locally irrational disposition, had a more general disposition such that, when they desired that p and thought that ϕing was required for p, they would not ϕ.

Setiya's response is to suggest that general dispositions of this sort can be ruled out on Davidsonian grounds: if we are to be counted as minded at all, most of our dispositions must be rational, and a general disposition of this sort would mean that that couldn't be the case[41] But the pressing question is why? Davidson's view at least involved a sharp distinction between the genuinely causal relations expressed in physical concepts, and the normative relations expressed in mental concepts. But Setiya wants the best of both worlds: that is, he wants to hold on to Davidson's claims about holism, while allowing a normatively-neutral kind of causation to be genuinely illuminate the relationships between items that are to be described in essentially mental or normative terms. Davidson at least maintained that genuine causal explanation involved describing events in physical terms that showed them to be instances of general causal laws—in other words, in terms that are not available to us just insofar as we are rational agents who know what we are doing. But given that the idea—central to reductive accounts—that *acting for a reason* is independent from the normativity that grounds the holism of the mental and the constitutive ideal of rationality, why should it ever be subject to such constraints? Simply insisting that it is leaves us in the dark

One final way of bringing out this point is to imagine another species of logical alien. Unlike the Contrarians, these creatures would generally be rational, and as such their actions and attitudes would for the most part be characterized by the constitutive ideal of rationality. But unlike us, these aliens have the following, quite special, ability: they can create dispositions in themselves at will, and so form self-referential beliefs with the form of SR to do anything on the basis of anything.

Even if Setiya is entitled to his appeal to Davidsonian considerations, it is not clear that they apply here. For we are stipulating that, when these aliens exercise this ability, they acknowledge that p is in fact no reason to ϕ. Nonetheless, they are making use of their

[41] See [32, 65].

peculiar ability to form the relevant dispositions at will to make it the case that they are ϕing because of that belief.[42] The actions and attitudes of such a person would generally be compatible with the constitutive ideal of rationality; and, for all I can see, their self-referential beliefs would be too, because given Setiya's account its not at all clear what kind of rational constraints there could be on such beliefs, beyond those involved in certain aspects of their content. That is, we might be able to trace rational constraints that have their source in the belief that p, which is explanatory of the act; and perhaps we could trace rational constraints that have their source in the belief that one is ϕing. But since the self-referential belief is, as I would put it, merely explanatory, then so long as a person *is* capable of making it the case that they are doing ϕ because of the belief that p, there is nothing irrational in their believing that this is what they are doing.[43]

Compare this with the belief that one is ϕing for the reason that p, understood as committing the agent to the idea that p *is* a normative reason to ϕ, i.e. as playing more than a merely explanatory role. This *would* entail further constraints that ruled out the agent also holding the belief that p is *not* a reason to ϕ. So such a person could not hold the following two beliefs, on pain of irrationality:

- I am hereby doing ϕ on the grounds that p.
- p is no reason to ϕ.[44]

Here the incompatibility is clear: but it involves the idea that p provides grounds for—and is thus a normative reason for—the relevant action.

[42] Perhaps this is how Setiya ought to understand his example involving the dime and the pencil sharpener.

[43] It will help here to imagine that the aliens in question only exercise this ability on special occasions—perhaps as a joke. Most of the time they are just like us.

[44] I take that this to mean that p somehow speaks in favour of ϕing, even if it doesn't in fact make it the thing to do all things considered.

Summary

What I hope to have shown in this section is that the distinction between normative and motivating reasons—which, once again, emerges from the overall form of reductive accounts, and lies behind Setiya's normatively-neutral characterization of their causal underpinnings—by itself introduces the possibility of logical deviance and alienation described here. Indeed, considered in themselves, causal-psychological accounts have to treat these cases as perfectly good examples of *acting for a reason.* This by itself provides us with grounds for dissatisfaction with such accounts, and in subsequent chapters I will be arguing that it is only by deploying normative notions in characterizing *what it is to act for a reason* that we are able to to make sense of (or reject) the imaginary scenarios described here.

CHAPTER 7
NON-REDUCTIVE ACCOUNTS

7.1 A Non-Reductive Account?

In the last chapter I argued that any reductive account of reason-giving explanation will be unable to distinguish acts subject to such explanations from deviant cases, and so fail to provide its proponent with what they want from it: a substantive response to our philosophical questions (MQ) and (EQ). Since these difficulties arise directly from the reductive form of the proposed account, they also lend support to the search for a non-reductive (or perhaps even primitivist) account of such explanations. In this chapter I shall consider one strategy for providing such an account: the 'normativism' of philosophers such as Eric Marcus and Sebastian Rödl.

As we shall see, the normativist strategy promises to develop the basic insights from Wittgenstein' and Anscombe's work discussed in §3 into a full response to our two philosophical questions. Though I shall argue that this appearance is ultimately illusory, I shall also suggest that the careful attention to the strengths of the normativist strategy, and the way that strategy is framed, together point towards a more promising way of connecting these insights to the contemporary discussion.

Normativism

As we saw in §2.3.3, 'Normativism' names a particular strategy for accounting for the distinctive character of reason-giving explanations, along with the acts they describe. The basic form of a normativist account is as follows. First, acting in general is identified with representing the action in question as having a normative status such as 'good' or 'to-be-done'. Second, 'acting for a reason'—i.e. doing A because one is doing B—is identified with representing one action as having this status because the other has this status, i.e. representing doing A as to-be-done because doing B is to-be-done.

The normativist understands this strategy to involve commitments to two specific claims:

1. Rational explanations depend on a distinctive form of causality, one that is constituted by the agent's representing of normative relations, and so cannot be understood independently of those relations.[1]

2. The fact that the causality consists in the agent's representings explains why she is in a position to provide an explanation of her action, i.e. to answer Anscombe's question 'Why?'.

For instance, Sebastian Rödl begins his account of intentional action from the following two propositions [30, 138]:

- Doing something intentionally is representing doing it as good.

- Acting intentionally in a certain way is representing acting in this way as good.

Eric Marcus makes analagous claims:

- [A]ctions just are agents' representings of to-be-done-ness.[2]

- Representing ψing as to be done, when successful, just is ψing.[3].

Claims of this form provide the basis for an account of rational explanations of action. The normativist maintains that when a person's actions are subject to rational explanations, the causal-explanatory relations represented in those explanations are *identical with* representings of normative relations on the part of the subject. As Eric Marcus puts it, "[a]cting-for-a-reason is representing ϕing as to be done as a consequence of the to-be-doneness of ψing" [24, 111]. This is a part of the claim that action-explanations of the form

[1] This is a fundamental difference from causal-psychological accounts, which agree with the normativist in treating rational explanations as causal explanations, but claim that the causality involved can be understood independently of normative relations.

[2] [24, 72]

[3] [24, 169]

S is doing A because she is doing B

depend on a distinctive kind of causality that is constituted by the subject representing ⟨doing A as to-be-done because doing B is to-be-done⟩:

> [I]f someone is doing A because she wants to do B, . . . then her instrumental thought constitutes the causality the explanation represents. . . . [A]n action explanation "She is doing A because _____" is true, only if she who is doing A thinks "I should do A because _____". If someone is doing A because _____, then her thought that she should do it because _____ is the causal nexus that the first "because", the "because" of action explanation, represents. [29, 48]

In both cases, these claims about the causality of rational explanation are taken to explain the agent's first-personal knowledge of what she is doing, and why she is doing it. This time I will just quote Marcus:

> To act for a reason is to represent the to-be-doneness of an action as following from the to-be-doneness of another action. . . . Because acting-for-a-reason consists in an agent's so representing, agents can say, as Anscombe emphasized, what they are doing and why, yet not on the basis of observation or evidence. The ability to do what is to be done as a consequence of another action's being to be done is. . . necessarily self-conscious. [24, 7-8]

Normativist accounts thus purport to give substantive answers to our philosophical questions (MQ) and (EQ). In response to (MQ), they claim that to act for a reason is for one's actions to be the result of a distinctive form of causation, which consists in the act of a 'rational ability'. Furthermore, since the acts of that ability consist of representings of normative relations between one's actions, the ability further explains how it is we come to have first-personal knowledge of our reasons for acting: to act for a reason is to represent one's action as inheriting a normative-status. This means that, in contrast to primitivist accounts, the normativist has something substantive to say about the explanatory relations involved in acting for a reason.

The normativist strategy therefore seems to promise us the best of both worlds: a way of bringing out what is helpful in the Wittgensteinian ideas we have considered in previous chapters—particularly the role that normativity plays in the actions described in reason-giving explanations—but in a way that explicitly connects those ideas to the philosophical questions that shape the contemporary literature.

However, in this chapter I shall argue that this promise is illusory. Normativist accounts in fact face a dilemma that is internal to the normativist strategy: if they represent themselves as providing substantive responses to our philosophical questions, then the specific normativist claims become mysterious; but if they domesticate those claims, they no longer serve as substantive responses to our philosophical questions, and in fact serve to obscure the character of the rational abilities they purport to describe.

7.2 The Normativist Strategy

7.2.1 Normativity and Explanation

Before developing these objections, it will be helpful to get a general overview of the normativist strategy. They key ideas behind the normativist's approach are as follows:

a. Acts that are subject to reason-giving explanation involve an understanding on the part of the agent of the relevant form of normativity, and are therefore

b. known by the agent as acts that involve that form of normativity.

These points should already be familiar from our previous discussion. To return to our earlier example, an act of reading is an instance of a kind of responsiveness to a particular form of normativity, which might be expressed in the rules for rendering a particular language. That the reader knows her acts *as* acts involving this form of normativity is shown not just by the acts themselves, but by the descriptions she provides of them, and the questions which she knows make sense to ask of them.

The normativist strategy involves an distinctive approach to these ideas. This is most explicit in Rödl's work, and it will be helpful to put it in his terms. His account involves three distinct ways of describing an action:

1. The movement or happening: Doing A
2. The self-predication: I am doing A
3. The representing-as-good: Doing A is to-be-done. (Or: I ought to do A.)

All three of these are introduced as ways of describing one and the same action, meaning that the 'thoughts' described in (2) and (3) are identical to the movement in (1): (2) because realizing the action-concept *doing A* is identified with a self-predication of that concept, and (3) because the form of that self-predication is signified by characterizing it as a representing of goodness or to-be-doneness. Thus (1)-(3) provide three different ways of specifying or describing the same act, providing the basis for the central normativist identity statements:

- To do A *is* to think I am doing A
- To think I am doing A *is* to think A is good to do, or to-be-done.

This gives us the main moves of the normativist strategy as it applies to action:

1. The idea of acts of representing something as having normative status;
2. The claim that those acts of representing are identical to acts of practical abilities.

To understand the role played by normative status in this approach, we have to see how it is tied to the role that normativity plays in reason-giving explanations. Before considering the one kind of case the normativists focus on, it will be helpful to put this point in more familiar terms. Earlier I suggested that a reading explanation of the form

S said "…" because$_R$ "_____" is written on the page,

belongs together with a rule of the form

"…" is to be read "_____",

which can be understood as an expression to the normativity in terms of which the reader understands her acts. The subscript in our explanations marks the fact that the acts it describes are acts of this kind. This means that the explanation as a whole represents the acts it describes as a manifestation of an ability to respond to this form of normativity, a fact that is further reflected in the ways in which the reader can explain, justify, evaluate, or correct her act. In this sense, our subscripts are an attempt to mark the form of the explanation—something that is more naturally expressed in the grammar of descriptions like

S read "…"

and the further questions it makes sense to ask of them, e.g.

And where was she reading from?

This grammar reflects our understanding of acts of reading as manifestations of a kind of responsiveness that belongs to a particular kind of practice.

The normativist's appeal to representing of normative status can be understood as a way to capture these same points. Whereas I used these subscripts as a marker to indicate that the explanation as a whole represented a particular kind of act which involved an understanding of a particular form of normativity, the normativist captures these same ideas through their descriptions of 'rational abilities'. Thus, whereas I merely suggested that the explanation and the rule belong together, the normativist would describe the act represented in the explanation as a representing of the normativity expressed by the rule. An act of reading could then be described as an act of representing "…" as to-be-said because "_____" is written on the page,

or perhaps,

representing "…" as to-be-said because "_____" is to-be-read.

The key point is to capture the idea that the act is a response to a particular form of normativity, and known as such by the agent. Describing the agent as 'representing normative status' is the normativist's strategy for expressing this idea.

Application to Intentional Action

Of course, normativists are concerned with a much broader class of explanation that description of acts of reading, and rather than focusing on simple instances of rule-based responsiveness, they focus on such things as complex intentional actions. Developing an insight from Anscombe's *Intention*, one of the ways in which they aim to capture the distinctive form of normativity involved in such action is by way of a comparison with practical inference. In her work, Anscombe notes that the practical syllogism reveals "an order that is there whenever actions are done with intentions" (§42). Developing this point we can understand a particular syllogism as playing a role analogous to our 'reading-rule' above. Consider the following example from *Intention*:

Vitamin X is good for all men over 60
Pigs' tripes are full of vitamin X
I'm a man over 60
Here's some pigs' tripes

Despite the fact that it involves an idea of what is good, the first premise of this syllogism does not by itself represent the end pursued—it simply states a general fact. Its role in specifying an end, and showing the good of an action relative it, both become apparent once we know the conclusion of the syllogism, i.e. whether this reasoning terminates with e.g.

preparing the pig's tripes, or throwing them in the bin. Supposing the former, we might see the syllogism as e.g. showing the point of preparing pig's tripes relative to the end of preparing a healthy dinner. This would be to apply the syllogism as part of a description or explanation of a particular person's activity, just as we earlier paired the rule with a reading explanation. In this case, it could be paired with an action-explanation:

> S is preparing the pig's tripes because he wants to eat a healthy dinner

To understand the syllogism we must see the work the premises are called on to do in showing the good or use of an action for which they set out grounds—and to see that we need to know what action was grounded by this reasoning, and what end it served. The reasoning will make sense to us if we can see how, together, these premises do show (or might be thought to show) the good of that action relative to that end, and it will puzzle us if those relations remain opaque.

Of course, the syllogism does all of this without itself explicitly representing any end as to-be-pursued, or any action as to-be-done. Although the syllogism is in a sense 'organized' by its purpose of showing what good or use an action is, relative to a particular end, it does not include either a specification of that end, nor a description of the concluding act as good, as premise or conclusion. A real example of such a syllogism would do its work relative to a particular occasion, in which it might be used to represent the grounds of some particular act of preparing pig's tripes. There the syllogism could be said to show the good of that act, relative to a given end, and one would understand the syllogism if one could see the premises working together to do this. That means understanding how, together, they show the point of the act they purport to ground, relative to the end that is being pursued.

Although, as we have seen, a practical syllogism does not explicitly represent an action as 'to-be-done', its playing the role of practical reasoning consists in its premises showing the point of some particular action, given some further end—and, in that sense, in showing why that action is to-be-done. In pairing such a syllogism with particular action explanations, we bring out the fact that the act represented in those explanations is based on an understanding of the normativity expressed in the syllogism. Thus our explanation,

S is preparing the pig's tripes because he wants to eat a healthy dinner,[4]

represents the agent's act as teleologically-related to his goal of eating a healthy dinner, and the syllogism lays out grounds for thinking it such. That the agent is acting from an understanding of this normativity could be shown by the explanations or justifications he might provide for his action (i.e. citing grounds from the syllogism), and also the way he reacts to certain things, e.g. learning that pig's tripes are actually detrimental to men's health.

Normativists develop this line of thought further, by claiming that we should understand 'acting for a reason' as a kind of 'practical inference'. Here is Marcus on this point:

> The rational 'because', in the first instance, links the elements of theoretical or practical inference. In central theoretical cases, a thinker extends the status 'to be believed' from one or more propositions to another proposition. In practical cases, an agent extends the status 'to be done' from one action to another. The 'because' in a rational explanation refers to this inferential nexus. To say S was ϕing because q is to say that S made a practical inference, the conclusion of which *was* her action, and which she could express by saying "I am ϕing because q" or "I'll ϕ because q". This is the way rational explanations explain and the kind of causation they postulate. [24, 233-4]

On the theoretical side, Marcus' claim is that explanations of the form "S knows p because she knows q" represent an inferential connection between the specified contents,

[4] The strategy we have deployed would suggest writing this explanation as,

S is preparing the pig's tripes because$_I$ he wants to eat a healthy dinner,

where the subscript " $_I$" marks the fact that the acts are (understood by the agent as) teleologically-related to each other, a relation whose basis could be laid out in some practical syllogism.

and describe exercises of a rational ability “to represent a known fact as a proposition whose to-be-believedness establishes the to-be-believedness of another proposition, which represents a further, thereby known fact” [24, 15]. Here the ability involves representing particular propositions as having the normative status ‘to be believed’ on the basis of further propositions represented as having the same status. Acts of this ability *are identical with* the causal relations underpinning explanations of this form, and the result of these acts are states that involve “a representing of a relation between doxastic deontic facts about propositions and at the same time also a representing of why one believes what one does” [24, 31]. On the practical side, explanations of the form “S is ϕing because she is ψing” also represent an inferential connection (this time between actions), and describe exercises of a rational ability having to do with the representation of a different normative status: “To say that someone is ϕing because she is ψing is to say that she represents the to-be-doneness of ϕing as a consequence of the to-be-doneness of ψing” [24, 107]. Here it is actions that are represented as having the normative status ‘to be done’. The idea of an ‘act of inference’ is then supposed to capture what it is for a person to knowingly act on the basis of particular considerations. For the normativist, it is to represent their action as having the status to-be-done because of those considerations.

7.2.2 Form and Grammar

These ideas can be connected to our discussion of the results of Anscombe’s investigation into the concept of ‘intentional action’ in §3.1. There we saw that her key claim was that “the term ‘intentional’ has reference to a *form* of description of events” and that “[w]hat is essential to this form is displayed by the results of our enquiries into the question ‘Why?”’. The normativist strategy is essentially an attempt to develop this line of thought. Our key ideas (a) and (b) above can be put by saying that the agent knows her acts under descriptions of the form Anscombe identifies, and that this is reflected in the way she gives expression to those acts, and the further things she knows it makes sense to say of them.

The normativist tries to capture these ideas by equating the acts themselves with representings of normative status that belong with this form. To see this, think about how the normativist would describe our example from the earlier chapter:

> James is sliding on the ice because he is retrieving his coat

becomes

> James is representing sliding on the ice as to-be-done because fetching his coat is to-be-done.

What this aims to capture is the idea that James' action is a pursuit of the relevant end, and understood by him as such. We capture the idea that a description of an act is of a particular form, and the act itself known by the agent under a description of that form, by identifying the act with a representing of a normative status associated with that form. Here the form in question is illustrated in the order revealed by Anscombe's question 'Why?', and by fragments of practical reasoning, both of which show why an action was 'to-be-done' relative to some further end. The agent was acting from an understanding of these normative relations, and knows his action under a description of this form (and so *as* an intentional action subject to the relevant sense of the question 'Why?', etc.). The normativist aims to capture all of this by identifying the act itself with a representing of this normative status.

There are thus two key ideas behind the normativist strategy:

1. Knowing an action under a description of a particular form (or a description that has a particular grammar) involves some kind of representing of that form; and

2. that representing can be explained in terms of representing the action as having a normative status that belongs to the form.

Together these provide the basis for the normativist's account of reason-giving explanations and the acts they describe.

7.3 Rational Abilities and Rational Explanation

As we saw in §7.1, the application of the normativist strategy to reason-giving explanations appears to provide substantive responses to our philosophical explanations (MQ) and (EQ). This rested on the claim that

1. acts of 'rational abilities' constitute a distinctive form of 'rational causation' represented in reason-giving explanations

2. such acts are representings of normative status .

Point (1) appears to provide an answer to (MQ), since it suggests that *what it is to act for a reason* is *to be an instance of rational causation.* Point (2) appears to answer (EQ), since it suggests that *we have first-personal knowledge of what we are doing and why* because *our actions are identical to representings that express our understanding of them.* As such, the normativist strategy appears to provide the basis for a broad and powerful account of our rational capacities, since it seems to explain both *in virtue of what* something counts as an instance of the act of such a capacity, and *how it is* that we come to have knowledge of such acts.

However, in the rest of this chapter I aim to show that this impression is illusory since, on closer inspection, the characterization of the acts of rational abilities that is at the heart of the normativist's account poses some fundamental problems. In this section, I shall argue that the nature of these 'rational abilities' is either a mystery in need of explanation, or the normativist's talk of 'rational abilities' and 'representings of normative status' are nothing but a way of describing the acts of our familiar practical abilities. Either way, the normativist's claim to provide a substantive response to our philosophical questions is undermined. For if the nature of these abilities remains unexplained, then they cannot provide the basis for a resonse to (MQ) and (EQ); and if they are merely our familiar practical abilities, then the normativist's response to (MQ) and (EQ) will turn out to be circular and tautologous. In §7.4 I go on to argue that the normativist strategy does not give us the kind of reflective perspective we need to attain philosophical clarity, since identifying the acts of our familiar practical abilities with representings of normative status obscures, rather than clarifies, the nature of those abilities.

7.3.1 The Normativist's Dilemma

The dilemma I shall be concerned with applies to the normativist strategy in general, but takes on a specific form when applied to the account of reason-giving explanations. This dilemma will become apparent as we look more closely at the specific normativist claims that purport to provide substantive answers to our philosophical questions:

1. acts of 'rational abilities' constitute a distinctive form of 'rational causation' represented in reason-giving explanations
2. such acts are representings of normative status

An assessment of the normativist strategy therefore depends on the extent to which we can make sense of the relevant ability and its acts. The problem that arises is that each of these two claims can be understood in two ways, one of which makes it utterly mysterious, while the other undermines its explanatory role. So it will turn out that either:

a. the idea of a 'rational ability' is mysterious, or

b. it is no more than a way of describing our familiar practical abilities.

Either option undermines the idea that a presentation of this 'ability' serves to characterize a distinctive form of causation that shows what it is to act for a reason – i.e. the idea that it provides a substantive response to (MQ). Likewise, it will turn out that either:

c. the idea that an act of that 'rational ability' is a representing of normative status is mysterious, or

d. it is no more than a way of saying that the bearer of such abilities can just say what she is doing and why.

Once again, either option undermines the idea that a presentation of the 'acts' of this 'ability' will serve to explain how it is we come to have first-personal knowledge of those acts – i.e. the idea that it provides a substantive response to (EQ).

To make matters worse, if we do understand the normativist strategy as aiming at (b) and (d), then the specifically normativist elements of that strategy serve to obscure the character of the rational abilities it aims to describe.

7.3.2 Characterizing the Normativist 'Rational Ability'

Eric Marcus' book *Rational Causation* [24] aims to provide an account of rational explanations of action and belief that takes the form we have just sketched. His goal is to describe a "distinctive sort of causation, which is neither efficient, physical causation nor some sort of telekinetic analogue" [24, 2]. As above, the real connections that make up this form of causation are acts of 'rational abilities' that involve the representing of normative relations.[5]

Marcus takes his lead from Anscombe's work by focusing on action-explanations of the form "*S* is doing A because she is doing B". These explanations represent their targets as parts of a teleologically-articulated whole, i.e. some particular action is represented as a part of some further action. This part/whole relation is what Anscombe articulated in her discussion of the A-D order, and if she is right in her claim that this is an "order that is there whenever actions are done with intentions", then explainability in terms of the relations characterizing that order is intrinsic to our understanding of what it is to be an intentional action.

As in the sketches above, Marcus bases his account on the idea of acts that explicitly

[5] Marcus argues that these abilities are a species of disposition ("albeit one that is distinctive in crucial respects"[24, 50]), meaning that the explanations that Marcus focuses on are a species of *dispositional explanation.*

represent the normative-explanatory relations that characterize this order. Here is Marcus' description of the 'rational ability' involved in action-explanation:

> Rational action-explanations appeal to the exercise of an ability to do what is to be done because something else is to be done. To say that someone is ϕing because she is ψing is to say that she represents the to-be-doneness of ϕing as a consequence of the to-be-doneness of ψing. When this ability is exercised succesfully, the agent's representing the relevant action as to be done is also her performing them. [24, 107]

If we understand Marcus to be providing a substantive response to (MQ), we must see him as asserting that something counts as an instance of acting for a reason *in virtue of* being an instance of this kind of causation, which is to say *in virtue of* being an act of this 'rational ability'. Understanding these acts and the abilities they manifest therefore becomes essential to the account: for they are what promises a substantial answer to our philosophical questions.

A Distinct Ability?

An assessment of Marcus' account thus depends on our understanding the nature of these abilities, and the role they are supposed to play in intentional action. We can take our lead here from a comment Anscombe makes in the course of her investigation:

> When we ordinarily speak of practical knowledge we have in mind a certain sort of general capacity in a particular field; but if we hear of a capacity, it is reasonable to ask what constitutes an exercise of it. E.g. if my knowledge of the alphabet by rote is a capacity, this capacity is exercised when I repeat these noises, starting at any letter. §48

Faced with Marcus' characterization of his 'rational abilities', it is reasonable to ask what constitutes their exercise. And even before we consider the normativist description of those acts as ‘acts of representing’, it could seem that the exercise of these rational abilities must be very different from the acts of abilities in the more familiar sense. For instance, in introducing the idea of explanations that appeal to the exercise of an ability, Marcus uses the example of an ability to do a handstand,[6] which is made manifest in an act of ⟨doing a handstand⟩, i.e. by a familiar instance of action. A direct analogy with this case would suggest that an ability to do what is to be done because something else is to be done will be made manifest, not by a particular act of ⟨doing A⟩, but rather by an act of ⟨doing A because one is doing B⟩. But here we must ask: what kind of *act* is that?

The problem is that if we treat the acts of the ‘rational ability’ as genuinely distinct, akin to the act of doing a handstand, they become mysterious. The puzzle that arises here is a version of the one I described in §4.3.1. Marcus introduces the notion of a ‘rational ability’ by way of a comparison with other kinds of disposition and ability, focusing in particular on our everyday practical abilities such as the capacity to do a handstand. However, the ‘ability’ he describes turns out to ultimately seem to have a fundamentally different character, which is not illuminated by the other cases that were used to introduce it. In effect, Marcus’ ability is ‘the ability to act for a reason’ or ‘the ability to draw practical inferences’—but if it makes sense to talk of such an ability, it will be in a different sense from the way we talk about familiar practical abilities. For instance, unlike Marcus’ own example of a practical ability—doing a handstand—his rational ability is not something that I can practice, and if I can be said to learn how to do it, or to teach it to others, it will be in a different sense from the way I might learn or teach abilities of the more familiar sort.

Perhaps the clearest way of bringing out this difference is to show that, if they are distinct from acts of our practical abilities, acts of this ‘rational ability’ cannot themselves be subject to action-explanations, nor indeed reason-giving explanations of any sort. A particular act

[6] Marcus calls it a ‘rational ability’, but for clarity’s sake, I shall reserve the term ‘rational ability’ for the special kind of ‘ability’ that play a central role in Marcus’ account, and call these more familiar abilities ‘practical abilities’ (though there will turn out that there *is* a sense in which the practical abilities, and the rational ability, are one and the same).

of James' practical ability to do a handstand can be explained by an action-explanation of the familiar sort: James is doing a handstand because he is practicing yoga. But what about an act of Marcus' rational ability? Its act is not e.g. ⟨doing a handstand⟩, but rather ⟨doing a handstand to practice yoga⟩. We can, perhaps, envisage asking questions that included all of that in their scope, i.e.,

> Why did James do a handstand to practice yoga?

There are possible uses for a question phrased in this way: it might be a way of asking what role handstands have in yoga, or why James chose this way of practicing yoga today, instead of working on his downward-facing dog. But neither of those questions is directed at the 'act' of our rational ability: they rather ask about the more specific acts described in the *explanans* and *explanandum*, i.e. doing a handstand and practising yoga. To ask for an explanation of the act of the 'rational ability' would be like asking "Why did James act for a reason?" or "Why did James draw a practical inference?"—phrases which have been given no use.

This problem is made particularly acute by the specific application of the normativist strategy to action-explanation. As we saw above, this involved the claim that what is represented in such explanations are acts of practical inference, and that such acts involved representing an action as having the status to-be-done, relative to some further end, because of the considerations laid out in the syllogism. One natural way of hearing the normativist's talk of 'representings of normative status' is on the model of e.g. a capacity to represent things as coloured. This capacity manifests in various ways—perhaps in the fact that I have colour-vision, and certainly in the fact that I can talk about objects as coloured, etc. But its basic form involves representing an object as having a particular property. Marcus' describes his rational abilities in similar terms "[h]umans represent propositions to themselves as to be believed and represent actions to themselves as to be done". This makes is sound as though both capacities involve contentful acts of representing: just as I can e.g. represent a bottle as blue, I can also represent an action as to-be-done, or a proposition as to-be-believed.

But this cannot be what Marcus means. If such acts are both acts of practical inference, *and* contentful representings of normative relations between propositions, then we ought to include the content of that representing among the premises of our practical syllogisms. For instance, suppose a syllogism concluded in an act of inference that was a representing of ⟨doing A as to-be-done because doing B was to-be-done⟩. If this act is a contentful judgement about normative relations between actions, then it belongs among the premises of the syllogism, since it is among the grounds for ⟨doing A⟩. But of course, were we to include this premise, we would be immediately find ourselves at the beginning of a regress. For drawing the conclusion that included this new premise would require a further act of inference, whose content must be added as another premise—requiring another act of inference, etc.

A Way of Describing Practical Abilities?

Thus, if we treat the normativist's 'rational ability' as something distinct from the particular acts described in the *explanans* and *explanandum* of a reason-giving explanation, the character of that ability becomes mysterious and threatens to undermine the whole account: it is fundamentally different from other practical abilities, but no positive account of that difference has been provided. Our other option is to treat this aspect of the normativist strategy as providing us with a way of describing particular rational abilities. To show what I mean, I will temporarily set aside the explicitly normativist characterization of 'rational abilities', i.e. as abilities whose acts consist in *representings* of normative status, and instead bring out a way in which one aspect of the normativist strategy is analaogus to my own approach.

To see what I mean here, it will be helpful to return to the more specific example provided by 'reading'. Sticking for the moment to a semi-normativist characterization of rational abilities (i.e. one that does not describe them as acts of representing), we could say that the ability whose 'acts' were depicted in 'reading explanations' was "an ability to say what is to-be-said because$_R$ such-and-such is written on the page". If we think of this as a rather roundabout way of describing *an ability to read*, then we have a model for how we are to understand "an ability to do what is to-be-done because something else is to-be-done". For the acts of this ability would be nothing other than the agent's acts of reading.

This would mean that this aspect of the normativist's strategy was no different from our introduction of reading-explanations as a way of characterizing the ability to read. What such explanations—or this way of characterizing the ability to read—make explicit is that acts of reading intrinsically involve a certain kind of explanatory relation. That is why the idea of a 'reading explanation' (or this way of describing the ability) seems artificial: they articulate an explanatory unity that we would normally represent as a whole, i.e.. by describing a person as reading.

If we understand Marcus' 'rational ability' in the same way, then its acts must be *nothing but* the acts of particular practical abilities, i.e. actions in the familiar sense. When, in doing a handstand as part of his yoga practice, James exercises his ability to do what is to be done because something else is to be done, he does so *by doing a handstand.* This can be seen as an 'act' of the rational ability in two senses. First, if he is successful, his ⟨doing a handstand as part of his yoga practice⟩ (and thus his doing what is to be done because something else is to be done) will involve no further action besides this handstand. Second, because doing a handstand is *itself* an instance of doing what is to be done because something else is to be done, since doing a handstand is a complex act. (To do a handstand, James must e.g. establish a firm base with his hands, kick one leg up and then the other, etc.)[7]

There is, then, a sense in which *any* practical ability that can be exercised in intentional actions can be described as an ability 'to do what is to be done because something else is to be done': first, because the acts of that ability can be parts of a teologically-articulated

[7] A similar point can be made with our other examples as well. In each case, when the agent exercises the relevant 'rational ability', he does so by exercising a practical ability: saying "kæt", writing "2, 4, 6, . . . ", turning left, etc. Describing these as 'acts' of 'rational abilities' marks the action as belonging to a particular kind: saying "kæt" as an act of reading, writing "2, 4, 6, . . . " as an act of rule-following, turning left as an act of sign-following, and so on. And by the same token, describing James' inversion as an 'act' of a rational capacity marks the fact that it was an intentional action, i.e. that James *intended* to invert himself, and did it by doing a handstand.

whole; and second, because in most (perhaps all) cases, the act itself will have a complex teleologically-articulated structure that will involve various sub-acts. In this sense, the ability to do a handstand *is* an ability to do what is to be done because something else is to be done, because *qua* intentional actions its acts can be done for the sake of realizing some further end, and because those same acts can be divided into sub-acts that have the same form. Describing an intentional action is then describing something that intrinsically involves a particular explanatory structure *qua* the kind of thing that it is.

Characterizing the act of a particular 'practical ability' (or, more generally, any attempt to realize an intention) in terms of this semi-normativist 'rational ability' forces us to adopt a particular perspective on that act: it brings out the fact that it intrinsically involves the A-D order that Anscombe articulated in *Intention*, with various sub-acts unified insofar as they are done for the sake of a particular end. Anscombe expresses this same point in the continuation of the quote above, i.e. in her characterization of what we mean when we think of 'practical knowledge' as a 'certain sort of general capacity in a particular field':

> In the case of practical knowledge the exercise of the capacity is nothing but the doing or supervising of the operations of which a man has practical knowledge; but this is not *just* the coming about of certain effects, like my recitation of the alphabet or bits of it, for what he effects is formally characterised as subject to our question 'Why?' whose application displays the A-D order we discovered. §48

To describe an action as the 'act' of a 'rational ability' in this semi-normativist sense is to say nothing more than that it is formally characterized as subject to Anscombe's question 'Why?', i.e. that it intrinsically involves a particular kind of normative-explanatory connection.

A Substantive Response?

However, if we understand the normativist account as providing a way of describing our familiar practical abilities, we undermine its claim to provide a substantive response to (MQ). For now it turns out that, rather than pointing to a distinctive kind of ability that was going to tell us *in virtue of what* something counted as an instance of acting for a reason, we instead find out that what we have is a way of describing such acts that makes explicit the kind of acts they are.

In this sense, the aspect of the normativist strategy I have described is analogous to our use of subscripts in writing out particular forms of explanation. For instance, the subscript $_R$ indicates that the following explanation represents its target as an act of reading:

S said "..." because$_R$ "_____" is written on the page.

By the same token, the subscript $_I$ could be used to indicate that the following explanation represents its target as the intentional pursuit of an end:

S is doing A because$_I$ she is doing B.

These subscripts made explicit something that was already comprehensible in our use of these explanations, i.e. what kind of act they represent. For it is a feature of this kind of act that it is subject to this kind of explanation. But making that explicit was not an effort to say *in virtue of what* something counted as an instance of that kind of act. Indeed, using these subscripts in such an account would amount to tautology: something counts as an instance of the kind of act represented in this explanation in virtue of being the kind of act represented in this explanation.

If we understand the basic idea behind the normativist's description of 'rational abilities' as a different way of capturing the idea that the acts of our familiar practical abilities are subject to particular forms of explanation *qua* the kind of act that they are, then the 'abilities' and 'acts' they describe lose their air of mystery: they are nothing but the acts of our familiar practical abilities. But at the same time, this also removes the impression that the normativist account is providing a substantive response to the question ‘what is it to act for a reason?’. For on this interpretation, the normativist’s response would amount to: to act for a reason is for one’s act to be subject to reason-giving explanations.

7.3.3 Acts as Representings

The full normativist strategy involves an additional element that takes it beyond my use of subscripts. Whereas I have relied on these subscripts to capture the idea that the agent acts from an understanding of the relevant form of normativity, as in

> S said “. . . ” because$_R$ “_____” is written on the page,

the normativist tries to capture that idea by describing the agent as representing the relevant normative relations, e.g.

> “. . . ” is to-be-said because “_____” is written on the page.

This properly normativist characterization of these ‘rational abilities’ complicates this story, since it involves identifying actions in the familiar sense with acts of representing normative status and relations. Indeed, as Marcus makes clear, what makes them ‘rational’ is that they “involve representational powers that animals (as far as we know) lack”:

> Humans represent propositions to themselves as to be believed and represent actions to themselves as to be done. Non-human animals do not. Hence the etiology of their behaviour—for all its similarities to human thought and behaviour—must be understood differently. [24, 12]

Up to now I have been relying on the semi-normativist characterization of Marcus’ rational ability: an ability to do what is to be done because something else is to be done. My suggestion was that we understand this as a way of describing the acts of familiar abilities, one that emphasised an explanatory structure that was intrinsic to them *qua* acts of practical abilities. That meant that we could see the 'act' of this ability as nothing but an action of the familiar sort, e.g. a handstand. But the fully normativist account introduces a wrinkle. For now we must characterize the rational ability, not as an ability to do what is to be done because something else is to be done, but as an ability to *represent* something as to-be-done because something else is to-be-done. If the account of rational abilities provided so far is to hold up, we need to find a way to make sense of this idea that actions in the familiar sense are identical with acts of representing.

As we have seen, Marcus frames this aspect of the normativist's account as part of an explanation of why it is that rational agents are capable of answering Anscombe's question 'Why?', i.e. of saying what it is they are doing, and what their grounds are for doing it:

> [J]ust as "I believe that p because q" can be the verbal incarnation of believing-for-a-reason, "I am ϕing because I am ψing" can be the verbal expression of acting-for-a-reason. Just as believing-for-a-reason is expressible because it is the believer's representing one proposition as ⟨to-be-believed as a consequence of another proposition's to-be-believedness⟩, acting-for-a-reason is expressible because it is the agent's representing an action as ⟨to-be-done as a consequence of another action's to-be-done-ness⟩. [24, 72]

If we understand Marcus to be providing a substantive response to (EQ) here, we must see this idea of a 'representing' as explaining *how it is* we come to have first-personal knowledge of our reasons for acting. This would be to understand it on the model of the other accounts we have considered: as identifying the causality involved with a representation that grounds the agent's knowledge of that causality. In §6.3, we saw that Setiya appeals to a self-referential desire-like belief, SR, which is both the cause of the action, and the grounds for the agent's knowledge of that action. If we understand Marcus' talk of 'representing' in the same way, then we must understand the 'representing' as both the cause of the action, and the grounds for the agent's knowledge of that action. For if these are identical then we can understand how the agent has knowledge of her action just insofar as she acts in that way. The question then is if we can understand the relevant notion of 'an act of representing' as providing grounds for knowledge.

Representing an Action as To Be Done

To see why this is difficult, it is helpful to compare Marcus' acts with something we might also call an 'act of representing an action as ⟨to-be-done as a consequence of another action's to-be-done-ness⟩'. Suppose, for instance, that I was planning a trip, and part of that planning involved writing a 'to do' list of tasks I needed to complete before I left, with each task having further sub-tasks required for its completion. We could then understand an 'act' of the sort that Marcus describes as that of writing a particular task somewhere on this list: e.g. if I write 'buy phrasebook' under the heading 'learn French', then I am representing buying a phrasebook as to-be-done as a consequence of the fact that learning French is to-be-done. Such a list could even play a role in explaining *how I knew* that I had to do a particular task for the sake of the trip. If you asked me why I needed to buy sunscreen ('Don't we have some already?'), I can point to the fact that it is written on the to-do list to justify my knowledge that it needs to be done.[8]

Clearly this cannot be what Marcus has in mind, since 'representing an action as to-be-done because another action is to-be-done' in this sense has to be a distinct act that is separate from the action that is represented: after I write 'buy a phrasebook' on my list, I still have to go out and buy it.[9] This points to precisely what needs explaining in Marcus' account. Given the arguments presented above, it seems that the best way to understand this 'act of representing' is to see it as identical to the act of the practical ability it represents. To take our earlier examples,

[8] For an application to another of our examples, imagine someone writing down a table of rules for how to read a strange script. Writing a description of a particular sound next to a particular sign would be an act of representing that sound as 'to-be-said' in response to that sign.

[9] The same is true of simply 'representing an action as to-be-done'.

⟨representing doing a handstand as to-be-done because doing yoga is to-be-done⟩

is just

doing a handstand because one is doing yoga,

i.e. doing a handstand as a part of one's yoga practice. By the same token,

⟨representing "kæt" as to-be-said because "C-A-T" is written on the page⟩

is just

saying "kæt" because$_R$ "C-A-T" is written on the page,

i.e. reading "cat". But in each case, to describe the relevant act (doing a handstand, reading "CAT") as an 'act of representing normative status' threatens to render that act mysterious. For clearly these acts are not acts of *producing representations*, as was the act of writing something on our to-do list.

The solution lies in seeing Marcus as describing a *capacity* that the agent has *insofar* as she is acting in this way. To talk of a 'representing' in this sense is to talk of a capacity to (among other things) produce a certain kind of representation: e.g. a capacity to *express* my reasons for action by saying "I am doing a handstand because I am doing yoga". The acts of such a capacity give expression to the understanding from which the agent acts on a particular occasion, i.e. her knowledge of what she is doing and why.[10] As before, this is given a normativist twist: rather than talking of a 'representing' of the form 'I am doing A because I am doing B', the normativist's 'representing' is of the form 'Doing A is to-be-done because doing B is to-be-done'. Once again, this can be seen as an attempt to make explicit the idea that the act is known *as* an intentional action, and thus as intrinsically involving

[10] Since acts of this capacity gives expression to the agent's understanding of her action, we might also think of it as involved in everything that manifests her intelligent realisation of that action. However the possibility of expressing this understanding in language is an essential mark of the way in which it differs from the understanding of a non-rational animal.

the relevant form of normativity. So to say that the agent is ⟨representing doing a handstand as to-be-done because doing yoga is to-be-done⟩ is to recognize that, if she says "I'm doing a handstand because I'm doing yoga", her expression of her reason for acting describes her act in terms of that normativity.

If we understand Marcus' 'representings' in this sense, then they are identical with the relevant act, since the agent has that capacity just insofar as she acts in that way. Here too, then, the normativist's strategy turns out to be a *way of describing* the acts of our familiar practical abilities, rather than the description of an act of a distinctive ability. This makes the normativist's main point analogous to the way in which I have characterized the various practical abilities I have been concerned with. For instance, I said that it was a mark of the ability to read that its bearers could respond to such questions as 'why did you say such-and-such?' or 'were you reading or quoting from memory?'. But as before, in describing this feature of the ability I did not purport to be providing any account of how its bearers came to be able to answer these questions. Any such response would be just as tautologous as the response to (MQ) suggested above, i.e. we can answer questions about the acts of this ability because it is a mark of the ability that we can answer such questions.

7.3.4 Applying The Dilemma

This helps us apply the dilemmas described in 7.3.1 to the normativist's view. Remember, the normativist seemed to provide a substantive response to our philosophical question (MQ) because they claimed that

- acts of 'rational abilities' constitute a distinctive form of 'rational causation' represented in reason-giving explanations.

However, it now turns out that the 'acts' involved in such causality are either a mystery in need of further explanation, or they are merely a way of describing the familiar acts of our practical capacities. Thus, the normativist was either saying

a. to talk of ‘rational abilities’ is to describe a distinctive kind of ability, and it is in virtue of being an act of that ability that something counts as an instance of acting for a reason; or,

b. to talk of ‘rational abilities’ is a way of describing our ordinary practical abilities that makes the character of their acts explicit

(a) promised to provide a substantive response to (MQ), but the content of that response was mysterious, since the character of the relevant ability and its acts was unexplained. In contrast, (b) did not purport to provide a substantive response to (MQ), since any response would have the tautologous form ’to act for a reason is for one’s act to be subject to reason-giving explanations’. This means that the most plausible interpretation of the normativist’s strategy undermines its claims to provide substantive answers to (MQ). For it turns out that the normativist’s ‘answers’ are circular and tautologous.

In the same way, the normativist seemed to provide a substantive response to (EQ) because they claimed that

- acts of rational abilities are representings of normative status

However, it again turns out that the ‘representings’ involved in our being able to say what we are doing and why are either a mystery in need of further explanation, or they are merely a way of describing our familiar practical abilities. Here, the normativist was either saying

c. to identify the act with a ‘representing’ is to say that the act is, in some sense, a representation, and that this representation grounds the agent’s knowledge of her act; or,

d. to identify the act with a ‘representing’ is a way of describing our capacity to say what we are doing and why just insofar as we act

(c) promised to provide a substantive response to (EQ), because it promised to specify a ground that could explain our knowledge, and to show how we had that ground just insofar as we acted; but the content of this repsonse was mysterious, since it was unclear how the act could *be* a representation in this sense. In contrast, (d) did not purport to provide a substantive response to (EQ), since it simply describes the phenomena that stands in need of explanation: i.e. it describes, rather than explains, our capacity to say what we are doing and why. Once again, any response would have the tautlogous form ‘we can answer questions about the acts of our abilities because it is a mark of those abilities that we can answer such questions’.

This shows that there was some truth to the objection that we considered in §2.3.3: that normativism seems to collapse into primitivism. For we can now see that the claims the normativist provides in response to (MQ) ultimately have the form of tautologies: they tell us such things as,

> To act for a reason is for one’s act to be explainable by a reason-giving explanation,

which amounts to saying,

> To act for a reason is for one’s act to be an instance of acting for a reason.

For it is the normativist’s attempts to emphasize the aspects of his view that distinguished it from primitivism that led to the accusation that the view was mysterious. The difference between normativism and primitivism supposedly lay in the fact that normativism described a distinctive form of causation that consisted in special acts on the part of rational subjects. But it now turns out that, on further investigation, attempts to characterize this causation and the acts involved in it collapse into the same tautologies put forward by the primitivist.

7.4 Assessing the Normativist Strategy

The normativist's dilemma can be summarized as follows. The normativist appears to be providing a substantive account that states that what differentiates reason-giving explanations from other kinds of explanation is that the former represent instances of 'rational causation'. However, attempts to provide a positive characterization of this 'rational causation' render it mysterious; and attempts to domesticate it collapse into the circular and tautologous claims that characterized primitivism. Ultimately the normativist is making the baroque but empty claim that what differentiates reason-giving explanations from other kinds of explanations is that the former represent acts that are explainable by reason-giving explanations and explain them as such.

This emptiness can be traced back to the conception of (MQ) that I claimed was shared by all three of the accounts surveyed in §2.3. There I argued that these accounts sought a response to (MQ) that would cover *any* instance of reason-giving explanation; I further suggested that this would involve a description of what was immediately represented by those explanations that would show why it counted as an instance of acting for a reason. We can now see clearly what the primitivist has known all along: that the only descriptions of specific acts available at this level of generality and abstraction will take the form of primitivism's circular claims.

Of course, such claims are not false. But they hardly provide a helpful perspective on (MQ) and (EQ), or the puzzles that led to them. Moreover, the argument of §4 suggests that taking these abstract claims as our starting point for reflection on 'acting for a reason' will tend to engender philosophical confusions. This can be seen at two different points in the normativist's account, both related to their reliance on a single uniform notion of 'rational causation' as the common element in what is represented by all reason-giving explanations. The first place is in the original application of the normativist strategy: the identification of the acts explained with representings of normative status. To apply this point at the level of generality that the normativist desires, the relevant notion of an 'act of representing' has to be common to both what is described in the *explanandum* of explanations of the form,

> S believes that p because she believes that q,

and explanations of the form,

> S is doing A because she is doing B.

For though the 'object' (a proposition, an action) and 'content' ('to-be-believedness', 'to-be-doneness') differ, the basic kind of act—an 'act of representing'—is the same. However, this creates a problem for the normativist, since they want to *identify* that act with the acts described in the *explanandum.* But those 'acts' are quite different in character: one is an action, and one is a thought. An action is something that takes some time to unfold, and can be interrupted before it is completed; but a thought is not. This is reflected in the different grammars of the possible substitutions in our explanations. For instance, it makes sense for me to ask 'How long did it take you to do A?'; but not (in the same sense at least) 'How long did it take you to believe p?'. By simply identifying both acts with 'representings', the normativist threatens to elide this difference.

The second place where the normativist seems to fall into analogous difficulties is in their more specific claims about 'rational causation' in the practical case. Here Marcus is eager to differentiate his notion of 'rational causation' from contemporary notions of 'efficient causation'. Part of the difficulty here is that, in doing this, he seems to overstate the difference between his 'rational abilities' and other forms of dispositions.[11] But more urgently, in emphasizing this difference he is led to strongly identify 'rational causation' in the practical

[11] For instance, Marcus states that, in the case of physical dispositions like fragility, the triggering cause is 'external' to the manifestation of the disposition:

> On this model [of dispositional causation], a cause must be *external* to, since it is the cause of, the manifestation of the disposition. But [...] rational causation is *internal* to the manifestation of the ability under discussion here. The analogue to the tablet's dissolving [because it is soluble] is a causal connection between q and S's believing that p (when S knows q). On the view I defend, the relevant ability's manifestation is not the *effect* of a rational cause; rather, it is *itself* the rational-causal connection. [24, 7]

This seems to miss the fact that, with a dispositional explanation such as,

> The vase broke because$_D$ it was dropped.

sphere with the teleological explanations that are his primary focus. This means that he is strangely silent about the other forms of reason-giving explanation we surveyed in §3.3. This by itself points to a gap in his account – but if my earlier argument was right, among those further forms of reason-giving explanations will be some that represent the acts they describe as a reason-based response. Even if such explanations cannot be made to fit available theories of 'efficient causation', they will be closer to some generic notion of 'efficient causation' than they will to the kind of teleological cases that Marcus focuses on.

The problems that beset attempts to treat (MQ) and (EQ) at the level of generality and abstraction that characterize these accounts are therefore twofold:

1. The only descriptions available at this level of generality and abstraction are circular, and do not provide the basis for a substantive or informative account.

2. Those descriptions tend to attract various forms of philosophical confusion, leading us to both elide or over-emphasize the differences between specific cases.

Looking back, these are versions of the problems anticipated in §4.3: the normativist strategy both fails to articulate certain key differences, and overstates others.

7.5 Abilities and Explanation

Despite these criticisms, in this final section, I shall argue that one of the normativist's core ideas—once separated from the normativist strategy—can be combined with the Wittgensteinian material from chapters 3-5, to suggest an alternative approach to our questions. The core idea from Marcus that I wish to retain is that explanations such as,

one could understand the 'manifestation of the disposition' as encompassing the whole explanation, and not just what is described in the *explanandum.* Here too there is an 'internal relation' between cause and effect, since it matters that the former explain the latter *qua* manifestation of fragility. Seeing 'The vase shattered because it was dropped' as a dispositional explanation means seeing *explanans*, *explanandum*, and the explanatory relation between them, all in terms of the disposition to fragility. It is this disposition that makes the dependence of *explanandum* on *explanans* the appropriate sort for this to count as a example of fragility.

I1: James is doing a handstand because$_I$ he is practising yoga

should be understood as analogous to explanations such as,

D1: The vase broke because$_D$ it was dropped,

insofar as both depend on an appeal to a particular disposition or ability. Just as (D1) explains the act described in its *explanandum qua* manifestation of fragility, (I1) explains the act it describes *qua* intentional action. Indeed, we already applied a version of this insight in our treatment of 'reading-explanations' in chapter 5. For there we claimed that an explanation such as,

RE: S said "…" because$_R$ "_____" is written on the page,

explains the acts it describes *qua* acts of reading.

Part of the argument against reductivist accounts in chapter 6 involved showing how grasping such explanation involves seeing the whole in terms of some form normativity associated with the relevant subscript. For instance, to understand (D1) we must understand that it describes a manifestation of fragility, which means seeing it as related to judgements such as,

(**D2**) Objects that are fragile break when dropped.

(**D3**) The vase was dropped.

(**D4**) The vase is fragile.

which have the form,

(**D2**) Os that are **D** do ϕ in circumstance **C**.

(**D3**) S was in circumstance **C**.

(**D4**) S is **D**.

Such judgements give us part of the grammar of our thought and talk about objects with dispositions of this kind. For instance, if an object we know to be fragile does not break when dropped, it makes sense to enquire why not. This is the sense in which there is some minimal notion of normativity associated with the subscript in these dispositional explanations – with physical dispositions, we know that ϕ-ing is to-be-expected in circumstance C.

The subscript in D1 indicates how the phrase as a whole is being used: as a dispositional explanation. This is something that would usually be clear from the original context of use. Grasping that the phrase is being used in this way is essential to understanding it *qua* dispositional explanation. This point is related to the fact that the same phrase can be used to describe cases that are not manifestations of the relevant dispositions, i.e. deviant cases.[12]

Analogous points applied to our treatment of the subscript in our reading explanations. Understanding those explanations depended on seeing the whole as representing an act of reading. This meant seeing an explanation such as (RE) as belonging together with further judgements such as,

(**R1**) "_____" is to-be-read "...".

(**R2**) "_____" is what is written on the page.

(**R3**) S can read.

Understanding (RE), and with it the relevance of (R1-3), involved locating what is described in (RE) within a practice of reading. This took us beyond what was immediately represented in (RE), just as (D2-4) took us beyond what was immediately represented in (D1). But the circumstances involved in there being a practice of reading are nonetheless very different from the circumstances involved in there being fragile objects.[13]

[12] For instance, (D1) could be used to describe a case in which the vase was not fragile, but was instead rigged with a mechanism that would make it shatter when dropped.

[13] This in turn reflected in further differences. For instance unlike (D2), (R1) can be used to evaluate whether S's action was *correct*. As we saw in § 5.3, an act that can be evaluated in this way involves a different form of disposition from an act that is merely 'to-be-expected'.

As we saw in § 5.3, part of what is involved in characterizing this ability is setting it alongside such things as the system of norms that are involved in alphabetical writing. This idea is captured in the 'rule of reading' that is supposed to be expressed by (R1). Statements expressing these norms play an analogous role in characterizing *what it is* to be an ability of this sort as statements such as (D1) do in characterizing a disposition like fragility. They also determine our understanding of the causal relations involved in specific acts of reading – knowing that someone is reading, we look for that which they are reading from, and recognize it as playing a particular causal role. This is the idea captured by (R2). But we also saw how characterizing such an ability involved more than abstract systems of rules that give expression to the norms informing its acts – it involved various kinds of descriptions of the lives in which those norms found application.[14]

The key point in all of this is that seeing that (RE) is being used as a reading explanation—and, by extension seeing that what it describes is an act of reading—means seeing it as belonging together with judgements such as (R1-3) and the circumstances in which they would have application. That means, in particular, that a full understanding of (RE) depends on seeing it in terms of the rule expressed in (R1), and the recognitional judgement expressed in (R2). Full grasp of these judgements, and therefore full understanding of (RE), depends on knowing how to read. However, even someone who could not read the language would know that the fact that (RE) was being used a reading explanation meant that it must belong together with *some* judgements of the form (R1) and (R2).

On the normativist account described in §7, (I1) works by an analogous appeal to the normativist's rational ability,

A1: The ability to do what is to-be-done because something else is to-be-done,

[14] In §5.3.1 I suggested that this became clearer when we considered acts of reading in the full familiar sense, rather than the attenuated cases that Wittgenstein has us focus on. Here we are concerned with understanding what is written, and so with the idea that it *says* something. Thus (R1) becomes

(**R1**) "______" says that . . . .

whose acts are representings of one action as to-be-done because something else is to-be-done, and constitute instances of rational causation. However, it proved impossible to provide a positive characterization of this 'ability' at the level of abstraction and generality required by the normativist account.

In response, I suggested the following semi-normativist strategy: rather than treating (A1) as a description of a specific and unique rational ability, we should think of it as a way of describing our various practical abilities. Once we make that switch, we can think of the ability whose act is described in the *explanandum* of (I1)—i.e. the ability to do a handstand—as *itself* an ability to do what is to-be-done because something else it to-be-done. In understanding (I1), we grasp that the *explanandum* describes the act of an ability of this sort, and explains it *qua* such an act by showing that for the sake of which it is 'to-be-done'.

Described in this way—without the normativist's identification of these acts with representings of normative status—this core idea simply gives us a way of making explicit something that we already grasped in understanding this explanation. It makes explicit that the explanation as a whole depends on the idea of a normative relation between what is described in the *explanans* and the *explanandum*; and further that the act described in the *explanandum* is being explained *qua* subject to this normativity. In other words, the semi-normativist's descriptions serve to make explicit the fact that an explanation of the form,

S is doing A because she is doing B,

describes an intentional action that is teleologically-related to some further action. In grasping a specific version of such an explanation, e.g.,

I2: The gardener is moving his hand up and down because he is replenishing the water supply, I grasp that the actions described in the *explanans* and *explanandum* are intentional, and that the latter is being explained *qua* intentional action that is teleologically-related to the former.

As with fragility and the ability to read, part of what is involved in grasping an explanation such as (I2) is seeing it as belonging together with further judgements. In this case, rather than a simple rule these judgements could be arranged in a practical syllogism that shows the point of moving one's hand up and down, relative to the end of replenishing the water supply:

> To replenish the water supply, work the pump.
>
> To work the pump, move the handle up and down.
>
> This is the pump handle!

Understanding (I2) involves seeing it as belonging together with the judgements in this syllogism, which again involves circumstances that take us beyond what is immediately represented in (I2).[15] This can therefore be seen as showing us something of what it is to be an act of this sort, just as our earlier cases showed us something of what it is to be an act of reading or a manifestation of fragility.[16]

Here we see that part of what characterizes such actions is that our grounds for them can be laid out in a practical syllogism that shows the good or point of that action relative to some further end. The generic idea of such an ability is of something whose acts can be directed towards an end based on such grounds; and we understand particular acts by relating them to particular ends by way of those grounds. What it is to be an instance of

[15] As with reading, there is room here for different degrees of understanding. At a minimum, grasping that (I2) represents an intentional action depends on grasping that there must be *some* set of judgements that show the point of the action described in the *explanandum*, relative to the end described in the *explanans*.

[16] Note that these categories are not mutually exclusive – an act of reading can also be an intentional action, representable in an explanation such as,

RE2: S is reading the recipe to tell T what to do next.

Thus it belong to our ability to read that its acts can be intentional actions.

‘doing what is to-be-done because something else is to-be-done’, in the sense we are concerned with, is to be the kind of thing that belongs together with expressions of such grounds in things like syllogisms, rules, or other expressions of normativity, and descriptions of agents as directed towards the associated ends. In this sense, the idea of a practical syllogism plays a role analogous to judgements like (D2-4) or (R1-3) in characterizing the kind of disposition we are concerned with, and determining our understanding of its particular acts (including the causal relations involved in those acts).

The semi-normativist strategy aims to make all of this explicit by redescribing (I2) as representing,

NI2: The gardener doing what is to-be-done because pumping the water is to-be-done,

or redescribing (RE) as representing,

NRE: S saying what is to-be-said because “_____” is to-be-read,

These descriptions try to ‘build in’ the fact that the act belongs together with the syllogism or the rule, by describing the act itself in terms of the normativity expressed by that syllogism or that rule.

In this sense, these semi-normativist descriptions play a role analogous to my use of subscripts – where the normativist might have explanations such as (NI2) or (NRE), I would have,

S is doing A because$_I$ she is doing B,

or,

S is saying “. . . ” because$_R$ “_____” is written on the page.

As I have used these subscripts, their role is to indicate how these phrases are being used: as explanations of manifestations of fragility, acts of reading, or intentional actions. Part of what is involved in seeing that the phrase is being used in this way is seeing that the descriptions in these sentences belong together with e.g. the considerations in the premises of the practical syllogism, or the rules that characterize a practice of reading. For this is part of the grammar of these descriptions in this particular use.

By itself, the difference might seem trivial. But it reflects a more general approach to what is involved in making explicit the form (or grammar) of these explanations.

The 'semi-normativist' strategy works by trying to make the normativity involved in an act, and shown by the explanation, explicit all at once in a description of that act. This is reflected in the way the semi-normativist strategy characterizes the abilities that are manifested in those acts. Because the normativist reaches for an account that applies to *any* explanation of the form,

S is ϕing because$_R$ $\ldots$,

they seek characterizations with the maximum level of generality and abstraction. This is why the characterizations they provide of the relevant abilities are not only circular, but also peculiarly empty. To capture every explanation of the form 'S is doing A because she is doing B', the semi-normativist strategy talks of such explanations as appealing to abilities 'to do what is to-be-done because something else is to-be-done'. Once we understand this formulation, we know that it is telling us that the action it describes is intentional, and somehow teleologically-related to some further action. But it does not help us attain any clarity about what that comes to in a particular case.

Where normativism abstracts away from the particular case as quickly as possible, in order to reach the level of generality that it thinks (MQ) demands, the Wittgensteinian approaches we considered spend more time attending to the character of the particular acts described in such explanations. We have already seen this in the case of reading. Were a normativist to treat such cases, he might immediately abstract away to the kind of description suggested by the normativist strategy:

An ability to do what is to-be-done in response to such-and-such.

There is a sense in which this captures everything that we covered in our account of the ability to read: for acts of reading are indeed acts that inherit their normative status from that to which they are a response. But in occluding the circumstances of the act that were central to our account of the ability to read, and focusing instead on an abstract description of the specific act, the normativist has shown us nothing of what this form of 'acting for a reason' comes to.[17]

It might well be the case that this 'normativist' description could correctly describe all the further cases of reason-based responsiveness that we set alongside reading in §7.2.1 – e.g. cases such as following a sign, obeying an order, and so on. But whereas our Wittgensteinian approach helped us to see the similarities *and differences* between the various cases, the normativist description hides all of this from view. In this way, it invites the form of philosophical confusion described in §4.2 and §4.3.[18]

Conversely, when Anscombe introduces the A-D order, whose parts are describable in the normativist's terms, she is careful to show that we cannot see what this order amounts without keeping the circumstances of the act in view. For in the case she considers, the four descriptions of the act that constitute the A-D order are all descriptions of the same piece of behaviour! In this case, the only action that doing B consists in is doing A, but "more circumstances are required for A to be B than for A just to be A" (§26). Indeed, in her case we have "one action with four descriptions, each dependent on wider circumstances,

[17] Of course, the normativist does not actually attend to explanations of this form, focusing exclusively on teleological explanations and apparently assuming that these cover every practical case of 'acting for a reason'. On this, see more below.

[18] If we connect these points with the overview of Wittgenstein and Anscombe's work in chapters 3 and 4, we can think of the work done in clarifying the various forms of reason-giving explanation as showing us part of the grammar of descriptions of our abilities and their acts. Part of what characterizes our abilities and their acts is that they involve our understanding. This is shown by the fact that the descriptions that figure in the *explanans* of reason-giving explanations can also be used to express principles that are independent of any particular action. In other words, part of what it is to be an ability of this sort is to be the kind of thing whose acts can be explained by 'reasons for action' in the justificatory of evaluative sense – that is, rules, recipes, practical syllogisms, etc.

and each related to the next as description of means to end". Seeing what the A-D order amounts to in this particular case is therefore seeing these acts in their circumstances.

This point then extends to what it is to see the A-D order in other cases. For this will come to different things on different occasions. As with the example of reading, the point of Anscombe's example is to help us to see what this order comes to in one example, which we can then set alongside others to bring out the similarities and differences in what we would call 'doing A because one is doing B'. For example, we see (I2), and the syllogism that goes along with it, as in some ways analogous to (and, in this case, belonging together with),

I3: The gardener is replenishing the water supply because he is part of a scheme to bring down the government,

and the syllogism,

> The only way to bring down this government is to assassinate its leaders.
> We can do that by poisoning them.
> T laced the water supply to the house with poison.
> Etc.

But what 'doing A because one is doing B' comes to is something different in each case – to mark just one difference, if S is succesful in (I2) the the act described in the *explanandum* will be over at the same time as the one described in the *explanans*, whereas in (I3) it might take more time and work to realize the end described there.

Putting this in the normativist's terms, we might say that 'doing what is to-be-done because something else is to-be-done' also encompasses a variety of different cases, and it is in attaining a reflective perspective on this that we see what intentional action is. Even if the normativist's description is correct, it does not work as a response to our concerns because it hides all of this from view.

CHAPTER 8

CONCLUSION

In the previous chapter I suggested that normativism ultimately collapses into a form of primitivism. This can be traced back to the shared conception of (MQ) I described in §2.4, which suggested that any response must explain what differentiates *any* reason-giving explanation of the form,

RA: S ϕed because$_R$ _____,

from other explanations, including perhaps apparently identical explanations,

NR: S ϕed because _____,

that do not describe instances of 'reason-giving explanation'. Given this conception, our accounts were led to seek some common feature of every act that could be explained by an explanation of the form (RA).

The arguments of chapters 6 and 7 aimed to show that primitivism was right in its claim that the only common feature of all acts explainable by explanations of the form (RA) was that. . . they are explainable by explanations of the form (RA). Specifying *this* feature is hardly informative, and certainly doesn't seem to provide us with a helpful perspective on our original puzzles.

Does this mean that primitivism is our only option? In this conclusion, I shall argue that it is not. Indeed, as I suggested in §7.5, we already have everything we need to outline an alternative approach to our topic. However, adopting this approach means rejecting the conception of (MQ) and (EQ) shared by contemporary accounts, and the attendant ambition to explain 'what it is to act for a reason' all at once, via a uniform and substantive account of specific acts. Instead, we must attend to the details of particular cases to which we would apply the description 'acting for a reason', and allow the emerging clarity about their similarities and differences to constitute our answer to (MQ).

8.1 A Different Conception of (MQ)

As stated above, the contemporary accounts we have been concerned with all take (MQ) to demand a response that is:

1. maximally general, insofar as it covers any reason-giving explanation.
2. substantive, insofar as it shows why specific explanations belong in this category, and thus why the acts they describe count as instances of 'acting for a reason'.

In §2.4 I suggested that this would lead these accounts to abstract away from the character of the specific acts described in these explanations, and the circumstances in which they are used, and instead seek an general characterization of what was immediately represented in these explanations that would satisfy (1) and (2).

I have subsequently argued that the only descriptions that fit both criteria are those of the primitivist. But we anyway had grounds to be suspicious of this general approach based on the discussion in §3. For there we saw that the contemporary literature was including two different forms of explanation under the general heading of 'reason-giving explanation'. First, those whose *explanans* specified an end that the agent was pursuing, and represented the act described in the *explanandum* as teleologically-related to that end; and second, those whose *explanans* specified a ground for the action, and represented the act described in the *explanandum* as based on that ground. Furthermore, we saw that even at this level of abstraction, each of these forms of explanation could gather together a variety of different cases. For instance, some end-specifying explanations represent one finite action as teleologically-related to a further finite action; and within these, some represent the first action as a part of the second (so that when the first is complete, the second is not), whereas others represent the first action as a way of doing the second (so that when the first is complete, the second is too). Even this much reflection shows that the form of explanation that the normativist focuses on, S is doing A because she is doing B,

does not exhaust the category of 'end-specifying explanations', and admits of further specification in particular cases. Analogous points applied to the explanations that the reductivist and the primitivist focused on,

S is doing A because p,

which could be used to represent everything from the quite general knowledge involved in a craft or skill, to reactions to specific features of their immediate circumstances.

If instances of 'acting for a reason' include any act that could be subject to what we could call a 'reason-giving explanation', it will include all of these categories (and no doubt many more). That should already lead us to be suspect of the idea that we could provide a maximally-general account of all such explanations that was anything other that circular.

Although this undermines the pressure to provide a maximally-general account, it does not undermine the idea that we need a substantive response. For someone could admit that 'reason-giving explanation' encompasses a variety of related form of explanation, and still think that we need descriptions of what is immediately represented by each more specific category to show why it belongs in that category.

The arguments of §3.2 and §5 were supposed to relieve this pressure. They do this partly by showing that any description of what is immediately represented these explanations will either be circular, or fail to capture the fact that what is described is subject to an explanation of this sort. But they also aim to show that our recognizing something as a particular kind of explanation—or our recognizing an action as the kind of thing we would call 'acting for a reason'—has as much to do with the circumstances of that action as it does with anything that is immediately represented by its explanation.

In §7.5 I suggested this was related to the fact that reason-giving explanations in the practical sphere represent the acts of our practical abilities. Understanding *any* dispositional explanation involves more than what is immediately represented in the explanation itself –

for it involves recognizing that what is described is being explained *qua* manifestation of that disposition. But as we saw in our discussion of reductivism in §5.3 and §6.4, characterizing the specific dispositions and abilities whose acts are represented in reason-giving explanations requires seeing those explanations as working by an appeal to specific forms of normativity that characterize human life.

This was shown in some detail in the discussion of reading in §5. The upshot of that discussion was that recognizing something as an act of reading depended on seeing it together with circumstances that showed it to be such an act: that is, circumstances in which we could see it as part of a practice of reading and writing. Since it is in recognizing it as an act of reading that we recognize it as the kind of thing to which we would apply the description 'acting for a reason', it is these circumstances that show us what 'acting for a reason' comes to in this particular case.

Of course, this was just one kind of case – not every reason-giving explanation explains its target by immediately locating it in a practice. Nevertheless, as I argued in §7.5, any reason-giving explanation will depend on an appeal to *some* form of normativity, and represent the act it explains as based on the agent's understanding of this normativity. Seeing the act in terms of that normativity will involve circumstances beyond what is immediately represented by that explanation. This was why the normativist's attempt to capture this point by describing a generic ability to represent normative status seemed empty: it obscured those circumstances from view, and in doing so occluded the different forms of normativity involved in different reason-giving explanations.

The approach I have suggested is more piecemeal, beginning with attention to particular explanations and the acts they represent, and working up to a characterization of the abilities those acts manifest. This then provides a basis for comparison with other cases. As we saw, further reflection on 'what it is to be an act of reading' revealed various similarities and differences with other kinds of acts, including responding to a sign-post or a spoken order. These too were cases to which we would apply the description 'acting for a reason'—

and setting them alongside our discussion of reading emphasized particular aspects of what that came to in each case. However, we only concerned ourselves with cases that involved particular kinds of reason-based response to an immediate fact about the environment – in other words, to one very particular subset of the variety of cases encompassed under the general heading of 'reason-based explanation'. Other cases will involve striking differences as well as similarities. For instance, many ground-giving explanations do not explain their target by representing it as a response to some immediate feature of the environment. They might instead explain it by representing it as based on an understanding of the general principles of some craft or skill. An account of this form of reason-giving explanation would therefore aim to show what an act's being a manifestation of such knowledge comes to.

Nevertheless, I would suggest that clarity about our particular range of cases constitutes real progress towards answering (MQ). Of course, this is a different form of response from the one sought by contemporary accounts. Rather than trying to come up with a general and substantive characterization of 'acting for a reason', we look to particular cases to show us what 'acting for a reason' comes to in them. This does have the form of a response to (MQ) – i.e. to the question 'what is it to act for a reason'. But the content of that response is something like: to act for a reason is to act like *this*.... The point of such a response is not to provide us with a substantive and informative account, but to give us a reflective perspective on what we already understand in the multiplicity of cases for which we would use the description 'acting for a reason'.

REFERENCES

[1] G.E.M. Anscombe. Causality and determination. In *Metaphysics and the Philosophy of Mind:*, volume II of *The Collected Papers of G.E.M Anscombe*. Blackwell, Oxford, 1981.

[2] G.E.M Anscombe. On the grammar of 'enjoy'. In *Metaphysics and the Philosophy of Mind: Collected Philosophical Papers Volume II*. Blackwell, 1981.

[3] G.E.M. Anscombe. The reality of the past. In *Metaphysics and the Philosophy of Mind*, volume II of *The Collected Philosophical Papers of G.E.M. Anscombe*. Blackwell Publishers, 1981.

[4] G.E.M. Anscombe. *Intention*. Harvard University Press, second edition edition, 2000.

[5] G.E.M. Anscombe. The causation of action. In *Human Life, Action and Ethics*. St. Andrews, Studiels in Philosophy and Public Affairs, 2005.

[6] G.E.M. Anscombe. Practical inference. In *Human Life, Action, and Ethics: Essays by G.E.M. Anscombe*. St. Andrews, Studies in Philosophy and Public Affairs, 2005.

[7] G.E.M Anscombe. Frege, wittgenstein, and platonism. In Luke Gormally Mary Geach, editor, *From Plato to Wittgenstein: Essays by G.E.M. Anscombe*, St. Andrews Studies in Philosophy and Public Affairs, chapter 11, pages 127–135. Imprint Academic, 2011.

[8] G.E.M Anscombe. Ludwig wittgenstein. In *From Plato to Wittgenstein: Essays by G.E.M. Anscombe*. Imprint Academic, 2011.

[9] G.E.M. Anscombe. A theory of language? In Luke Mary Geach, editor, *From Plato to Wittgenstein: Essays by G.E.M. Anscombe*, St. Andrews Studies in Philosophy and Public Affairs, chapter 17, pages 191–203. Imp, 2011.

[10] G.E.M. Anscombe. Wittgenstein: Whose philosopher? In Luke Gormally Mary Geach, editor, *From Plato to Wittgenstein: Essays by G.E.M. Anscombe*, St Andrews Studies in Philosophy and Academic Affairs. Imprint Academic, 2011.

[11] Stina Bäckström. What is it to depsychologize psychology? *European Journal of Philosophy*, 25(2):358–375, 2017.

[12] Jason Bridges. Meaning and understanding. In *A Companion to Wittgenstein*. Wiley-Blackwell, 2017.

[13] Jonathan Dancy. *Practical Reality*. Oxford University Press, 2002.

[14] Jonathan Dancy. Two ways of explaining action. *Royal Institute of Philosophy Supplement*, 55:25–42, 2004.

[15] Donald Davidson. Actions, reasons, and causes. In *Essays on Actions and Events*. Clarendon Press, Oxford, 1980.

[16] Donald Davidson. Freedom to act. In *Essays on Actions and Events*. Clarendon Press, Oxford, 1980.

[17] Donald Davidson. Introduction. In *Essays on Actions and Events*. Cl, 1980.

[18] Wayne A. Davis. Reasons and psychological causes. *Philosophical Studies*, 122:51–101, 2005.

[19] Cora Diamond. The face of necessity. In *The Realistic Spirit: Wittgenstein, Philosophy,and the MInd*. The, 1991.

[20] David Finkelstein. Wittgenstein on rules and platonism. In *The New Wittgenstein*, pages 83–100. Routledge, 2000.

[21] Anton Ford. The arithmetic of intention. *American Philosophical Quarterly*, 52(2):129–143, 2015.

[22] Alvin Goldman. *A Theory of Human Action*. Princeton University Press, 1970.

[23] John Hyman. *Action, Knowledge, and Will*. Ox, 2015.

[24] Eric Marcus. *Rational Causation*. Harvard University Press, 2012.

[25] Neil McDonnell. The deviance in deviant causal chains. *Thought: A Journal of Philosophy*, 2015.

[26] John McDowell. Functionalism and anomalous monism. In *Mind, Value, and Reality*. Harvard University Press, 1998.

[27] Rush Rhees. On continuity: Wittgenstein's iidea, 1938. In *Discussions of Wittgenstein*. Tho, 1970.

[28] Rush Rhees. 'the philosophy of wittgenstein'. In *Discussions of Wittgenstein*. Thoemmes Press, 1970.

[29] Sebastian Rödl. *Self-Consciousness*. Harvard University Press, 2007.

[30] Sebastian Rödl. The form of the will. In *Desire, Practical Reason, and the Good*. Oxford University Press, 2010.

[31] Bertrand Russell. The limits of empiricism. *Proceedings of the Aristotelean Society, New Series*, 36:131–150, 1936.

[32] Kieran Setiya. *Reasons without Rationalism*. Princeton University Press, 2007.

[33] Kieran Setiya. Reasons and causes. *European Journal of Philosophy*, 19(1):129–157, 2009.

[34] Kieran Setiya. Reply to bratman and smith. *Analysis*, 69(3):531–40, 2009.

[35] Michael Thompson. Apprehending human form. In Anthony O'Hear, editor, *Modern Moral Philosophy*. Cambridge University Press, 2004.

[36] Michael Thompson. *Life and Action: Elementary Structures of Practice and Practical Thought*. Harvard University Press, 2008.

[37] Peter Winch. *The Idea of a Social Science and its Relation to Philosophy*. Routledge: London, second edition edition, 1990.

[38] Ludwig Wittgenstein. *The Blue and Brown Books: Preliminary Studies for the 'Philosophical Investigations'*. Blackwell Publishers, 1958.

[39] Ludwig Wittgenstein. Cause and effect: Intuitive awareness. In Alfred Nordmann James Klagge, editor, *Philosophical Occasions: 1912-51*. Hackett, 1993.