Wittgenstein on Reading

Amos Browne

October 4, 2015

Introduction

1. My aim today is to show the ongoing philosophical significance of Wittgenstein’s brief discussion of reading in his *Philosophical Investigations*. Wittgenstein’s interlocutor in these sections aims to provide a particular kind of account of *what it is to read*. Despite the apparently esoteric nature of its topic, the form of this account is shared by contemporary work on more familiar philosophical issues. I argue that Wittgenstein’s response to his interlocutor demonstrates that no such account is possible, and that the points he raises here apply equally to other accounts of the same form.

To demonstrate this, I connect these sections of Wittgenstein’s work to contemporary accounts of *what it is to act for a reason*, and accounts of how explanations that make such acts intelligible do their work. By building on material from Wittgenstein’s discussion of reading, I show a general difficulty that attends any such account of a form of reason-giving explanation and the acts it explains. To do this, I introduce a somewhat artificial kind of reason-giving explanation, related to acts of reading, and show how the general difficulties raised in Wittgenstein’s discussion apply to it. I then suggest that these problems generalize to many other kinds of reason-giving explanation, and draw a number of conclusions from this claim, bringing in material from the surrounding sections of the
Part I: §156-§171: Wittgenstein on Reading

2. The discussion of reading comprises a short sequence of passages from §156-§171, which fall in the middle of an extended investigation of understanding and related notions. At the beginning of the excursus, Wittgenstein provides the following brief description of what he will mean by ‘reading’:

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 general shape of the dialectic that will concern us is then given in a compressed form in the rest of §156. Wittgenstein begins the excursus by contrasting an experienced reader with 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.—We should also say that he had read a sentence if he spoke neither aloud nor to himself during the reading but was afterward able to repeat the sentence word for word.

---

1Or up to §178 if you include the seven sections on guidance
2Given 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.
word or nearly so.—He may attend to what he reads, or again—as we might put it—function as a mere reading-machine: I mean, read aloud and correctly without attending to what he is reading; perhaps with his attention on something quite different (so that he is unable to say what he has been reading if he is asked about it immediately afterward).

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).

Wittgenstein suggests that if we focus our attention on the beginner, we will be inclined to give a certain kind of account of what reading consists in: we shall be inclined to say “it is a special conscious activity of the mind”. Here Wittgenstein is giving voice to the idea that there must be characteristic marks—in this case, sensations—in virtue of which a person counts as, or knows themselves to be, reading. And yet, he goes on, if pressed on the matter we would have to admit that any “conscious activity of mind” we might treat as constitutive of reading might also be present in the mind of someone who was merely pretending to read.

Faced with this, the interlocutor responds by insisting that there must nevertheless be some difference between what goes on with an experienced reader, and what goes on in someone who merely pretends to read: if not in their conscious experience, then in some other fact about them:

Now we should of course like to say: What goes on in that practised reader and in the beginner when they utter the word can’t be the same. And if there is no difference in what they happen to be conscious of there must be one in the
unconscious workings of their minds, or, again, in the brain.—So we should like to say: There are at all events two different mechanisms at work here. And what goes on in them must distinguish reading from not reading. §156

Here we have the idea that there *must* be something that distinguishes the two cases (one can bring out the insistence by emphasizing the right words: what goes on *can’t* be the same, there *must* be a difference, and it *must* distinguish reading from not reading). Struck by the idea that there must be *something* that reading consists in, the interlocutor goes on to insist that we can know this much without being able to specify what the relevant difference might be. Later he suggests that this inability comes “only because of our too slight acquaintance with what goes on in the brain and the nervous system” (§158):

> 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”.—And it presumably must be like that—for otherwise how could we be so sure that there was such a connexion?

This dialectic, which appears in compressed form in §156, is then taken up and repeated in different ways throughout the subsequent sections. Its general shape might be summarized as follows: Wittgenstein’s interlocutor suggests a possible candidate for what *reading* consists in; it is shown that this candidate might also be present in cases where there was no *reading*; in response, the interlocutor insists that nevertheless there must be *something* that distinguishes the two cases—even if it is something that we are as yet unacquainted with.

3. In all of this, Wittgenstein’s interlocutor repeatedly comes back to the idea that his account of *what it is to read* should be spelled out in terms of a *felt or conscious experience*. But why should Wittgenstein (or anyone else) take such an account seriously? Doesn’t it
suggest a fairly naive and implausible approach to psychological concepts and the philosophy of mind in general? Perhaps there were some British Empiricists who said this kind of thing, but one would certainly be hard-pressed to find contemporary philosophers defending accounts of any of the psychological concepts that Wittgenstein is concerned with in such terms as these. If anything, one finds philosophers complaining about the emphasis on felt or conscious experience amongst Wittgenstein’s students, as when Kieran Setiya complains that G.E.M. Anscombe’s understanding of the significance of her own concept—mental cause—is distorted by “a bad philosophy of mind—the illicit focus on what we feel” (Setiya 2007: 47).

My contention is that there is a lot more going on in these passages of the *Philosophical Investigations* than such a reading suggests—in fact, they exemplify a manner of engaging with philosophical positions that is characteristic of the book as a whole. I take the *PI* to be a reflection on the activity of doing philosophy, and the kinds of account we find ourselves tempted to provide when we pursue it. The positions taken by the interlocutor, and the ways in which Wittgenstein responds to them, should be taken as providing an outline of patterns that will be found more broadly in philosophical discourse. The interlocutor provides a primitive and perhaps implausible version of a particular kind of philosophical account—and Wittgenstein’s responses are supposed to help us see why no such account can meet the ambitions it sets for itself. This is why it should come as no surprise if his discussion of a concept like *reading*, which might seem to be of apparently marginal interest, should have clear parallels with central debates in contemporary philosophy.

4. Let’s return to the first stage of the dialectic: the interlocutor takes some feature of a particular case of *reading*—say, the kind of sensations that accompany reading when we do it slowly and deliberately—and suggests that *reading* consists in those sensations.

---

3 My reading of the *Philosophical Investigations* in this way has been influenced by Warren Goldfarb’s paper *I Want You To Bring Me A Slab* (Goldfarb 1983), along with comments made by Jason Bridges in a class on the *Philosophical Investigations* in Winter 2014.
Perhaps no serious philosopher would propose such a simplistic account, but attention to its overall tendency is illuminating.

Why should the interlocutor be tempted by the possibility of providing an account in terms of the *sensations* that accompany reading? Here is a suggestion: it is because the interlocutor also thinks of these sensations as something *given, occurent, and self-identifying*. This, as we’ll see, makes them excellent candidates for the kind of account that he wants to provide.

Let’s begin with the idea that these sensations are something *given*. As the *Philosophical Investigations* continues, we see that the interlocutor is strongly attached to this idea: he thinks that sensations are something 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. One might say that they provide an already given basis for our application of those terms, one which serves to provide a foundation for our use of them. In the present discussion, this means that the interlocutor thinks of sensations as something that can be specified independently of the circumstances in which they occur: the *reader* either experiences the relevant sensation, or she does not, and this can be determined (by her, at any rate) without any attention to the broader context in which the act of reading takes place.\footnote{Here and throughout the book we get a picture of activities of the mind as things that can be explained *mechanistically*—that is, in terms of concepts that can be applied without attending to the broader context in which an object is embedded. Think, for instance, of the difference between understanding the inner working of a watch qua mechanism, and understanding it as an artefact to be used for telling the time. For more on mechanistic conceptions of the mind in the *Philosophical Investigations*, and their connections with the interlocutor’s appeals to sensations and mental imagery, see Bridges (Forthcoming).} This is of course connected to the idea that sensations are something *occurent*: they occur *as we read*, which means that they provide a mark of *reading* that is fully present with the activity itself. In other words, there would be *something*—specifiable independently of the broader circumstances, and fully present at the time of the act—*in virtue of which* it would count as an act of *reading*. What’s more, since it is a characteristic feature of sensations, as the interlocutor conceives of them, that they can be immediately identified by their possessor, this might suggest that any appeal to them will be sufficient to capture
the self-conscious character of the phenomena for which the interlocutor wants to provide an account. If I am suffering a sensation of pain, I know that I am suffering it; and likewise, if I am reading, I know that I am reading. The interlocutor might try to explain the latter fact on the model of the former.\textsuperscript{5}

As such, an appeal to particular sensations might seem to be sufficient to provide a certain kind of account of what it is to read. First, since sensations are something given and as such specifiable independently of our other concepts, it would be an account that explains the concept reading in terms that do not depend on it, while still capturing its essential characteristics. That is, it would be the basis for an account that explained what it is to read in terms of something that was prior to and independent of the concept reading. Second, it would explain what it is to read in terms of a feature or mark in virtue of which the particular act would count as an act of reading—and this mark would be specifiable independently of the broader circumstances of the act, while also being something that occurs while we read. Third, it would provide an account that served to explain the self-conscious character of reading: the fact that, when a person reads, she knows that she is reading.

This is the basis of my claim that the apparently naive the move that the interlocutor makes as his opening gambit is one that might be found in any kind of philosophical investigation into what something consists in, or what it is to be such-and-such. A philosophical account would share something with the account provided by the interlocutor if it tried to explain what, for example, acting for a reason consisted in in terms that were supposed to be graspable independently of, and prior to, our understanding of acting for a reason.\textsuperscript{6} I shall call accounts of this sort \textit{reductive}, since they aim to provide an analysis in terms of notions

\textsuperscript{5}See Bridges (Forthcoming)

\textsuperscript{6}Such an account needn’t maintain that these further terms were radically prior, in the way the interlocutor seems to think that sensations are; it might merely maintain that these terms can be taken for granted in providing this particular account. Indeed, this is a move made by some of the contemporary philosophers we will be considering in the next section (see below). Given the shape of such accounts, one might think that the interlocutor’s version is particularly clear-sighted: he is appealing to something that he thinks is given not just in the context of this particular investigation (i.e., something he can take for granted here, but which he might need to provide an account of on some other occasion), but given in a radical sense. This contrasts with the way in which a concept like “causation” appears in contemporary account of acting for a reason: the concept is taken for granted, but with the admission that an account of it might be provided on some other occasion. See for instance Setiya (2009).
that are supposedly available independently of the concept under investigation. It is characteristic of such accounts to look for a mark or feature, graspable independently of the relevant concept, *in virtue of which* that concept can be said to apply.

5. As we saw above, Wittgenstein responds to the interlocutor by showing that the criteria by which he hopes to explain the concept *reading* might well be present in cases where we would not apply the concept, and as such cannot be sufficient basis for an account.

In the first instance, this line of argument is applied to the sensations that the interlocutor identifies: someone could experience *them*, and yet not be reading. For instance, in §159 the interlocutor focuses on the “characteristic sensations” that accompany reading, and contrasts them with “characteristic sensations” that might accompany, say, reciting a passage one has memorised in a language one cannot read. The suggestion is that *reading* consists in the presence of the former. 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”—§160). Wittgenstein’s further suggestion is that we can always imagine any such “characteristic sensation” to be felt in the wrong kind of case.

In other words, it looks like there is no particular sensation we can identify that will be present in all and only cases of *reading*; and even if we could, it seems as though *qua* sensation there is no reason to suppose that it could not occur on occasions when we would not count the person as reading. The very fact that made sensations seem suitable for providing an account here—their radical independence and priority—also means that, by the interlocutor’s own lights, they should always be separable from the phenomena under investigation.

This same pattern of argument is applied to the interlocutor’s more “sophisticated” suggestions as well. For instance, in §162 we find the suggestion that reading involves “deriv-
ing” 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”. 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. 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 without appeal to the concept *reading*.

Anscombe provides a succinct summary of the difficulty here:

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?’ (Anscombe 2011: 211)

It is important to see that this claim does not involve a denial of the thought that e.g. there are sensations characteristic of reading. In particular circumstances, a particular sensation could quite correctly be called a sensation of reading; and it could strike us as an important fact about the place reading has in our lives that it involve such sensations. But that truth about the sensation belongs with other truths about what led up to and followed its occurrence, not to mention truths about the person to whom it occurred, and the culture and society in which they have been raised. Our characterization of something as a sensation of reading belongs with all of this; it is, in an important sense, dependent on and posterior to our concept of *reading*. We don’t recognize this as a case of reading in virtue of recognizing particular sensations that accompany it—in fact, we don’t recognize something as a case
of reading in virtue of anything at all, at least not as the interlocutor would understand that phrase.

Again, there may be some truth to the claim that in reading one derives what one says from the words on the page—in fact, I think one might hear that as a truism, a grammatical statement that tells us what reading is in an unproblematic sense. But this is only true because our grip on what we mean by “derive” here goes via our concept of reading, and cannot be fully specified apart from it: one derives what one says in the way characteristic of reading. As soon as we do try to take the two senses apart, we lose our grip on the sense of “derive” that applies to all and only cases of reading; and there is no way of recovering it without appeal to that concept.

The difficulty seems to be that our understanding of the relevant sense of “deriving”, or on what is characteristic of the “special way in which words come to one”, belongs with our understanding of the concept of reading. This means that we cannot appeal to these ideas and provide the kind of account that the interlocutor aims to give: an account that explains what it is to read in terms that are independent of that concept.

Part II: Reading and Reading Explanation

6. This general overview of §156-§171 was intended to bring out some key moments in the discussion, and to provide us with a model of a particular kind of philosophical account, and a way of responding to it, that might prove useful in the face of more recognizable philosophical problems than the question ‘What is it to read?’. But its usefulness will surely depend on whether we can find other debates taking the same form. In this section, I will show that contemporary discussions of ‘What it is to act for a reason’ can be helpfully compared to Wittgenstein’s investigation of reading. I will do this by developing certain strands in Wittgenstein’s account in a way that brings out these parallels.
The quote from Anscombe above provides a good starting point. There, in an essay titled *Wittgenstein: Whose Philosopher?*, Anscombe provided a nice summary of the overall tendency of Wittgenstein’s responses to his interlocutor. But elsewhere in her work, we find Anscombe offering similar objections in her own voice in response to theses held by Donald Davidson.

Davidson opens his 1963 paper *Actions, Reasons, and Causes* (Davidson 1980) with the following question:

**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?

His answer was that we should understand this relation in terms of a causal connection holding between certain psychological states of the agent and his action. Over time Davidson came to see certain problems with this account, encapsulated in the possibility of ‘deviant causal chains’ on display in his well-known climber-case. Anscombe’s abrupt response to Davidson’s suggestion in her 1974 paper *Practical Inference* (Anscombe 2005) picks up on this problem:

[Davidson’s] solution lacks acumen. 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 the grounds* of the belief. Davidson indeed realized that even identity of description of *act done* with *act specified in the belief*, together with causality by the belief and desire, isn’t enough to guarantee the act’s being done *in pursuit of the end* and *on the grounds of the belief*. He speaks of the possibility of ‘wrong’ or ‘freak’ causal connexions. I say that any recognizable causal connection would be ‘wrong’, and that he can do no more than postulate a ‘right’ causal connexion in the happy security that none such could be found. If 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?’.

(Anscombe 2005: 110-1)
We should compare this interchange with the model provided in the sections on reading. Davidson suggests a possible candidate for an account of *what it is to act for a reason*: to act for a reason is for one’s action to be caused by a particular pair of psychological states. Anscombe’s response is that this criterion could be met in cases where a person does not act for a reason—the criterion provided by the account does not succeed, since for any act caused by such states, we can always ask: “But was the action done for the sake of the end and in view of the thing believed?”. In other words, we could always ask: “But was the act an *intentional action*?” Anscombe’s question is supposed to bring out the fact that *any* account specified in the terms Davidson allows himself could not settle the question of whether a person was acting for a reason.

It is perhaps unfair to characterize Davidson as trying to provide an account here, but many contemporary philosophers explicitly set themselves that aim, defending a ‘causal-psychological account of acting for reasons’. Rather than focusing on the interchange between Davidson and Anscombe, I shall take the accounts of contemporary authors such as Kieran Setiya and Wayne Davis as my model. Both argue, like Davidson, that what it is to act for a reason should be understood in terms of a causal connection that holds between certain psychological states and the agent’s actions. And both develop this claim with the aim of providing a reductive account of what it is to act for a reason.

This is perhaps clearest in the way such authors introduce the notion of causality that they call on in providing their accounts. It is a key claim of such accounts that the relevant notion of causation can be understood *independently* of an understanding of what it is to act for a reason. For example, Kieran Setiya appeals to a generic notion of psychological causality, one that is common to *any* case where a psychological state causes an action, including the kind of ‘deviant causation’ involved in the climber case, meaning that it is independent of any account of acting for a reason. Elsewhere in his work, again dealing

---

7Setiya (2009); see also Setiya (2007) and Davis (2005)
with deviant causation, he makes an analogous 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.\(^8\)

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. (Setiya 2007: 32)

Setiya is thus quite explicit about the reductive form his account takes. Describing the shape of his approach, he states:

\[\text{[W]e can ‘build’ acting for reasons from materials present in deviant causality, along with others that are not themselves composed or defined in terms of such action.}\]  
Setiya (2009: 135)

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

7. Taking their lead from Davidson’s Question, most causal-psychological accounts take as their starting points explanations that make a person’s action intelligible by providing their reasons for it. Setiya mentions the following three expressions, which we might call Rational Explanations:

- A is doing \(\phi\) because \(p\)
- A is doing \(\phi\) on the ground that \(p\)

\(^8\)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.
Setiya’s approach here is fairly typical of the literature as a whole: these schemata are
supposed to help us isolate a general class of explanations that are distinguished by the
fact that they explain an agent’s acts by giving her reasons for them, and as such they are
supposed to get the topic of investigation into view all at once. To make this point explicit,
Setiya appends a subscript indicating that the reason specified in the explanation was the
agent’s reason:

- A is doing $\phi$ because $R\ p$.

One added twist is that many authors (though not, it has to be said, Setiya) motivate their
causal-psychological accounts by pointing out that there can be cases in which a person
takes $p$ as their reason for acting, and does something on this basis, though it in fact turns
out that $p$ was false. Since here too we have cases of people acting for reasons—bad or
false ones—some proponents of causal-psychological accounts will suggest that the basic
form of explanation is in fact:

- S is doing $\phi$ because she believes that $p$.

since this schema has the advantage of applying to both good and bad cases.

To parallel this approach, let’s focus on a distinctive class of explanations I’ll call reading explanations. What I mean by reading explanation is simple enough: I have in mind the fact that, when a person is reading, we can explain her utterances by reference to signs she reads. So, for instance, when asked why she said “$kæt$”, she might explain her act by pointing to the written sign “CAT”. An explanation of this sort makes her utterance intelligible as an act of reading.

Suppose, following the model of rational explanation described above, we tried adopting the following as a schema for reading explanation:

$S$ uttered “…” because of the sign “_____”
This captures the idea that we are explaining a person’s utterance as a response to a particular sign. The trouble is that explanations of this form will be appropriate to many utterances that are not acts of reading: §151-§171 give us a whole slew of cases in which people make utterances in response to signs, yet are not counted as reading. We might, then, adapt our schema to indicate that we are only interested in cases where the sign explains the utterance in the way it does when a person is reading by adding a subscript that parallels Setiya’s:

\[ S \text{ uttered “…” because}_{R} \text{ of the sign “…”} \]

Let this be our schema for reading explanation.

8. Reading Explanation, and the acts it explains, have a number of interesting parallels with the broader class of rational 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 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.

Note that 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. This relation displays all the marks of causality cited by causal theorists. So reading explanations are a kind of causal explanation.

Note further 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. 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. Once again, we can mark this limited sense by appending a 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 play a justificatory role.

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

---

\(^9\)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 “COT”, we would be perplexed.
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 reponse 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 above 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 misreading, 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 taking them to be acts of reading when in fact they are not. 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.

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.$\text{10}$

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 misreads a sign. We can divide such cases into two general groups: first, there will be those in which the person misper-veives 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.

But 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 “$køt$” “CUT”. In such a case, we get the following instance of our schema:

$$S\text{ said }“kæt”\text{ because of the sign }“\text{CUT}”.\text{10}$$

Here the reason given for the person’s utterance is not a reason$R$ for that utterance; it is a

$\text{10}$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.
reason$_R$ to say “CUT”, not “CAT”.$^{11}$ Since this is a defective case, we would nevertheless expect an explanation such as the following to be true:

\[ S \text{ said “} kæt \text{” because she thought sign was “CAT”}.^{12} \]

Here we do have mention of something that \textit{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:

\[ S \text{ said “…” because$_R$ she thought the sign said “”}. \]

9. We can now reframe Wittgenstein’s investigation into \textit{reading} in terms of the notion of reading explanation outlined above.

First, there is the following reformulation of Davidson’s Question:

\begin{quote}
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?
\end{quote}

As we’ve seen, a person can make a noise in response to a sign and yet not be reading it, such that even if we say that the sign was the reason she made that utterance, it will not

---

$^{11}$Remember, this is compatible with the claim that the sign “CUT” was the reason our reader said “CAT”. The claim here is just that this cannot be the “reading” sense of reason: reason$_R$.

$^{12}$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.
be her reason. So what is it that differentiates the cases we are interested in?

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

\[ S \text{ said } "\ldots" \text{ because}_R \text{ of the sign } "\ldots". \]

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:

\[ S \text{ said } "\ldots" \text{ because of the sign } "\ldots". \]

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?”

10. With these preliminaries out of the way, we can now come up with sketches of reductive accounts of what it is to read. Their basic form will be as follows: in order to provide an explanation of what reading is or consists in, they 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-reductive 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:

\[ S \text{ uttered "..." because "" was written on the page.} \]

but with further specification we will be able to show that the causes of reading are sig-
nificantly different from the causes of other utterances. The substance of any particular
causal-reductive account will lie in a specification of those causes.

11. 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

\[ S \text{ uttered "..." because "" written on the page.} \]

is common to both the beginner and the reader, a plausible initial move is to try to iden-
tify 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:

\[ S \text{ uttered "..." because she saw the sign "" written on the page.} \]

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.\(^{13}\) 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.

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

---

\(^{13}\) 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.
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:

\[ S \text{ uttered } "\ldots" \text{ because she saw the sign } "\ldots" \text{ and was in state } \sigma \]

and further provide an account of what it is to read in the following form:

To read is for one’s utterance “\ldots” to be caused by a perception of the sign “\ldots” and state \( \sigma \).

I will not attempt to come up with a plausible candidate for such a state, but its 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 \( \sigma \) 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:

\[ S \text{ uttered } "\ldots" \text{ because she saw the sign } "\ldots" \text{ written on the page and believed that in saying } "\ldots" \text{ she would be reading } "\ldots". \]

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 “\ldots” to be caused by the perception of a sign “\ldots” and the belief that in saying “\ldots” one would be reading “\ldots”.

However, for now let us temporarily set aside these divergent accounts and grant the reductivist his claim that an appropriate state \( \sigma \) can in principle be specified without appeal to the concept reading—would this yield an adequate account?
Again, 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.\(^{14}\)

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 it might seem as though 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 \(\sigma\) could not suffice—for that same state could be present, and cause an utterance—and yet the utterance not be an instance of reading!\(^{15}\)

\(^{14}\)Nic Koziolek helped formulate this example.

\(^{15}\)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 out loud. 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 (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.
12. 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. 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.

This point is easily obscured. An interlocutor could insist 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. 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.

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.

The interlocutor’s insistence that there must be some such mechanism warrants further attention than I will give it here. For our purposes, the important point about this move...

\[\text{Cf. (Davis 2005: 70)}\]

\[\text{Note that this was also a possible response when we asked the interlocutor to specify a psychological state σ unique to readers—he might say that, though we are unable to specify such a state now, the progress of neuroscience will eventually allow us to do this.}\]

\[\text{Davis (2005: 70). The idea of an ‘r-result’ in the text echoes Davis’ strategy:}

\[\text{We know that there is a characteristic way in which intentional action results from desire, even though there is a great deal we do not know about it. I will say that φing is an f-result of the desire to φ when the action results from the desire in that way.}\]

\[\text{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:}

\[\text{Note also how sure people are that to the ability to add or multiply or to say a poem by heart etc., there must correspond a peculiar state of the person’s brain, although on the other hand they know next to nothing about such psycho-}\]
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?\(^\text{20}\)

This difficulty was already implicit in the bare specification of an “r-result”. This clearly 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 be 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.

13. Of course, a philosopher might try to provide more than the bare specification of a causal connection. One common move in the literature on acting for a reason is an appeal to the idea of a disposition or ability as the source of non-deviant acts of the relevant kind.\(^\text{21}\)

Accounts of this form emerge partly in response to the problem of causal deviance, the hope being that deviant cases can be ruled out by the fact that they do not originate in the relevant disposition.

The problem with such counts should by now be obvious: how are we to describe the relevant disposition other than as a ‘disposition to read’? Simply stating that the act emerges from a reliable disposition in the agent won’t do, since philosophers can arrange for all manner of deviance. We can imagine, for instance, a non-reader who reliably associated particular words with the shapes of various signs, without being able to read any of them.

\(^{20}\) 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.

\(^{21}\) Setiya appeals to this notion in sketching his response to the problem of causal-deviance. See Setiya (2009). Marcus (2012) defends a non-reductive account that relies more explicitly on the notion of a disposition, ability, or capacity.
Such a person’s acts would spring from their disposition, and might appear to the casual observer as acts of reading, even though they are not.\footnote{Likewise we can imagine our teacher in the case described above as having an unfortunate syndrome that leads her to blurt out particular words when she gets anxious. That the source of her anxiety happens to be an occasion for reading the word she comes out with doesn’t mean that she read it. Its source was in some other disposition than her ability to read.} Once again, if appeal to the concept \textit{reading} is forbidden, it is difficult to see how such accounts could get off the ground.

14. As we saw above, Wittgenstein’s interlocutor in the \textit{Philosophical Investigations} tries a range of further approaches: reading is a \textit{special mental process}, in reading one \textit{derives what one say from what is written}, in reading the words \textit{come to one in a special way}. In every case, the difficulty is that the interlocutor is unable to say anything satisfactory about these notions without an appeal to the concept \textit{reading}.

In effect, what the subscript on our schema indicates is that the act the explanation aims to make intelligible is an act of reading. It is only through recognizing it as such an act that a person will be able to appreciate the full significance of such an explanation. The point can be made more clearly if we temporarily set aside our artificial schema, and imagine a particular scenario. Suppose we came across someone staring at some strange marks on a wall and making a series of 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 is only 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. For in our puzzlement we could already see that there was some connection between the markings and their noises: our problem was that we could not tell what it was.\footnote{We might also imagine a variation on this case, in which a child who has no understanding of what it is to read comes across someone reading. Suppose she is fascinated, and delighted by the strange sounds she asks “why did he make that noise?”. How might we 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 \textit{reading} is—and, arguably at least, a full grasp of that concept requires learning how to read.}

Thus, someone might argue that even after being told that the person was reading cuneiform script, we wouldn’t \textit{fully} understand this act \textit{qua} act of reading unless we too could read the script, and see the relation between the particular marks and the noises he made. We will return to this thought below.
The subscripts on our schema can then be seen as doing the work that might also be done by just being told that what we have before us is a case of reading. So instead of:

\[ S \text{ said ‘‘…’’ because}_R \text{ of the sign ‘‘____’’}, \]

we might say

\[ S, \text{ reading, said ‘‘…’’ because of the sign ‘‘____’’} \]

Or most naturally of all:

\[ S \text{ read ‘‘…’’}. \]

The first reformulation brings out the idea that there is a dimension of intelligibility here that depends on recognizing the target of the explanation as an act of reading. And the second shows that to describe something as an act of reading is not merely to classify it, but to provide some minimal form of explanation: in characterizing it as an act of reading, we already say something about how it is to be explained.

15. We are now in a position develop what I take to be an important lesson from Wittgenstein’s discussion. I have suggested that reading explanation is a distinctive kind of explanation, one that makes sense of acts \textit{qua} acts of reading. If this is correct, it casts our use of the subscript \textit{’’}_R\textit{’’} in an interesting light. Earlier I insisted that it was merely a useful shorthand to help demarcate our topic. But one thing that is becoming apparent is that we might end up appending it to any of a number of closely related words: because\(_R\), reason\(_R\), justification\(_R\), cause\(_R\), derive\(_R\), etc. In each case, it is shorthand for something like “take this in the way we use this word when we apply it to cases of reading”.

It might seem then that the sense of these familiar words is being modified or transformed by their application to cases of reading.\(^{24}\) Of course, its not that these words have taken

\(^{24}\text{This way of putting the point may be misleading. It will be helpful to keep in mind the following stricture from Anscombe:} \]
\[ \text{Where we are tempted to speak of ‘‘different senses’’ of a word which is clearly not equivocal, we may infer that we are in fact pretty much in the dark about the character of the concept which it represents.} \]
\[ \text{§1 (Anscombe 2000)} \]
on a radically different sense, so that they are now ambiguous between two completely
distinct meanings. And yet a person will not be able to fully grasp the way these words
are used unless they have mastered the concept *reading* (and *mastering this concept* arguably
means *learning to read*). In other words, they will not know what it is for a sign to be a
reason$_R$ for an utterance, how a sign might provide justification$_R$ for an utterance, or how
an utterance could be made because$_R$ of a sign, unless they have some understanding of
what it is to *read*.

Obviously this does not preclude a mastery of these words as they apply to other cases—if
anything, it presupposes it, since it would be rather odd if a person came to master con-
cepts such as *cause* or *justification* only as they applied in cases of *reading.* And needless
to say, humans were surely offering explanations that could be articulated by such con-
cepts before the invention of reading.

This brings out something important: the concepts marked by our subscripts are in fact
extensions of concepts that a person can grasp without any understanding of the con-
cept *reading*. This is part of what motivated the reductivists claim that e.g. the notion of
causality involved in cases of *reading* was one that could be grasped independently of that
concept. After all, they might point out, a person who was unfamiliar with the practice of
*reading* might still manage to work out that readers were making utterances in response
to particular markings; and he might conclude from this that those markings *caused* those
utterances, perhaps even that they *justified* them, etc. But while this is certainly true, it is
also true that the observer would not be able to fully appreciate the sense in which the
markings caused or justified the utterances *unless* he knew what reading was. This sug-

---

25 As Rush Rhees put it in a different (though related) context: “[If] I acquire or adopt a new concept, I can do this only within a *system*
of concepts, only where I recognize the continuity or ‘identity’ between the concepts I had learned already and the concept I have now”
(Rhees 1972-3: 38). Also cf., from Anscombe (1981):

The word “cause” itself is highly general. How does someone show that he has the concept *cause*? We may wish to say: only by having such a word in his vocabulary. If so, then the manifest possession of the concept presupposes the mastery
of much else in language. I mean: the word “cause” can be *added* to a language in which are already represented many
causal concepts. …But if we care to imagine a language in which there no special causal concepts are represented, then
no description of the use of a word in such a language will be able to present it as meaning *cause*.

26 In fact, one might even say that he could not fully understand the utterances unless he could read the language in question.
gests it doesn’t really matter if we think of our subscripted concepts as new concepts, or as extensions of our old ones. The important point is that they involve a kind of transformation in our understanding of the old notion—one that allows us to see how it applies to these new cases.

This thought, if correct, helps us to see why it seemed so difficult to provide an account what it is to read in terms that were independent of the concept reading. We can now see that there certainly are marks or features that are distinctive of cases of reading: the utterances in question are made because of particular signs, where this means that they are caused by perception of particular signs, and justified by reference to those signs. We might also, perhaps, say that when a person reads they derive what they say from the signs in front of them, and that this is a particular process that is distinctive from other ways of formulating utterances on the basis of signs. That is, we may say many of the things that Wittgenstein’s interlocutor wanted to say. But we can also now see that, in every case, an understanding of the mark or feature given in our account of what it is to read itself depends on an understanding of the concept reading. This is what our subscripts have been marking, and what we meant when we said that the utterance was “caused”, “justified”, “derived”, etc. in the way such things are in cases of reading.

But if an understanding of these features or marks depends on the concept reading, then they cannot be used to provide an informative account of what it is to read. Anyone in a position to understand such an account would already need to know what it is to read, and to apply this knowledge in deploying the account. This is why the reductive accounts of reading explanation that we considered above were bound to fail. Their whole approach was framed by a need to find some distinctive mark or feature in virtue of which instances of reading would count as such, but no such mark can be specified apart from an understanding of what it is to read.
Part III: Normativity and Explanation

16. So far we have seen that there appear to be deep problems facing any reductive account of reading explanation, and of course I mean this to further suggest that similar problems will also face any attempt to provide a reductive account of rational explanation more broadly considered. To lend further plausibility to this claim, in this final section I will outline what I take to be the real source of the problems confronting reductive accounts, and show that the same considerations will apply directly to any species of rational explanation.

The problem stems from the fact that the relations that such explanations depend on are both normative and explanatory. In a paradigm instance of reading explanation, the fact that explains the person’s utterance—that such-and-such a sign was written on the page—also provides what is generally called a ‘normative reason’ for that utterance. Classifying the act as an act of reading already involves recognizing it as such as to be explained in this way: that is, we expect the act to be explained by something that stands as a reason \( R \) for it.

Thus the explanatory power of reading explanations depends on an understanding of the normative relations involved in reading. This suggests that—contrary to the claims made by many reductive approaches—the notion of a reason that figures in an explanatory role in such explanations cannot be separated from the notion of a reason that figures in an understanding of what it is to read.\(^{27}\) This further means that the explanatory features in virtue of which such explanations are true—for instance, causal relations holding between certain facts and the acts of an agent, or the psychology involved in realizing this relation—also cannot be properly characterized without appeal to the normative dimension of reading. To demonstrate what I mean by this claim, I will return once more to an attempt to characterize the psychological underpinnings of acts of reading.

\(^{27}\) For claims that the normative and explanatory sense of reason are distinct, see Davis (2005) and Setiya (2007).
17. We can start by returning to the last suggestion considered on behalf of reductive accounts: that **reading explanations** should be understood as relying on appeal to a particular disposition. As we just saw, there seems to be little prospect of specifying the relevant disposition without appeal to the concept *reading*, which undermined any hope of formulating a reductive account. But setting aside such ambitions, the idea of a *disposition*—and related notions such as *ability*—might seem to provide a promising starting point for further investigation.

In fact, Wittgenstein himself, in the sections immediately prior to the discussion of reading, hints that the topics he is discussing might be clarified by some such idea:

> 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)

And 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.

However, a general appeal to dispositions or abilities only gets us so far. To see why, we need only point out that all sorts of dispositions could provide exemplifications of our generic schema:

> $S$ said “…” because of the sign “_____”

In fact, what gave considerable plausibility to the reductive account was that acts of reading seemed to share a great deal with a whole range of other acts that we might broadly classify as “perceptually-based responses”, i.e. cases where some act is caused by something presented in perception. Perhaps it would be a mistake to assume all such responses to be based in dispositions, but surely some are. Cases could range from a tendency to

---

28Note that the idea of mastery of a technique shows up again in §199, on the other side of the remarks on reading. On this and surrounding sections, see below.
flinch when a bright light was shone in one’s eye to tendency to recall a particular memory when presented with a certain symbol.

This suggests that a generic appeal to the notion of a disposition—where that means something like a reliable transition—will be insufficient for our purposes. What is required is a closer examination of the kind of disposition that is involved in acts of reading.

18. The broad or generic notion of a disposition I take as my starting point encompasses almost any reliable transition.29 As such, dispositions of this sort are very easy to find, both within and without the psychological and behavioural fields that are our main focus. Any reliable transition between psychological states, or between such states and some particular behaviour, will count as such a disposition.

This generic notion of a disposition as a reliable transition is often appealed to in the philosophical literature, and has some important features. Most significant, for our purposes, is the idea that acts manifesting that disposition are ‘no accident’: they have their source in some enduring fact about the agent, ensuring that e.g. transitions between mental states that are manifestations of dispositions are “part of the ordinary course of [the agent’s] mental life” (Setiya 2007: 32). 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.30 Appeal to a disposition ensures that this cannot be true of acts characterized by the account.

But the generic notion of a disposition will not do for our purposes (nor, I would argue, for the purposes of such accounts). It lacks some of the fundamental features that characterize

29Some work might be needed to specify what would count as ‘reliable’. But since I am merely providing a sketch of this generic notion, I will not speculate as to how such an account might be provided.

30So, 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. Setiya (2009), particularly his discussion of the case from Freud (pp.33-4) and the example from Nagel (pp.64-5).
an ability to read, and acts that manifest it are not caught up in the same kind of patterns as we find in acts of reading.

First, note that such dispositions provide only minimal normativity. It is true that, based on attribution of such a disposition, we would expect its bearer to behave in particular ways in various circumstances. For instance, if the disposition involves a transition from state A to state B in circumstance C, we know when to expect such a transition. The only normativity here comes from the disposition itself: it provides a standard, in the sense that it leads us to expect certain behaviour in certain circumstances, and to look for an explanation if such behaviour is not forthcoming. And as long as there is some such reliable transition, we have enough to fit the characterization of a generic disposition.

Suppose, for instance, that I had some idiosyncratic associations connected with particular signs, which were reliably triggered whenever I saw them. Recognizing and characterizing such a disposition needn’t involve anyone beyond me: there is nothing incoherent in imagining that I am the only person in the world who will ever have it. And so long as I reliably make the relevant transitions, the generic notion of a disposition seems appropriate.

Note too that this disposition provides only a minimal standard of normativity by which to evaluate my behaviour. It leads us to expect certain transitions, and perhaps to expect some further explanation if they are not forthcoming. (Have I lost the disposition? Was this sign not the right type? Am I distracted? Etc.) But while, given the disposition, my behaviour is to-be-expected, it is not, in any sense, correct.

Introduction of a concept like ‘reading’ into the characterization of the disposition brings with it a new dimension of normativity. For now not any reliable transition between signs and utterances will count as a disposition of the right sort. Only dispositions that involve correct rendering of signs are relevant. So in attributing an ability to read, we are not simply saying that a person has various dispositions to transition between signs and utterances;
and the behaviour we expect is not any old response to the perceptual stimulus. We expect the reader to render signs *correctly*, and expect some explanation if she fails to do *that*.

This appeal to correctness partially explains the need to appeal to the concept *reading* in specifying the relevant disposition. For the standards by which we evaluate the acts of the reader come from our understanding of what it is to read, and in this sense involve a source of normativity that goes beyond the particular generic dispositions of particular agents.

The correct way of reading might be explained in terms of the acts of a competent reader: but in imagining such a reader, we are engaged in an activity that is, in an important sense, independent of this or that particular reader, except insofar as we chose to take some particular reader as an example of what a competent reader can do. In other words, our grasp on what the competent reader does is in one sense prior to our counting particular individuals as competent readers.\(^{31}\)

19. In fact, to properly characterize the dispositions involved in reading we need to see them as belonging to a *practice*. Of course, this idea, along with related notions such as *custom* and *institution*, are explicitly referenced in the sections of the *Philosophical Investigations* following the discussion of reading, particularly §197-§202. So, famously, at §198, Wittgenstein references a practice involving the use of signs to indicate directions:

> Let me ask this: what has the expression of a rule—say a sign-post—got to do with my actions? What sort of connexion is there here?—Well perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react. But that is only to give a causal connexion; to tell how it has come about that we now go by the sign-post; not what this going-by-the-sign really consists in. On the contrary; I have further indicated that a person goes by a sign-post only insofar as there exists a regular use of sign-posts, a custom. (§198)

\(^{31}\)Compare this to a case where we I am trying to pass on my idiosyncratic associations to others. There too I provide a standard by which to judge the associations of others. But this standard involves essential reference to me, as a particular individual, and the particular associations I happen to have. On the priority of the practice over its bearers, as well as its dependence on them, see below.
The parallels with our topic should be clear. What we provide, in offering reading explanations, is not a merely causal connection; we further indicate that there is a practice of reading, that this agent is a bearer of it, and that the particular act in question belongs to this practice. This is part of what was brought out by comparing the generic class of explanations of the form:

\[ S \text{ said } \ldots \text{ because of the sign } 
\]

with the more narrow class of reading explanations:

\[ S \text{ said } \ldots \text{ because } R \text{ of the sign } 
\]

A “merely causal” connection would suffice for an instance of the first kind of explanation; but a reading explanation further indicates that there is a practice of reading, and that this act should be seen as belonging to it.

Thus, while it would not be wrong to class reading explanations with the broader class of explanations that appeal to dispositions, the fact that acts of reading belong to a practice brings with it various differences that set the dispositions involved apart from the more generic notion. Besides the standards of correctness stemming from the practice, there is also the fact that all the particular dispositions of various agents have a common source in the agent’s inculcation into the practice (whereas the source of a generic disposition is irrelevant to its characterization), and the fact that the practice has a certain kind of priority over, and independence from, the dispositions and acts of particular readers.\(^32\)

20. If this is right, it suggests a particular source for the distinctive character of reading explanations. What accounts for the distinctive normative and etiological character of such explanations is that they implicitly rely on appeal to a practice, such that grasp of the

\(^{32}\)In another dimension, however, the practice depends directly on the dispositions and acts of particular readers, since it only acquires actuality through those acts and dispositions. On this, cf:

A practice, however we are to understand such a thing, does not exist except through people’s acting and being disposed to act in accordance with it. And on any account a disposition must have some sort of explanatory standing in respect of what happens when it is manifested in an individual action. (Thompson 2008: 160)
explanatory relations laid out in them fundamentally depends on acquaintance with the practice in question.

This point generalizes to other practices as well. To see this, consider the following three explanations:

1. S said “…” because of the sign “___”
2. S turned left because the arrow pointed in that direction.
3. S φed because she said that she would

I mean each explanation to be taken as follows: (1) is an explanation of an act of reading; (2) is an explanation of an act of sign-following; and (3) is an act of fidelity, i.e. keeping a promise.

Note that (2) and (3) display some of the key features we outlined with respect to reading-explanation. First, seen from outside the practice, the normative, and thus explanatory, relationship is obscured. With reading, this amounted to the claim that we would not be able to see exactly what the relationship was between the written sign and the utterance made by the agent. A similar point could be made in the other two cases. Someone unfamiliar with any practice of signposting 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. (To make this last case more vivid, imagine that everything else seems to speak against the person doing φ.) In each case, 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.

Describing the situation from the perspective of such an outsider brings out further difficulties with our purported explanations. For from such a perspective, certain fundamental
descriptions of both the act and its cause will be unavailable, meaning that the relevant explanation simply could not be understood. For instance, in the case of reading, an outsider from a culture in which there is no writing will not understand the relevant sense in which the marks are a *sign*. Or take the case of sign-following: the outsider will 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 said was keeping that promise.

In each case, certain reason-giving—and thus explanatory—relations, along with certain descriptions of acts and their causes, depends fundamentally on a familiarity with the relevant practice. It is the practice that figures as the source of the normative relations that characterize acts of reading, and as such any example of *reading explanation* will implicitly rely on appeal to some such practice. The practice itself plays an explanatory role. This in turn means that a full understanding of the causal relations and psychological features that explain a particular action also depends on the practice, such that they cannot be characterized independently of it.

21. I have suggested that in *reading explanation* we see a kind of reason-giving explanation in which a practice plays an explanatory role. And I have further suggested that there will be many other such kinds of explanations, more or less similar to *reading explanation* in character, all of which equally depend on an appeal to some practice.

There are at least two relevant points that follow from these claims: first, a distinction between explanations that appeal to dispositions belonging to practices, and explanations that appeal to other kinds of disposition or ability. This introduces an abstract, formal difference between kinds of reason-giving explanation.\(^{33}\) Second, the specific distinction

\(^{33}\)Note that there might also be room for further distinctions between *kinds of practice* at this formal, abstract level. My discussion here is intended to be a development of points made by Michael Thompson in his discussion of the practice of promise-keeping (Thompson 2008: Part III). But, as a moral or ethical practice, promise keeping might seem to have various features that are absent in a practice like reading. In particular, the problem of a suitable transfer-principle between the goodness of the practice and the goodness of acts belonging to it simply does not arise for a practice like reading.
introduced by characterization of a particular practice, and the kind of intelligibility it brings with it. The generic notion of a practical disposition, or a disposition that belongs to a practice, does not by itself provide materials for articulating the specific normative and causal relations that characterize acts of reading. If anything, the general notion of a practical disposition emerges from reflection on particular practices, and the intelligibility that belongs to them.

My final claim is that this suggests that the general question that most contemporary accounts start with—‘what is it to act for a reason?’—is ill-posed, because the broad category of rational explanation admits of all kind of internal differentiation that is fundamental to an appreciation of its overall unity and character. This goes beyond the arguments directed at reductive accounts in the first part of this paper. If the line of thought outlined above is correct, we have reason to expect there to be (1) an important general difference between reason-giving explanations of acts belonging to a practice and other reason-giving explanations, and (2) important local differences between explanations of acts belonging to specific practices. Some of what falls under the general schema of rational explanation falls within the category of practice-based explanation—e.g. “S turned left because the sign for the bathroom pointed that way”. These are genuine and familiar instances of explanation with none of the artificiality that came with my notion of reading explanation.

But many explanations that seem to fit the general schema for rational explanation do not depend on appeal to a practice. Some such explanations involve appeal to other general categories playing an analagous role, e.g. explanations that make their targets intelligible by representing them as characteristic activities of a life-form. For instance, if I say that “The squirrel ran up the tree because it saw the wolf approaching”, I represent sight of the approaching wolf as providing a reason for the squirrel to flee, and represent its act as one of flight. Grasp of these normative relations depends on specific knowledge.
of the lives of squirrels, and via this, more abstract knowledge of the characteristic activities of animal life.

Furthermore, many acts that might be explained as responses to specific reasons can also be represented as part of the pursuit of some specific end. For instance, we might say “\( S \) turned left because he was going to the bathroom”, or “The squirrel ran up the tree because it was fleeing from the approaching wolf”. Explaining an act in this way involves representing it as part of the pursuit of some end, where grasp of the normative and explanatory relations involves both a general grasp of the teleological structure on display, and specific knowledge of the instrumental relations holding between the actions described (e.g. that running up a tree is indeed a plausible way of escaping a wolf).

In each case, the normative relations on display in so called rational explanation, and thus the relevant form of ‘acting for a reason’, are grasped via an understanding of the general category—practice, life-form, intentional action—as well as an understanding of the specific details of the relevant instantiation of that category and the circumstances of the act.

At this point many contemporary philosophers would probably respond that it is one thing to show that all of this is true of normative, reason-giving relations; but that explanatory, causal relations, and the psychology that underpins them, is something else entirely. My aim, in focusing on the way in which a generic notion of a disposition has to be modified to apply to cases of reading, was to show that this line of thought is also suspect. The argument about dispositions shows that the explanatory relations, and the psychological and behavioural characterizations that underpin them, cannot be grasped independently of the normative relations laid out above. Whether the relevant underpinning is a disposition (like desire or intention, which might be understood as a disposition towards actions that further a particular end, a practice-related ability (like the ability to read), or a virtue (like fidelity), the characterization of the relevant aspect of our psychol-
ogy cannot be made in terms that do not depend on the relevant normative category. That means that the points made about normative relations cannot be separated from claims about explanatory relations.

All of this suggests that there is little hope for a completely general account of ‘what it is to act for a reason’ or of rational explanation. What falls under these general headings will be a family of more or less related cases, albeit cases that admit of articulation into abstract categories, as well as articulation in terms of various kinds of specific differences. An account of what is involved in a particular kind of reason-giving or rational explanation will involve attention to the details of specific cases, and the concepts those cases provide for articulating the relevant forms of normativity. And that, I hope, is a conclusion that is broadly in line with the Wittgensteinian considerations with which we started.
References


