Self-Knowledge in Consciousness

Conor McHugh

Submitted for the degree of PhD by Research

The University of Edinburgh

2008

I have read and understood the University of Edinburgh guidelines on plagiarism. This thesis was composed by me and is entirely my own work except where I indicate otherwise by use of quotes and references. No part of it has been submitted for any other degree or professional qualification.

Conor McHugh

Acknowledgements

I would like to thank my primary supervisor, Matt Nudds, and my secondary supervisor, Jesper Kallestrup, for more guidance and feedback than I can remember.
My thanks also to Chris Peacocke, who generously facilitated a three-month visit to Columbia University in 2007, and with whom I enjoyed a number of very helpful discussions.
Matthew Chrisman and Dorit Bar-On both took the time to read drafts of Chapter Two and provide extremely detailed comments. Thanks to them.
Many staff and postgraduates at Edinburgh have, knowingly or otherwise, contributed to the development of this thesis—sometimes in discussion after a seminar presentation, but more often in informal chats. I am grateful to all of them. To name everyone would be impossible, but Ezio Di Nucci is due particular thanks.
I was enabled to pursue my PhD by AHRC doctoral award 2004/107894, and by a scholarship from the College of Humanities and Social Sciences at Edinburgh University.
In getting through the process of writing this thesis, I could not have done without the extraordinary personal support of my parents, Brianne and Reggie, and of Lisa McKeown.

Abstract

When you enjoy a conscious mental state or episode, you can knowledgeably self-ascribe that state or episode, and your self-ascription will have a special security and authority (as well as several other distinctive features). This thesis argues for an epistemic but non-introspectionist account of why such self-ascriptions count as knowledge, and why they have a special status.

The first part of the thesis considers what general shape an account of self-knowledge must have. Against a deflationist challenge, I argue that your judgments about your own conscious states and episodes really do constitute knowledge, and that their distinctive features must be explained by the epistemic credentials that make them knowledge. However, the most historically influential non-deflationist account—according to which such self-ascriptive judgments are based on introspective experiences of your conscious states and episodes—misconstrues the unique perspective that you have on your own conscious mind.

The second part of the thesis argues that the occurrence in your consciousness of a state or episode of a certain type, with a certain content, can itself suffice for you to have a reason to judge that you are enjoying a state or episode of that type, with that content. Self-ascriptions made for such reasons will count as knowledge. An account along these lines can explain the special status of self-knowledge.

In particular, I show that a self-ascription of a *content*, made for the reason you have in virtue of entertaining that content, will be true and rational, partly because it is an exercise of a general capacity, which I call "grasp of the first-/third-person distinction", that is fundamental to our cognition about the world. A self-ascription of a particular *type* of conscious state or episode, made for the appropriate reason, will be true and rational in virtue of features distinctive of states or episodes of that type—features that contribute to determining which judgments are rational for a subject, without themselves being reasons that the subject has. I consider in detail the cases of perceptual experience and of judgment.

The thesis concludes by arguing that this kind of account is well placed to explain how self-knowledge fulfills its central role in the reflective rationality that is characteristic of persons.

TABLE OF CONTENTS

CHAPTER 1

THE PROBLEM OF SELF-KNOWLEDGE

1.1 Self-knowledge: the subject-matter

Each one of us has a special kind of knowledge of his or her conscious mind. In particular, each of us typically knows about his or her current conscious mental states and episodes—the particular states and episodes that make up his or her stream of consciousness. I now know that I am perceiving that (or apparently perceiving that) there is a computer screen in front of me. I know that I judge that it is windy outside. Self-knowledge of this sort is the topic of this thesis.

I said that self-knowledge is special, but in what way is it special? What is known, in both of the examples mentioned above, is a contingent, empirical fact. Yet I know these facts about myself in a different way to the way I can know about contingent, empirical facts in other domains—for example, facts about my environment (that it *is* windy outside), facts about other people's conscious minds (that *you* judge that it is windy outside), and certain other facts about myself (that I am six feet tall). I do not need to do anything to find out about my current conscious states and episodes. Despite not doing anything to find out about them, I am almost never mistaken about them. And my self-ascriptive assertions are typically not open to challenge from anyone who could claim to know better. In each of these respects, self-knowledge contrasts with knowledge in other domains.

The question addressed by this thesis is: how do we come to have knowledge, with this special status, of our own conscious mental states and episodes?

I will sharpen that question into a more precise philosophical problem in section 1.3, below. Doing so will depend on achieving a fuller characterisation of the specialness of self-knowledge, which I will attempt in section 1.2. In the remainder of the present section I want to make clear precisely what phenomenon is my explanatory target.

Let us start with examples. The following knowledge-states, when arrived at in the usual way, fall into the category that I will discuss. In each case, the state or episode known about should be understood as consciously occurrent.

- your knowledge that you are seemingly seeing that the leaves on a particular (visually presented) tree are yellow;

- your knowledge that you judge that the leaves on the tree are yellow;
- your knowledge that you are wishing you had gone out for a walk this morning;
- your knowledge that it strikes you that you need to be home soon;
- your knowledge that you decide to take the short-cut home;
- your knowledge that you are supposing that the leaves were green.

The following cases of self-knowledge are among those *not* included in this category:

- your knowledge that you are short-tempered or have some other personality trait;
- your knowledge, arrived at through psychoanalysis, that you have a certain repressed belief or other attitude;
- your knowledge that your going home instead of going to the party is caused by your anxiety about the possibility of seeing a certain person.

These latter cases do not count because the property self-ascribed does not necessarily consist in the occurrence of a conscious episode or state.

So much for particular examples. Let me now characterise self-knowledge of conscious mental states and episodes in more general terms.

By "conscious mental states and episodes" I mean occurrences in consciousness, not dispositional states whose manifestations include such occurrences. Thus, I talk about self-knowledge of judgments rather than of beliefs, and self-knowledge of conscious wishes rather than of desires. I will not discuss self-knowledge of beliefs, desires or intentions.

The knowledge I am discussing is propositional knowledge—knowledge that something is the case. It is not knowledge by acquaintance; that is, it is not mere awareness *of* your mental states and episodes. On some views, the propositional knowledge that you are enjoying a certain state or episode is epistemically based on an experiential awareness of that state or episode. But even if that view is correct (I will argue in Chapter 3 that it is not), the two are not identical.

It is present-tense. It is self-knowledge of *current* states, episodes and attitudes. Knowledge of what you were thinking or experiencing five minutes ago does not count.

The mere fact that a judgment (or belief)[1] first-personally self-ascribes a current conscious

1 I will talk about the judgments that constitute, or manifest, self-knowledge, rather than beliefs that constitute self-knowledge. Anyone suspicious of talk of 'knowledgeable judgments' can understand the phrase as referring to those judgments that are the conscious manifestations of beliefs that are

state or episode doesn't guarantee that it is an instance of self-knowledge of the special kind. A self-ascription may fail to count because, even though the property self-ascribed is one about which you would usually have self-knowledge of the right sort, the self-ascriptive judgment, or the way in which it is made, is unusual in some way. For example, self-ascriptions reached by observations of your own behaviour, or by accepting the testimony of your psychoanalyst, don't count. Thus, the category of self-knowledge that I am discussing is individuated in part by the way in which the knowledge is acquired.

I will be limiting my discussion to self-knowledge of states and episodes with intentional content. Conscious propositional thoughts uncontroversially fall into this category. I will also be assuming that perceptual experiences have intentional content—that they can be at least partly characterised in terms of what contents they represent as true to their subjects. I will not discuss self-knowledge of sensations or of pains; nor will I discuss emotional states.

Although some of the states and episodes I consider have a phenomenal character, I will not discuss the subject's knowledge of that phenomenal character. I will not discuss, for example, how a subject knows what a particular perceptual experience is like for her.

The knowledge I am discussing thus has a structured content with three conceptual components. It is knowledge that *I* am enjoying a certain *type* of conscious mental state or episode, with a certain *content*. It involves, firstly, the first person concept, 'I' (and not some other way of thinking of oneself, such as 'the person in the mirror'). Secondly, it involves the concept of a certain type of mental state or episode, M. And thirdly, it involves the concept of a certain intentional content C.

A full explanation of the knowledgeable status of a self-ascription will thus have three parts, answering three different questions. How does the subject know that *she*, rather than someone else or nobody, is in M with content C? How does she know that her state or episode is of type *M*, rather than M*? And how does she know that the content of her state or episode is *C*, rather than C*?[2] There is no guarantee that all three of these questions can be answered together. What's more, it may be that the second component requires different

---

themselves knowledge. See below, section 1.4.

2 The fact that we can think of a piece of knowledge as being compositional in this way does *not* mean that it is based on inference. The self-ascription 'I M that C' is not ordinarily arrived at by inference from three distinct propositions, 'I am enjoying a mental episode E', 'E is of type M', and 'E has content C'. You can know a complex proposition without that knowledge being inferred from (or otherwise based on) knowledge of simpler propositions. Consider the perceptual judgment that the cat is on the mat. Such a judgment is not based on inference: you simply see the cat on the mat, and take the content of that experience at face value. Nevertheless, we can ask: how you know that the *cat*, and not a cat-hologram, is on the mat; how you know the cat is on the *mat*, and not some disguised floorboards; and how you know that the cat is *on* the mat, not beside it.

explanations for different types of mental state. Indeed, the account I will offer (Chapters 5-6) does explain the different components, and different instances of the type component, somewhat differently.

I hope that it is now clear what the explanatory target of this thesis is. For the rest of the thesis, when I use the terms "self-knowledge" or "self-ascriptions", I am referring only to those self-ascriptive judgments (and beliefs) that fall within the range I just demarcated. In the next section I will try to say what is special about this sort of self-knowledge.

1.2 The specialness of self-knowledge

To appreciate the problem of self-knowledge, we must have a grasp of the way in which self-knowledge is special. This is not a straightforward matter. Unadulterated pretheoretical intuitions and observations are in short supply; and philosophers have disagreed over how self-knowledge is different from knowledge of the weather, of other minds, or of one's height. Nevertheless, I think we can point to a number of distinctive features that self-knowledge, uncontroversially, *appears* to have, and in virtue of which it appears to contrast with knowledge in other domains. I do not claim, yet, that it *has* all these features, but I do claim that even a philosopher who denies one of these features ought to accept that the appearance (illusory, as such a philosopher would claim) of the feature is real.

My characterisations, below, of these apparent features are not supposed to represent the opinion of pretheoretical intuition. They are infected by philosophical and phenomenological reflection. But this is not in itself a bad thing, as long as that reflection is not committed to any particular controversial theory.

In characterising the distinctive features I will talk as though self-ascriptions do indeed have these features. This is just for ease of exposition. My claim is only that they appear to have them.

(a) *Security*

Sincere present-tense self-ascriptions of conscious states and episodes do not easily go wrong. It is often claimed that knowledge, generally, is modally *safe* (see, e.g. Pritchard, 2005). Self-knowledge has an especially high degree of safety: it is *secure*.

When you make a judgment, that judgment is safe iff you could not easily have come in the same way to judge as you did and been wrong. For example, a perceptual judgment is safe iff

there are no sufficiently similar circumstances in which perception would have led you to make the same judgment falsely. Safety comes in degrees: a perceptual judgment can be safe enough to count as knowledge, even though there are circumstances in which you would have gone wrong in that judgment. For example, suppose that, in perfectly normal circumstances, you judge, based on a visual experience, that there are yellow leaves in front of you. Had you been a brain in a vat with indistinguishable experiences, you would have made the same judgment on the same basis[3] and been mistaken. Less radically, had you been the victim of a perceptual illusion, you would have gone wrong. These scenarios don't threaten the judgment's status as knowledge, because they are relatively modally distant.

The safety of self-knowledge appears to go far beyond that of perceptual knowledge. Sceptical scenarios do not seem to threaten it. When you self-ascribe a particular experience or thought, typically you would not have gone wrong *even if* you had been a brain in a vat—in the brain-in-a-vat scenario your experiences and thoughts come apart from the environmental facts,[4] but your self-ascriptions do not come apart from your experiences and thoughts. Nor could you go wrong by being the blameless victim of an illusion—if there is an analogue, for self-knowledge, of perceptual illusion, it involves serious and bizarre cognitive defects in the subject.[5] Barring such defects or cognitive lapses, it is difficult to envisage circumstances in which you would sincerely, but mistakenly, judge that you are currently having a seeming perception that the foliage is yellow, or that you currently judge that it is windy outside.

Security applies to all three components of self-knowledge. You could not easily be mistaken about *who* (if anyone) is enjoying a particular conscious state or episode, when you judge that you are enjoying it.[6] Secondly, you could not easily be mistaken about the *type* of state or episode that you are enjoying. If you self-ascribe the *judgment* that *p*, you could not easily have misidentified as judgment a *doubt* that *p*, a *wish* that *p*, or whatever. And finally, you could not easily be mistaken about the *content* of your mental state or episode. When you self-ascribe the judgment that it is windy, you could not easily be mistaken in virtue of the

3 Whether your basis is the same, or relevantly similar, in the brain-in-a-vat case as compared to the normal case, depends on how we conceive of the basis in the normal case. It doesn't matter for my purposes.

4 At least, they come apart in certain brain-in-a-vat scenarios, such as the scenario in which you have only recently been envatted, having previously been embodied in a normal environment.

5 See Burge (1996).

6 It has been claimed (Shoemaker, 1968; Evans, 1982) that self-ascriptive judgments, made in the right way, are immune to error through misidentification relative to the first person—they can't constitute knowledge that someone is enjoying a conscious state or episode, but be mistaken about who that person is. My claim here is just the weaker one that the first-person component of self-ascriptions is secure.

fact that you judge that it is sunny.[7] For none of the three components is there a sceptical scenario in which you would have gone wrong.

Security should not be confused with infallibility. Cognitive lapses occur. Perhaps errors can be made even in the absence of such lapses. The point is that these errors are not at all easily made, not that they are impossible.

(b) *Salience*

Salience is the converse of security. When you enjoy a conscious state or episode, you are exceptionally likely, modally speaking, to be in a position to know about it. You could not easily be enjoying a conscious perceptual experience that represents the foliage as yellow, or consciously judging that it is windy, and yet be unable to answer the question of what you are currently experiencing or thinking (assuming you have the conceptual abilities to answer it). You could not easily be 'self-blind' with respect to any particular conscious episode.[8]

Each of the three features of a conscious state or episode, corresponding to the three components of a self-ascription, is salient. You would not easily fail to know, of a conscious state or episode, that it is yours rather than anyone else's. Only in outlandish circumstances could you sensibly wonder, "Someone is judging that it is windy, but is it I?". You would not easily be blind with respect to what type of state or episode it is ("I am enjoying an episode with the content that it is windy, but am I *judging* that it is windy, or merely *supposing*?"). And you would not easily be blind with respect to its content ("I am making a judgment, but am I judging that it is windy, or that it is sunny?").

Salience does not imply that you *cannot* fail to know, if the question arises, what conscious state or episode you are enjoying. The claim is only that such failures are, usually, modally very distant.

(c) *First-person privilege*

The way in which you come to know that you are enjoying a conscious state or episode is a way that is available only to you, and it is available to you only with respect to your own

7 It has been claimed (Davidson, 1987; Burge, 1996) that certain self-ascriptions *cannot* go wrong about the content of the ascribed state or episode, because of a constitutive relation between the content of the self-ascription and that of the ascribed state or episode. I will touch on that claim in Chapter 2 (section 2.2), and consider self-knowledge of content more fully in Chapter 5. My present concern is only to claim that the content component of self-ascriptions is secure.

8 See Shoemaker (1994) for a fuller discussion of the impossibility of self-blindness.

conscious states and episodes. Each of us must, when ascribing conscious states and episodes to another person, base our judgment on the other person's behaviour or utterances, or on other evidence about the person. But each of us can come to know about his or her own mind in a quite different way, without relying on such evidence.

Thus, self-knowledge is not just knowledge about *facts* that happen to concern yourself, but *knowledge* of a sort that you have only about yourself. Only a subversion of the normal relation between you and your conscious states and episodes, such as by neural cross-wiring, could plausibly undermine first-person privilege.

Again, first-person privilege applies to all three components of self-knowledge. You know that *you* are enjoying a particular state or episode, you know what *type* of state or episode you are enjoying, and you know the *content* of that state or episode, in a way unavailable to others.

(d) *Authority*

By 'authority' I mean certain features of the way self-ascriptive utterances are, and ought to be, treated by interlocutors.

It is almost always inappropriate to raise a doubt over the truth of, or demand justification for, a self-ascriptive utterance within the range I demarcated. If you say, “It now visually appears to me that the foliage is yellow”, or, “I judge that the foliage is yellow”, it would ordinarily be inappropriate for your interlocutor to say either of two things: “No, you don't”, or “How do you know?”. It is part of our practice to accept such self-ascriptive utterances as true, without challenge. We do not expect the self-ascribing subject to be able to justify their self-ascriptive judgment, in the sense of offering some *further* consideration to justify their self-ascriptive claim. By contrast, if you say, “The foliage is yellow”, there is nothing in-principle unusual or wrong about an interlocutor replying, “No, it is green”, or “How do you know it is yellow?”. And we might expect you to be able to respond to the latter by saying something like, “I can see it to be yellow”, or, “That's how it looks from here”, thus justifying your claim that it is yellow.

Once again, authority holds for all three components. Each of the following replies to a self-ascription is normally inappropriate: “How do you know it is *you* that is seemingly seeing that the foliage is yellow?”, “How do you know that you are seemingly *seeing*, and not imagining, that the foliage is yellow?”, and, “How do you know that you are seemingly seeing *that the foliage is yellow*, and not seemingly seeing that it is green?”.

(e) *Immediacy*

When you make a knowledgeable judgment about another person's mental state, or about the colour of the leaves on a tree, you reach the judgment by acquiring or attending to grounds—even if those grounds consist only in the visually apparent colour of the leaves. For almost any contingent empirical fact, coming to know that fact involves a process of observation, of inference, or of some other manner of *finding out* what is the case. This finding out need not be a matter of inference; it may just be a matter of looking, and accepting what you see.

Self-ascriptions are different. There is no effort of acquiring, or attending to, grounds for judgment about your own conscious states and episodes. You seem to be able to know about those states and episodes just in virtue of enjoying them; once you have an experience or thought, you are in a position to know about it without having to do anything else. And when you come to make a self-ascriptive judgment, it would be deeply wrong to say that the judgment expresses something you have found out about yourself. Self-knowledge is not a discovery.[9]

The claim is not merely that self-ascriptive judgments are not arrived at by inference. It is that they are not arrived at by *any* process of acquiring or attending to grounds, inferential or otherwise.

We must distinguish the question of whether self-ascriptions are arrived at by acquiring or attending to grounds, from the question of whether they are epistemically based on grounds. Immediacy is a feature of the way in which you come to make self-ascriptive judgments, rather than of the epistemology of those judgments. Thus, I am not making any claim about the nature of the grounds, if any, for self-ascriptions. Nor am I claiming that self-ascriptions are epistemically groundless, as some have done (e.g. Wright, 1998). I will argue in the next chapter that that claim is false.

Again, this feature applies to all three components of self-ascriptions. You do not acquire or attend to grounds concerning *who* is enjoying a particular state or episode, concerning what *type* of state or episode it is, or concerning the *content* of that state or episode.

My claim, then, is that self-knowledge *appears* to have the five features of security, salience, first-person privilege, authority, and immediacy. In later chapters certain other features that

9 Velleman (1989) emphasises this point about self-knowledge of action.

self-knowledge appears to have, and refinements in the characterisations of the five features above, will emerge.

The characterisation I have offered of what appears distinctive about self-knowledge is meant to be relatively weak and uncontroversial. It was allegedly claimed by Descartes that self-ascriptive judgments, within a certain range, are infallible—that they cannot be mistaken. This seems to underestimate "the scope for human perversity", as Burge (1996, p. 96) puts it. You can, arguably, mistake wishing that something were the case for judging it to be so; and you can mistake the content of an experience, a thought, or whatever, for a closely related content.[10] Indeed, it seems that you can be mistaken even about pains. A friend of mine has told me about a childhood incident in which he and some other boys were playing on a hillside. One boy tumbled down the hillside and, on the way, struck his head on a rock. Holding his hand to his bleeding head, and in distress and pain, he believed that he had wounded his hand rather than his head. He was, it seems, mistaken about the location of his pain.[11] If we can be grossly mistaken about the locations of pains, we can surely be mistaken about almost anything.

It might also be claimed that it is impossible to enjoy a conscious state or episode without knowing, or being in a position to know should the question arise, that you are enjoying it. Williams (1978) attributes this claim to Descartes. Again, however, it seems at least conceivable that, due to confusion, pain, or whatever, you would fail to know about your current conscious state, or about some aspect of it.

My claim is thus relatively modest. What's more, I am suggesting that explaining apparent features (a)-(e) is a desirable feature in an account of self-knowledge, not that it is a *sine qua non*. The view I will develop (Chapters 5-6) in fact accounts for all of them, or so I will claim.

1.3 The philosophical problem

1.3.1 The nature of the problem

We are now in a position to state the philosophical challenge that this thesis takes up, and to appreciate why that challenge constitutes a philosophical problem.

10 See Martin (1998, p. 116) on mistakes about content. Note that these cases do involve some sort of lapse or irrationality on the part of the subject.

11 Thanks to Tom Roberts for the example.

The challenge is an explanatory one, of answering two related questions about self-knowledge:

The knowledge question Why are self-ascriptive judgments in the range I demarcated, when made in the usual way, typically *knowledge*?

The specialness question Why does this self-knowledge apparently have the distinctive features (a)-(e)?

The distinctive features of self-knowledge, as well as constituting an explanatory target in themselves, make this challenge problematic. They make it difficult to see how either part of the challenge could be satisfactorily answered.

Let me take the specialness question first. The contents of self-knowledge—that I am currently thinking that *p*, say—are empirical and contingent.[12] They are not analytic; nor are they the kinds of things that could be knowable *a priori*. Thus, it seems that self-knowledge—unlike, say, analytic knowledge—must depend on cognitively hooking up to empirical facts. But this is mysterious. How can we hook up to these facts with such extraordinary success, when we do not have comparable success in any other domain? Why can't the mechanism for hooking up to these facts go wrong as easily as that for hooking up to facts about the colour of foliage? Why do so few facts escape its grasp? And how can hooking up to these facts give rise to knowledge that is immediate: doesn't hooking up to a fact necessarily involve acquiring grounds?

The knowledge question is equally problematic. When a judgment constitutes knowledge, there is something in virtue of which it is knowledge. Ordinarily, if the judgment is not analytic, there is some justification or grounds, on which the judgment is based, in virtue of which it is knowledge. But self-ascriptive judgments are not analytic, and their authority and immediacy seem, on the face of it, to rule out justification or grounds: subjects cannot offer justifications for their self-ascriptions, and do not arrive at them by consulting grounds. But

12 As noted at the end of section 1.1, by 'self-knowledge' I mean knowledge of propositions of the form 'I M that C', where 'M' picks out a type of conscious state or episode. There are other propositions, that can be contents of self-knowledge understood more broadly, and that are arguably not contingent—such as the proposition that the pain I now have hurts.

that leaves us with a hole where we wanted an *explanans*. We seem to be driven to say that self-ascriptions constitute knowledge in virtue of nothing.[13]

I do not claim, of course, that these problems can't be solved. But solving them involves rejecting some of the assumptions that I used in formulating them, and it is not obvious how that can be done.

I hope I have said enough to show that there is a determinate philosophical task here, and to make clear what that task is. McDowell (1998) has expressed doubt about whether there is a philosophical problem of self-knowledge. Responding to an article of Crispin Wright's, McDowell argues that it takes more than merely pointing to a phenomenon, such as self-knowledge and its distinctive features, and asking that it be explained, to pick out a real philosophical task:

> "'How is it possible that…?' … is indeed a good way to express philosophical difficulties of a familiar kind, and some such difficulties may be worth tackling. Whether that is so depends on the specifics of the case. If a question of that shape is to express a determinate philosophical difficulty, it must be asked from a frame of mind in which there is at least a risk of its looking as though whatever the question is asked about is *not* possible." (McDowell, 1998, p. 57.)

McDowell goes on to suggest that there is nothing gripping about the idea that immediate[14], non-observational self-knowledge is not possible, except within the context of an objectionable Cartesian conception of the mental. Once we reject that conception, no determinate philosophical problem of self-knowledge remains.

If McDowell is right, then there is no philosophical task for the present thesis to address; it is framed in confusion. But I do not think he is right, for several reasons.

Firstly, I have argued that there *is* a risk of its looking as though self-knowledge, with apparent features (a)-(e), is impossible, and the argument did not depend on a Cartesian conception of the mind. It depended on the thought that being reliably right about contingent empirical facts requires a mechanism for cognitively hooking up to those facts, a mechanism that, in the case of self-knowledge, would have to be almost supernaturally insulated from

13 My formulation here is reminiscent of Boghossian's (1989) concern, that if we do not know our own mental states by observation or by inference, we must know them "by nothing". I have not relied on Boghossian's trilemma of 'observation, inference or nothing', but I am in similar territory. Of course, the solution to the problem will involve the rejection of that trilemma, and of the inference from authority and immediacy to the conclusion that there is nothing in virtue of which self-ascriptions are knowledge.

14 McDowell, following Wright, says "baseless" rather than "immediate". I think that is a mistake (see above, section 1.2), but the present point doesn't turn on it.

malfunctioning. And it depended on the thought that, if a judgment is knowledge, and is not analytic, the judgment is justified by grounds the subject has acquired. Neither of these thoughts is obligatory; indeed, I will eventually want to reject them. But both of them have a *prima facie* attraction that does not depend on a Cartesian conception of the mental.

Secondly, something's needing an explanation does not require its seeming to be impossible in any very strong sense. Of course, what counts as an explanation is relative to our concerns, and giving such an explanation is an interesting philosophical task only if it is not obvious what the correct explanation is. But there is a perfectly intelligible move of demanding an explanation for something one lacks an explanation of. I can intelligibly ask why some judgment constitutes knowledge, or why some type of action is morally wrong, without being even tempted to think that knowledge, or moral wrongness, is impossible in the case in question; all that's required is that I don't see why knowledge or wrongness is in fact instantiated in this case.

Thirdly, the mere appearance of a need for explanation is sufficient for a philosophical problem. It may be that the correct response is not to offer an explanation, but to show that the appearance of such a need is illusory—that the supposed *explanandum* is an illusion, or that some assumption in virtue of which it seemed to require an explanation is mistaken. But showing this is itself a philosophical task.

I conclude that there is a genuine philosophical problem of self-knowledge. I will take it, from now on, that the formulation I offered at the start of this section is a sufficient initial characterisation of that problem. Now I want to lay some foundations for subsequent chapters by distinguishing at a very general level between different approaches to the problem.

1.3.2 Approaches to the problem

An explanatory problem can be approached in either of two ways: it can be solved, by providing a satisfying explanation, or it can be dissolved, by showing that the appearance of a need for explanation is illusory. The natural way to show that the appearance of a need for explanation is illusory is to show that the apparent *explanandum* is illusory or has been importantly mischaracterised. Since our explanatory problem has two parts, corresponding to two *explananda*, two questions arise. First, are self-ascriptive judgments in the relevant range really knowledge? Second, are these judgments really special, roughly as characterised by the features (a)-(e)? If these questions are both answered in the affirmative, a third

question arises: are the distinctive features (a)-(e) to be explained by appeal to the epistemic credentials in virtue of which the judgments are knowledge, or do at least some of them have a non-epistemic explanation? We can classify approaches to the problem of self-knowledge according to their answers to these three questions.

(i) Approaches on which self-ascriptions are not knowledge

Strong Deflationism

A strongly deflationist account has it that self-ascriptive judgments do not constitute knowledge about one's own mental states and episodes, and do not have anything like features (a)-(e) as I characterised them. It also involves some account of why self-ascriptions appear to be knowledge and appear to possess features (a)-(e). The expressivist account that I will consider in Chapter 2, often attributed to Wittgenstein, is strongly deflationist.

Epistemic Deflationism

Epistemic deflationism is the view that self-ascriptive judgments do not constitute knowledge about one's own mental states and episodes, but do have features (a)-(e) roughly as I characterised them (*modulo* any parts of those characterisations where I assumed or implied that self-ascriptive judgments *are* knowledge). This view can be found in Bar-On and Long (2001), and in Jacobsen (1996).

In Chapter 2 I will offer a general argument that self-ascriptions of conscious states and episodes often constitute knowledge, thus ruling out both strong and epistemic deflationism in one go.

(ii) Approaches on which self-ascriptions are knowledge

If one accepts that self-ascriptive judgments constitute knowledge, one may still deny that they possess features (a)-(e). This is the view expressed by Ryle (1949) in *The Concept of Mind*. Ryle suggests that self-ascriptive judgments differ from judgments about others' mental states only in so far as you are in an especially good position to gather behavioural evidence about your own states, because you are always around to observe your own

behaviour.

Ryle's view is overwhelmingly rejected these days, and I will not treat the approach it exemplifies as a genuine option. It flies in the face of the data to claim that self-knowledge is genuine knowledge about one's mental states and episodes, but is not fundamentally different from knowledge of others' mental states and episodes. I will assume that, if one holds that self-ascriptions are knowledge, one also holds that they are special—that they have something like features (a)-(e), or at least most of them, roughly as characterised above.

This still leaves the third question I mentioned above: are features (a)-(e) to be explained in epistemic or non-epistemic terms?

Weak Deflationism

I term 'weakly deflationist' any view according to which self-ascriptive judgments are knowledge, and have (roughly) features (a)-(e), but the explanation of at least some of features (a)-(e) is independent of the explanation of why self-ascriptions are knowledge. On such a view, a significant part of the specialness self-knowledge, as characterised by (a)-(e), is not to do with its having exceptional epistemic credentials, but to do with some other, independent features of self-ascriptive judgments—for example, their expressive role. Bar-On (2004) can be read as offering a weakly deflationist account, as I will discuss in Chapter 2.

A weak deflationist may have to offer a somewhat revisionist characterisation of those elements of (a)-(e) that she claims are non-epistemic. I characterised them in epistemic terms. The weak deflationist would hope to capture what is correct about the appearances I described, while showing that at least some of those features are non-epistemic after all.

Epistemic Approach

A wholly non-deflationist approach would hold that self-ascriptive judgments constitute knowledge, that they have features (a)-(e), and that those features are to be explained by appeal to the epistemic credentials in virtue of which the judgments are knowledge. Such an approach offers a unified 'epistemic solution' to the problem of self-knowledge. It holds that self-knowledge is distinctive because it has distinctive epistemic credentials. The historically best known epistemic solution to the problem is introspectionism. I will reject introspectionism in Chapter 3. I will offer a non-introspectionist epistemic solution in

Chapters 5-6.

In sum, then, the available options include the epistemic approach, and a number of alternative approaches that are deflationary to various degrees. The epistemic approach aims to solve both parts of the problem of self-knowledge using the same resources. The weakest sort of deflationist approach holds that the two parts of the problem of self-knowledge are to be solved, but using independent resources. The epistemically deflationary approach holds that the second part of the problem of 'self-knowledge'—why self-ascriptions have features (a)-(e)—can be solved, but that the first part—why they are knowledge—should be dissolved. The strongest sort of deflationary approach aims to dissolve both parts of the problem.

1.4 Epistemological matters

My solution to the problem of self-knowledge will be an epistemic one. Although general epistemological questions are not the topic of the thesis, they will inevitably intrude. In giving an account of why some range of judgments constitutes knowledge, one must draw on a particular understanding of epistemic notions such as knowledge and warrant. I will discuss these notions directly in Chapter 4, when I outline a conception of a certain sort of warrant, before arguing that a warrant of that sort is operative in ordinary cases of self-knowledge. In the meantime, however, I will be relying on certain epistemological assumptions, which I hope are relatively uncontroversial. I want to state those assumptions explicitly now.

First, I have been talking, and will continue to talk, as though the 'unit' of self-knowledge is the self-ascriptive judgment—a conscious mental act—rather than the self-ascriptive belief that the judgment manifests—a mental state. That is, I will talk as though self-ascriptive judgments themselves constitute knowledge. This is in keeping with a tradition in the literature on self-knowledge, which has tended to focus on the special status of self-ascriptive judgments, often referred to as "avowals". Anyone who finds it objectionable to suppose that a judgment constitutes knowledge should understand that locution as shorthand for: the belief that the judgment manifests is knowledge.

I do assume that judgments themselves are open to epistemic assessment: if a judgment constitutes knowledge, or manifests a knowledge-constituting belief, it must be possible to explain why the judgment has or manifests that epistemic status. This does *not* entail that,

generally, in order to know that *p*, you must make a judgment that *p*. You can have a belief that constitutes knowledge without that belief ever being manifested in a conscious judgment. Perhaps there are knowledge-constituting tacit beliefs that *cannot* be manifested in a conscious judgment. Self-knowledge, however, is not typically tacit in that way: although you rarely bother to make self-ascriptive judgments, you are almost always in a position immediately to make them (provided you have the conceptual ability to do so). When you make such a judgment, the judgment itself has a positive epistemic status. That is the epistemic status I want to explain in this thesis.

Perhaps some judgments do not manifest beliefs; if so, and if belief is necessary for knowledge, then those judgments will never constitute or manifest knowledge. The fact remains that many judgments do manifest beliefs, and do constitute or manifest knowledge.

Second, I will assume that the primary epistemic condition whose fulfillment is necessary for knowledge is *warrant*. Following Burge (1993), I take *justification* and *entitlement* to be species of warrant.[15] Explaining why judgments in some range—self-ascriptive judgments, for example—are knowledge thus involves explaining why those judgments are warranted.

Third, when a judgment is warranted, there is something in virtue of which it is warranted. I will call that in virtue of which a judgment is warranted its 'epistemic credentials'. I introduce this term merely as a placeholder, but we can say some things to elucidate the notion.

We must distinguish between a judgment's being one for which there is a warrant, and a judgment's being warrant*ed*. That is, a judgment can lack epistemic credentials, and therefore not be warranted, even though there is a warrant for that judgment.[16] Thus, you might have excellent meteorological evidence that it will be windy today, and thus have a warrant to judge that it will be windy, and yet judge that it will be windy today because your horoscope suggests it. Your judgment, in that case, lacks epistemic credentials, even though there is a warrant for it.

What provides the warrant for a judgment will be some feature of the subject's situation or character—good evidence, a perceptually apparent fact, a virtuous cognitive practice, or whatever. What makes a judgment warranted, as opposed to merely being a judgment for which there is a warrant, is that the judgment is connected in the right way to whatever

15 Thus, I am not using the term 'warrant' in the sense coined by Plantinga (1993), who stipulates that warrant is the feature that, when added to true belief, yields knowledge.

16 As it is sometimes put: you can have propositional justification without doxastic justification. I do not assume that your warrant must take the form of a justification.

provides the warrant. It is this connection that constitutes the epistemic credentials of a judgment.[17]

It might be objected to this assumption that the immediacy of self-knowledge shows that some judgments can be knowledge even though they lack epistemic credentials. But this objection would be mistaken. What the immediacy of self-knowledge shows is that some judgments can be knowledge even though they are not reached by acquiring or attending to grounds, and do not express a discovery on the part of the knower. The correct conclusion to draw is that epistemic credentials are not always a matter of coming to judge by basing on observational or inferential evidence, or other grounds that are acquired or attended to. Of course, it is a philosophical challenge—one of the challenges addressed by this thesis—to give an account of epistemic credentials that is compatible with immediacy.

Fourth, I will be assuming that the warrant for self-knowledge involves rationality. That is, I will be assuming that a subject's knowledgeable self-ascriptive judgments are judgments that the subject is rational in making. There is a legitimate question, which I take up in Chapter 4, of what it *is* for a judgment to be rational; but for now I can rely on the intuitive notion of rationality. I take it that, if a judgment is the accepting of a content that the subject has reason to think true, then it is, to that extent, rational; on the other hand, if accepting the content is a mere leap in the dark from the subject's point of view, the judgment is, to that extent, not rational.

I do not claim that *all* knowledgeable judgments meet this condition—perhaps there is 'merely reliable' knowledge that does not involve rationality.[18] But our knowledge of our own conscious minds does not seem to be of the merely reliable sort; judgments about current conscious states and episodes are a paradigm of judgments that subjects are rational in making. When you make such a judgment about your current experience or thought, that judgment is not a leap in the dark from your own point of view.

If defence of this is needed, consider the role that self-ascriptive judgments can play in rationalising other mental acts. Often, when you judge that you have come to make a judgment on the basis of misleading evidence, it is thereby rational for you to revise the judgment that you made on that basis. Such a revision can be made rational by a self-ascriptive judgment only if the self-ascriptive judgment is itself rational. How could self-

17 I will say a bit more about this in section 2.2. These epistemological claims will be put to work in that section and the following one.

18 For reliabilist accounts of knowledge, see Dretske (1981), Goldman (1986). See also Sosa's virtue reliabilism (Sosa, 1980).

ascriptions contribute to the rationality of anything, if they were not themselves rational?

For this reason, I will be not be satisfied with any account of self-knowledge that merely shows that self-ascriptions are reliable, or that they meet some truth-tracking requirement. I will seek an account that also shows that self-ascriptions are rational.

Anyone who thinks it is not obvious that knowledgeable self-ascriptions are rational can think of my project as this: giving a plausible account of self-knowledge of conscious episodes, which allows that self-knowledge is rational. Whether we can give such an account is an important question, even if we would still then have to weigh it up against accounts according to which self-knowledge is not rational.

1.5 Looking ahead

Let me recap. I began by identifying the explanatory target of this thesis: present tense propositional self-knowledge of conscious states and episodes. I then characterised the specialness of self-knowledge in terms of five features that self-knowledge *appears* to have. This allowed me to state the two questions that this thesis addresses: why are self-ascriptions knowledge, and why are they special? This, I suggested, constitutes a philosophical problem, because answering the two questions seems to require positing a mysterious mechanism for hooking up to empirical facts, that somehow yields secure, immediate, authoritative judgments about those facts. I defended my claim to have identified a determinate philosophical problem against McDowell. I distinguished, at a very general level, between approaches to the problem—the epistemic approach, and various deflationist approaches Finally, I made explicit some epistemological assumptions that will guide the argument to come.

In this final section of the introductory chapter I want to give an outline of that argument.

Chapters 2 and 3 will be concerned with what general features the correct account of self-knowledge must have.

In Chapter 2 I will argue that the correct account of self-knowledge must be epistemic rather than deflationary to any extent. Self-ascriptions of conscious states and episodes are often knowledge, since they can be bases of knowledge-transmitting inferences. This implies that self-ascriptions are acts of judgment with epistemic credentials. This implication is inconsistent with the weak deflationist's attempt to explain self-ascriptions' specialness by appeal to expressivist resources

Traditionally, the dominant epistemic account of self-knowledge has been introspectionism: the view that self-ascriptions are based on introspective experiences *of* conscious states and episodes. In Chapter 3 I will show that introspectionism gets deeply wrong the nature of our perspective on our own minds. It renders mysterious the authority of self-knowledge; it is at odds with the point that we often come to self-ascribe conscious states and episodes by looking outwards, to their intentional objects, rather than by attending inwardly; and it allows for the possibility of a certain kind of self-alienation. I will argue that a state's or episode's being conscious is not the same thing as the subject's being conscious *of* it, and that a state or episode can epistemically ground self-ascriptions in virtue of occurring consciously, without any need for a further conscious experience of that state or episode.

Chapter 4 will set out an epistemological context in which the problem of self-knowledge can be solved.

I will outline a non-internalist conception of what it is to have a reason for judgment: a subject has a good reason for a judgment if, firstly, by basing that judgment on that reason she is likely to judge truly, and, secondly, in doing so she would be manifesting a rational sensitivity to the truth-connection between that consideration and the content of the judgment.

Chapters 5 and 6 will present and defend, within that context, an epistemic, non-introspectionist account of self-knowledge.

In Chapter 5 I will present my *moderate epistemic account* for the *first-person* and *content* components of self-knowledge, according to which enjoying a conscious state or episode with a particular content can give a subject a reason to judge that she is entertaining that content. In basing a self-ascription of a content on that content, you will self-ascribe truly, because you cannot judge for the reason given by that content unless you entertain that content. Self-ascribing in this way is rational because any subject who can do so will have a general grasp of the first-/third-person distinction—a practical appreciation, fundamental to our cognition, that the world one is aware of reflects both how the world is, and one's own perspective on it.

In Chapter 6 I will deal with self-knowledge of *type*: how you know what type of conscious state or episode you are enjoying. I will claim that states and episodes of different types have particular features that make a difference to what it is rational for their subjects to do, but that are not introspected by their subjects. I will give a detailed story for two cases: perceptual experience and judgment. Perceptual experience involves a distinctively

perceptual mode of presentation which I call 'directness'. Subjects can be rationally sensitive to the directness of a perceptual experience, without relying on introspection *of* the experience. Judgment, on the other hand, is a type of mental action. I will claim that self-knowledge of action is based on *control*. I will give an account of the way in which judgments count as controlled, and of how that control can ground self-knowledge of judging.

Finally, Chapter 7 will consider the role of self-knowledge in reflective rationality and personhood. I will argue that this role can and should be explained in part by the fact that self-ascriptions are warranted in the way claimed in the previous two chapters. There are views on which the role of self-knowledge is itself part of the warrant for self-ascriptive judgments; I will argue that such views cannot *both* make plausible claims about the role of self-knowledge *and* offer a satisfying explanation of the warrant for self-ascriptions

CHAPTER 2

DEFLATIONISMS

In the last chapter I distinguished several approaches to the problem of self-knowledge. On the one hand, there is the epistemic approach, which aims to explain why self-ascriptions of conscious states and episodes are knowledge, and to explain why self-knowledge is special by appeal to its epistemic credentials. On the other hand, there are various deflationisms: weak, epistemic and strong. Both strong and epistemic deflationism deny that self-ascriptions are really knowledge; the difference is that strong deflationism denies, while epistemic deflationism accepts, that self-ascriptions are special in something like the way spelled out in Chapter 1 above. Weak deflationism accepts that self-ascriptions are knowledge, but holds that the specialness of self-knowledge is in important respects a non-epistemic phenomenon. The heart of the various deflationist accounts is an expressivist treatment of self-ascriptive utterances.[1] In this chapter I will argue that no such brand of deflationism is plausible, and that the correct approach to self-knowledge is epistemic.

The chapter will go as follows. In section 2.1 I will outline the position that denies that self-ascriptions are knowledge, and show that that denial is mistaken. Self-ascriptions can play a role in reasoning, I will argue, that requires them to be knowledge. In section 2.2 I will draw out two useful implications of this argument: that self-ascriptions are acts of judgment, and that those judgments have epistemic credentials partly in virtue of how the subject comes to judge. In section 2.3 I will show that the weak deflationist's attempt to explain the specialness of self-knowledge by appeal to expressivist resources is incompatible with these implications. The conclusion will be that, while the expressivist treatment of self-ascriptive utterances may contain important insights, it does nothing to solve (or dissolve) the problem of self-knowledge.

2.1 Self-knowledge is genuine knowledge

When you think, or say, “It now looks to me as though there is a computer in front of me”, “I am thinking that the foliage is yellow”, or “I wish I had gone out earlier”, you appear to be

1 Crispin Wright (1991, 1998) is an exception. He proposes a view that counts at least as weakly deflationist, but that makes no appeal to expressivism. For space reasons I will not deal with Wright’s view. For the record, I think that, whatever its merits as an account of self-knowledge of the attitude of belief, it is not plausible as an account of self-knowledge of conscious states and episodes, because of its anti-realist commitments about the objects of self-knowledge.

manifesting knowledge about your conscious mind. Both strong and epistemic deflationism hold that this appearance is illusory. They are error theories about claims of self-knowledge.

Philosophers sympathetic to these brands of deflationism have been motivated in part by the thought that the apparent distinctive features (a)-(e) of self-knowledge (its specialness; see section 1.2) are inexplicable in a context in which we take it that self-knowledge is just another sort of knowledge, like perceptual knowledge or knowledge by testimony.[2] By taking a different view of the nature of self-ascriptions, we can see these distinctive features as unmysterious or even illusory; and we can see why they appear mysterious and real when self-ascriptions are mistakenly taken to constitute self-*knowledge*.

The strong or epistemic deflationist holds that self-ascriptions are not knowledge because they are not, and do not express, *judgments* about the subject's mental states or episodes at all. Instead, they are expressions of those mental states or episodes. So they do not qualify for epistemic assessment *qua* self-ascriptive judgments. This is the view I will refer to as *expressivism*.[3]

When you express a mental state or episode, you perform an action that is directly caused by—or even partly constitutive of—that state or episode, and that reveals that state or episode in an especially direct way. You can express a state or episode without judging or asserting that you are enjoying that state or episode. You express your pain by groaning. You express your wish to have some curry by reaching for the bowl that is full of curry; and you express your subsequent judgment or belief that the curry is delicious by making a certain facial expression. Alternatively, you can express this wish and this judgment by making certain linguistic utterances—by saying “Curry would be wonderful”, and then “This curry is delicious”. Neither of these latter utterances involve an assertion about a mental state or episode of yours: they are different from the respective assertions that you want curry and that you think this curry is delicious. Nevertheless, provided you are being sincere, an interlocutor can learn about your wish or your judgment from your expression of it as well as from an assertion about it.

Expressivism, as I will understand it, is a view both about speech-acts (utterances) and about acts of thinking (what we ordinarily take to be judgments). For ease of exposition, I will

2 Wittgenstein seems to have been tempted to offer such an error theory (Wittgenstein, 1953, pp. 189ff.; 1980, esp. §§ 450, 832). See also Ginet (1968) and Wright (1998).

3 The neo-expressivist view I will discuss later does not count as expressivist, in my terminology. Throughout, when I say ‘expressivism’ I am referring to the view presently being outlined, which is strongly or epistemically deflationist, and not to neo-expressivism, which, as I will understand it, is weakly deflationist. However, neo-expressivism does draw on the expressivist's account of self-ascriptive utterances.

assume that occurrent thoughts, including those that we ordinarily take to be judgments, consist of utterances to yourself (this is not to suggest that the expressivist is committed to this).

The central claim of expressivism is:

(**E**) For any speaker S, mental state- or episode-type M, and content *p*, in ordinary circumstances an utterance by S of "I M that *p*" is *not* an assertion by S that she Ms that *p*, but *merely* an expression by S of her M-ing that *p*.

For present purposes, M is restricted to those states and episodes within the range I demarcated in Chapter 1—conscious states and episodes with propositional contents. By 'ordinary circumstances' I mean to exclude those cases in which a subject comes to make a self-ascription through psychoanalysis, through observation of her own behaviour, or by some other deviant route.

The claim, then, is that when you say "I want to have curry", you are not asserting something about your mental state, but rather doing something akin to reaching for or pointing to a bowl of curry. Your interlocutor can thereby learn that you want curry, by perceiving your mental state being revealed, rather than by accepting a proposition you put forth. When you say "I think this curry is delicious", you are not reporting a mental episode or state, but expressing your judgment or belief that the curry is delicious. You are doing something akin to simply asserting that the curry is delicious, or to saying "Mmm".

There is an analogue of (**E**) that deals with self-ascriptions in thought, rather than utterances. It says roughly (remembering that we are treating occurrent thoughts as utterances to yourself) that what I have been calling "self-ascriptive judgments" are in fact *not* assertions to yourself, but merely episodes in which you express your mental state to yourself. When you think to yourself, "I think this curry is delicious", you are performing an inner "Mmm".

It is important that (**E**) is a negative thesis as well as a positive one. The expressivist makes the positive claim that self-ascriptive utterances and thoughts express the mental states they ostensibly ascribe. But one need not be an expressivist (or even a neo-expressivist) to agree with this. A non-deflationist can allow that a self-ascriptive utterance (say), as well as being an assertion about the speaker's mental state, is *simultaneously* an expression of that state: your statement that you think the curry is delicious reports on your judgment, but it also

expresses your judgment (or belief).[4] The difference between the two views is this: the expressivist denies, and the non-deflationist holds, that self-ascriptions have the role of asserting that the speaker or thinker is in the mental state in question, and thus of expressing the judgment or belief that she is in that mental state.[5] The expressivist denies this because if self-ascriptions are assertions about mental states, and thus express judgments or beliefs about mental states, the question arises of the epistemic credentials of those judgments or beliefs. The denial that this question can arise is central to the strongly and epistemically deflationist solutions to the problem of self-knowledge.

What sort of solution do they offer? The first part of the problem of self-knowledge, recall, is the knowledge question: why are self-ascriptions knowledge? The expressivist holds that this question is misguided: self-ascriptions are not really judgments about the subject's mental states and episodes, and so do not constitute or manifest knowledge about them, although we are deceived by their surface form into thinking that they do. The second part of the problem is the specialness question: why do self-ascriptions appear to have features (a)-(e)? The expressivist holds that the features, understood correctly, are not so special after all.

Security. Our self-ascriptions do not constitute extraordinarily safe judgments about our mental states and episodes, because they are not such judgments at all. However, the expression of a mental state is closely, perhaps constitutively, tied to being in that mental state, so expressions that take the form of self-ascriptive utterances will appear to be extraordinarily safe—just as the sincere utterance “This curry is delicious” is an extraordinarily safe indicator of the subject's judgment about the taste of the curry, because it simply expresses the judgment (or belief), rather than relying on anyone’s coming to know about the judgment.

Salience. Conscious states and episodes can almost always be expressed in behaviour; only in very abnormal circumstances is that not so. Since such expressions can take the form of self-ascriptive utterances, subjects will almost always appear, or be disposed to appear, to be right about their conscious states and episodes. But there is no genuine judgment, right or wrong, here.

4 The non-deflationist who took this line would be rejecting the thesis that Jacobsen (1996) calls 'Expressive Exclusivity'—the thesis that an utterance cannot express more than one mental state or episode. The neo-expressivist view that I will consider later can side with the non-deflationist on this point.

5 I assume here that if you sincerely assert that *p*, you express the judgment or belief that *p*.

First-person privilege. Only you can express *your* mental states and episodes. But this has nothing to do with a privileged way of coming to know about them.

Authority. Since self-ascriptions are not, and do not express, judgments about the subject's mental state, it would of course be inappropriate to gainsay or demand justification for them as such—just as, if you say, "This curry is delicious", an interlocutor might challenge you about the taste of the curry, but not about whether that is really your judgment.

Immediacy. Since there is no self-ascriptive judgment, but merely an expressing of your mental state, there is no question of acquiring or attending to grounds, or of finding anything out.

The sort of account I have been outlining can be developed in more than one way. Until recent years, it has been seen as a view about the *semantics* of sentences of the form "I M that *p*". Traditional expressivism[6] about self-ascriptions says that the utterance of such a sentence, in ordinary circumstances, does *not* put forth any proposition about the speaker's mental state or episode—the speaker does not say anything about her mental state or episode. Thus she does not say anything true or false about it, nor does she say anything that she knows or fails to know about it. This leads to a strongly deflationist view of self-ascriptions, since, if self-ascribers are not even saying anything about their mental states and episodes, my characterisations of distinctive features (a)-(e) are *radically* inaccurate.

But there is a more recent brand of expressivism which focuses on the *pragmatics* rather than the semantics of self-ascriptive utterances (this view can be found in Jacobsen, 1996; and Bar-On and Long, 2001). Expressivists of this stripe distinguish the *act* of uttering a sentence of the form "I M that *p*" from the *product* of that act—the token of the sentence thereby uttered.[7] In itself, (**E**) is a thesis about pragmatics—about self-ascriptive acts. One can endorse (**E**), claiming that self-ascriptive acts are not assertions about the speaker's mental

6 See Ayer (1936, Chapter 6) for a statement of 'traditional expressivism' about ethics. It is controversial whether this is the brand of expressivism to which Wittgenstein was attracted in his treatment of self-ascriptions (see Jacobsen, 1996). However, it has been attributed to him often enough that the view demands consideration.

7 This distinction is formulated and developed most clearly in Bar-On (2004). However, in that work Bar-On avoids committing to the brand of expressivism I am discussing here.

states and episodes, while allowing that such acts involve the production of sentences whose meaning concerns the speaker's mental states and episodes. Such acts put forth (in some sense) self-ascriptive propositions, but they do not put those propositions forth with assertoric force. This is a familiar phenomenon. You may say "You will close the window" as a way of instructing someone to close the window. In such a case you do not use your utterance as an assertion that your interlocutor will close the window, nor need you come to make it by making a judgment about her future actions; and you may not be prepared to offer a justification for the claim that she will in fact close the window. You may not be confident at all that your instruction will be followed. Nevertheless, *what you say* is true if and only if your interlocutor goes on to close the window. Similarly, according to this brand of expressivism, an utterance of the sentence "I M that *p*" puts forth a proposition that is true iff the speaker is M-ing that *p*, but such an utterance is not ordinarily an assertion of that proposition. It does not express the speaker's judgment or belief that she is M-ing that *p*. The speaker's *act* is merely one of expressing her M-ing, not one of reporting it.

This view thus allows that an occurrence of a sentence "I M that *p*" has truth-conditions concerning the speaker's mental state, while taking advantage of the expressivist treatment of self-ascriptive acts in order to explain the apparent distinctive features (a)-(e) of those acts. It can be an epistemically deflationist, rather than strongly deflationist, view, because the distinctive features (a)-(e) can be vindicated, albeit under slightly revised characterisations, rather than being shown to be illusory. For example, self-ascriptions are secure, in that the propositions they put forth are securely true—but they do not constitute secure *knowledge*.

I want to offer a general argument against strong and epistemic deflationism. It is tempting to meet these views with an incredulous stare—*surely* we can make knowledgeable judgments about our own conscious states and episodes. Those who are incredulous will probably not find the premises of my argument any *more* compelling than they already find its conclusion—that we can make knowledgeable judgments about our own conscious states and episodes. But, dialectically, I must offer some independently plausible grounds for that conclusion. I will appeal to facts about the role of self-ascriptions in our thought, our communication and our action—facts that I think are undeniable. I hope that the argument brings out some of what is behind the *prima facie* implausibility of strong and epistemic deflationism.

Let us compare two scenarios in which a self-ascription can be issued.

First, suppose you express yourself by uttering the sentence, "I think this curry is delicious". An interlocutor, hearing what you have said, can reason to the conclusion, "At least one

person at the table thinks the curry is delicious". Your interlocutor can thereby acquire knowledge. For, hearing your utterance gives her knowledge that *you* think the curry is delicious, and this knowledge can be transmitted across the competent inference to the conclusion that at least one person at the table thinks the curry is delicious. Call this method of belief-acquisition **IR**.[8]

Second, suppose that, instead of making a public utterance, you think or say to yourself, "I think this curry is delicious". If you do that, you can engage in some simple reasoning and come to the conclusion, "At least one person at the table thinks the curry is delicious". This procedure, which we can call **SR**, appears to resemble **IR** in that it involves reasoning from a fact about a person's thoughts—namely, your own.

People manifestly can and do engage in reasoning from facts about their own thoughts. A reflective subject can, for example, revise an attitude on the basis of a recognition that she has made a judgment on an inadequate basis; such reflective reasoning proceeds from the fact that she has indeed made that judgment. Equally, such a subject can reflect on whether her present experience gives her reason to believe that foliage in front of her is yellow; such reasoning proceeds from the fact that she is enjoying a certain experience.[9] To deny that people can reason from facts about their own conscious states and episodes would be to take an obviously false view of people's cognitive and rational abilities.

It seems that **SR** leads to knowledge. The conclusion that at least one person thinks the curry is delicious, arrived at by **SR**, is something you can know. You can rationally and successfully act on it. For example, if you reach that conclusion, and you face some crucial choice that depends on whether anyone thinks the curry is delicious, you are in a position rationally to make the right choice, because you know that at least one person thinks the

8 There is a question over what account we should give of your interlocutor's knowledge, in such a case, that you think the curry is delicious. On its face, it appears to be a case of testimony: she comes to know that you think the curry is delicious by accepting what you say. At any rate, it seems that we *can* learn about others' conscious states and episodes by their testimony. But this account is not available to the expressivist. Knowledge by testimony is acquired by accepting the content of another's assertion. The expressivist denies that, when you say "I think this curry is delicious", you are making an assertion. If that's right, then there *is* no testimony for your interlocutor to rely on. The expressivist must offer an alternative account of your interlocutor's knowledge. He must claim that your interlocutor comes to know about your judgment by taking your utterance to be an *expression* of that judgment or belief. Thus, the expressivist is committed to the claim that the appearance in such a case—the appearance that it is a case of testimony—is misleading, for such cases are *never* cases of testimony.

9 It is arguable that such reflective reasoning constitutively requires subjects to be capable of making knowledgeable self-ascriptive judgments (see Chapter 7; and see Burge, 1996). If that is right, then there is a simple argument here against strong and epistemic deflationism. I am offering a slightly different argument, however.

curry is delicious.[10]

If you were *not* in a position to know by **SR** that at least one person thinks the curry is delicious, you would be, in that respect, in a worse position than someone who hears you say "I think the curry is delicious". Indeed, if **SR** did not lead to knowledge, you could be, bizarrely, at a disadvantage relative to others with respect to whether anyone thinks the curry is delicious. For, suppose you think "I think the curry is delicious", and do not express your judgment that the curry is delicious in an utterance, but reveal it with an unconscious gesture or facial expression that you yourself don't notice. An observer, perceiving your expression, may thereby come to know that you think the curry is delicious, and, by inference, that at least one person thinks the curry is delicious. If **SR** is not available to you as a way of coming to know that at least one person thinks the curry is delicious, then you seem to have no way of coming to know it—or, at least, no way as simple as that available to the observer who sees your facial expression. But it would be insane to suppose that the person in the worst position to tell whether anyone thinks the curry is delicious is the very person who does think it is delicious. Clearly, then, when you think to yourself "I think the curry is delicious", you can come to know by **SR** that at least one person in the room thinks the curry is delicious.

Again, I emphasise that we do in fact engage in reasoning of this sort. We reflect on our conscious states and episodes and on the implications of our having them. And we come to know things by such reflection.

By expressivism's lights, you could not come to know by **SR** that at least one person thinks the curry is delicious. For the proposition that at least one person thinks the curry is delicious is the conclusion of an inference, among whose premises is the proposition that *you* think the curry is delicious. To carry out this inference, you must judge that you think the curry is delicious. In order for the inference to yield knowledge, your so judging must constitute or manifest knowledge—a judgment in the conclusion of an inference cannot be epistemically *more* satisfactory than the preceding judgment in any of the premises.[11] *Ex hypothesi*, **SR** begins from your thinking or saying to yourself, "I think this curry is delicious".

10 It might be said that successful and rational action need not be based on *knowledge*—it can be based on something that falls short of knowledge. But this sort of reply is no use to the epistemic deflationist. For, whatever the epistemic status of the conclusion of the inference, the epistemic status of the premise—that you think the curry is delicious—must be at least as good (see below). And the epistemic deflationist does not claim that self-ascriptions *fall short* of being knowledge; he claims that they are not up for epistemic evaluation at all. Self-ascriptions, for the epistemic deflationist, are not, and do not express, judgments, rational or irrational, warranted or unwarranted.

11 Provided you have no other reason to believe the conclusion.

Expressivism denies that this is a judgment that constitutes or manifests knowledge. It is therefore committed to denying that you can come to know by **SR**, that at least one person thinks the curry is delicious. But as we saw, you *can* come to know, by simple reasoning, that at least one person thinks the curry is delicious, when you think to yourself, "I think this curry is delicious". If you couldn't, there would be the possibility of everyone *except* you knowing that at least one person at the table thinks the curry is delicious.

In sum: we can engage in reasoning from facts about our own conscious states and episodes; such reasoning can lead to knowledge; it could not lead to knowledge if it did not involve knowledgeable judgments about those conscious states and episodes; therefore, self-ascriptions of conscious states and episodes are, at least sometimes, knowledge. The central commitment of strong and epistemic deflationism is false.[12]

The expressivist might reply that you can indeed know that at least one person thinks the curry is delicious, but this knowledge is not inferred from any special knowledge that you think the curry is delicious. He might claim that your knowledge that at least one person thinks the curry is delicious is based, not on the *content* of your self-ascription (since it could only be based on that if the self-ascription manifested knowledge with that content), but on the self-ascriptive *act*. According to this reply, you can come to know that at least one person thinks the curry is delicious by knowing about your act of saying or thinking "I think this curry is delicious", but this latter act should not be construed as a judgment from whose content you infer your conclusion.

But this reply is hopeless. On the one hand, you do not come to know that at least one person thinks the curry is delicious by finding yourself *saying* "I think this curry is delicious". Knowledge of this sort need not await public expressions of your mental state. On the other hand, the expressivist cannot claim that you come to know that at least one person thinks the curry is delicious by finding yourself *thinking* "I think this curry is delicious". For, if you are to come to know in this way, you must first have some epistemic access to your performance of this mental act of thinking "I think this curry is delicious". And so the question arises: what is this epistemic access? The expressivist cannot at this point appeal to some distinctive way that we have of coming to know our own mental states and episodes, for, if there is such a way, it can surely lead to knowledgeable self-ascriptive judgments; to allow for such a way

12 Note also that, by allowing that self-ascriptive utterances can be knowledgeable assertions, we can vindicate the appearance that we learn about others' mental states and episodes by testimony. See note 8 above.

would thus be to abandon both strong and epistemic deflationism.[13] If pursuing this reply, the expressivist seems bound to claim that you know about this mental act in virtue of knowing about some *further* act that expresses *it*. But, as before, this further act cannot be a public expression, so it must be a mental act, which you must know about in virtue of a further mental act, and so on. Clearly, this route leads nowhere helpful for the expressivist.

Note that I have not begged the question against expressivism. It is not a premise of the above argument that self-ascriptions constitute or manifest knowledge about your mental states and episodes. Rather, I argued that you can come to know things by inference from facts about your mental states and episodes, and that you must therefore be able to know about those states and episodes. The problem for the expressivist is that, in accounting for this fact, he must either acknowledge that there is a distinctive sort of knowledge in question here, or give a hopelessly implausible account of how we acquire that knowledge.

This is an argument against any view that denies that self-ascriptions constitute genuine self-knowledge. Thus, it rules out any strongly deflationist or epistemically deflationist view. The correct account must be either an epistemic or a weakly deflationist one. The rest of this chapter will be devoted to arguing that we should look for an epistemic solution.

2.2 Self-ascriptions as acts of judgment with epistemic credentials

As a preliminary to the argument that we should look for an epistemic solution to the problem of self-knowledge, I want briefly to say something about the implications of the conclusion that self-ascriptions are (constitute or manifest) knowledge. I will go on to argue that we cannot simultaneously respect these implications and give a non-epistemic, expressivist-inspired explanation of the distinctive features of self-knowledge.

2.2.1 Self-ascriptions are acts of judgment

Knowledge of a proposition, I assume, entails belief, or at least some attitude of acceptance of that proposition. When a self-ascription constitutes or manifests such knowledge, it is a manifestation of that propositional attitude. A self-ascription is also an act, in speech or

13 I am deliberately leaving open here what kind of epistemic access this would have to be—for example, whether it would have to be access to the *fact* that you are performing the mental act, or some sort of acquaintance with the act itself. Even if it were in the first place a sort of knowledge by acquaintance, it would thereby give the subject a way of coming to know facts about her mental states and episodes. The expressivist denies that there is such a way.

thought, of affirming the proposition that is known. An act of affirming a proposition, that manifests acceptance of that proposition, is an act of judgment. Thus, when you think "I think this curry is delicious", and that self-ascription constitutes or manifests knowledge, you are performing an act of self-ascriptive judgment; when you say "I think this curry is delicious", you are articulating your self-ascriptive judgment.[14]

It might be objected that not all self-ascriptions are judgments. I have shown that we *can* judge self-ascriptive propositions, and that such judgments can constitute or manifest knowledge. But it doesn't follow that every act of thinking a self-ascriptive content, or uttering a true self-ascriptive sentence, is an act of judgment—an act of affirming such a proposition. For all I have said, many such acts may not be judgments.

But I am not claiming that every utterance or thought that involves the tokening of a self-ascriptive sentence or content is an act of judgment. I am claiming that some central cases—in particular, those that constitute or manifest knowledge—are. We can and do make self-ascriptive judgments. Typically, you are in a position to make a self-ascriptive judgment about any of your conscious states and episodes, and such a judgment will constitute or manifest knowledge and will have the distinctive features of self-knowledge. That is the phenomenon I am trying to explain.

Nor does my claim here entail that all propositional knowledge involves judgment. That is, I do not deny that you can know a proposition (or fact) without making any explicit judgment.[15] Nor do I deny that, when it comes to propositions about your present conscious states and episodes, you typically do not judge them, but merely believe them in a dispositional sense. My claim is simply that, when you do make what appear to be self-ascriptive judgments, those judgments are exactly what they appear.[16]

---

14 Dorit Bar-On, the proponent of weak deflationism on whom I will focus, distinguishes between two senses of 'belief'. A belief can be a mere disposition to accept a proposition on considering it. Or, it can involve forming a judgment "where one has (and could offer) specific evidence or reasons for that judgment." (Bar-On, 2004, p. 363.) When I say that self-ascriptions are judgments, I do not imply that the self-ascribing subject has and could offer evidence or reasons for the judgment. I do imply, however, that there is a conscious episode of affirming a proposition. Indeed, Bar-On accepts that self-ascriptions may fall between the two notions she distinguishes (*ibid*., p. 365).
Note also that, in saying that an act of thinking "I M that *p*" is a judgment, I do not imply that it is based on several distinct judgments, "I M, rather than M*, that *p*", "I M that *p*, rather than that *q*", and so on. To judge "I M that *p*" need not be to *identify* one's conscious state or episode as M-ing rather than M*-ing, or to identify its content as *p* rather than *q*.

15 There is also an important sense in which you can know a fact that you would be unable to grasp conceptually. I take it that a mature thinker's self-knowledge is not like that, however.

16 As noted in section 1.4, I focus on judgments rather than the beliefs they manifest. So I don't say anything about why self-ascriptive beliefs that never get manifested in judgment are knowledge.

2.2.2 Epistemic credentials and ways of coming to judge

What consequences does the thesis that self-ascriptions are acts of judgment have?

An act of judging a proposition to be true is an exercise of conceptual capacities—the capacities that constitute possession of the concepts that compose the content of the judgment. The subject grasps the truth-conditions of the content in virtue of possessing those concepts. To judge a proposition is to accept that its truth-conditions are met; the act of judgment is an act of affirming that those conditions obtain.17

I am assuming (see section 1.4) that warrant is necessary for knowledge. A judgment constitutes or manifests knowledge only when it is warrant*ed*. A judgment is warranted when it has an appropriate connection to a warrant—to whatever feature of the subject's situation or character provides the warrant for it. That connection constitutes the judgment's epistemic credentials (*ibid*.).[18] This point is compatible with any conception of warrant that acknowledges the distinction between there *being* a warrant for some judgment and a judgment's being warrant*ed* by that warrant (an instance of this is the distinction between propositional and doxastic justification). This distinction applies as much to self-ascriptive judgments as to any other. If you are thinking that it is windy, but judge that you are thinking that it is windy because you accept the testimony of a phony clairvoyant, your self-ascriptive judgment will not be warranted.

What determines the epistemic credentials of a judgment? On one view, a warranted judgment is warranted in virtue of what the subject *could* do—that is, because she could give justifying reasons for it. But this view is too liberal, as the following passage from Sosa establishes:

However, it seems plausible that there is a very close connection between the explanation of why those beliefs are knowledge and the explanation of why the judgments that would manifest them constitute knowledge.

17 Some ethical expressivists might wish to deny that ethical propositions are tied to truth-conditions in this way. Such a denial is motivated by doubts about the existence of a realm of moral facts: ethical expressivists are suspicious of the idea that ethical propositions *have* truth-conditions. There is no parallel motivation in the case of self-knowledge—we do not wish to deny that there are facts about subjects' conscious states and episodes. I assume that a truth-conditional account of self-ascriptive propositions is correct.

18 Epistemic credentials can also attach to beliefs. But when a belief is acquired by the making of a conscious judgment, the warrant for the belief will depend on whether the judgment has good credentials (as long as the subject does not acquire some new reason to maintain the belief, besides that for which she made the original judgment).

> It would not do, however, to suppose that someone already knows something just because if they started thinking about how to defend their belief, they would *then* come up with a fine proof. Someone who guesses the answer to a complex addition problem does not already know the answer just because, given a little time, he could do the sum in his head. If he had not done the sum, if he had just been guessing, then he *acquires* his knowledge through reflection, and does not know beforehand. (Sosa, 2007, p. 121.)

Typically, the epistemic credentials of a judgment are determined in part by the way in which the subject *comes to judge*. The way in which the subject comes to judge typically determines the causal and constitutive relations into which the act of judging enters—it thus determines whether the act is appropriately connected to what warrants it.

The notion of a way of coming to judge can be illustrated by examples. You can come to judge that the foliage on a tree is yellow by accepting the content of a perceptual experience that presents that foliage as yellow. Your judgment, in that case, will typically have good epistemic credentials. It will be appropriately connected to a warrant-providing experience. If you came to judge that the foliage was yellow by accepting the content of a dream experience, the resulting judgment would lack epistemic credentials. It would not be warranted—even if you had had an experience that *would* have warranted that judgment, had you come to make the judgment in the right way. Similarly, suppose you have excellent meteorological evidence, that warrants the judgment that it will be windy today. If you come to judge that it will be windy today by being aware of that evidence and competently concluding that it will be windy today, your judgment will have good epistemic credentials. If you come to make that judgment by accepting the testimony of a manifestly unreliable interlocutor, it will not have good epistemic credentials.

In many cases, the notion of a way of coming to judge can be understood in terms of *transitions*: the transition from a perceptual experience with a particular content to a judgment of that content, or from awareness of certain evidence to the judgment that the evidence provides a warrant for. A judgment arrived at by appropriate transitions will have good epistemic credentials, and be warranted. [19]

Thus, the epistemic credentials of a judgment are partly a matter of coming to judge in an appropriate way.

This notion of a way of coming to judge should not be understood as implying that a judgment must be the upshot of some effort of investigation, deliberation or evidence-gathering. The transition from a perceptual experience to a judgment is relatively automatic

19 See Peacocke (2004) for this way of thinking of matters.

and effortless; but still counts as a way of coming to judge.

Furthermore, my claim here is not incompatible with Bar-On's thesis that self-ascriptions of conscious states and episodes exemplify, for the first-person, type and content components, the phenomenon of immunity to error through misidentification (Bar-On, 2004, Chapter 6).[20] The thesis is that, when you self-ascribe a conscious state or episode in the ordinary way, your self-ascription will not: be partly right, but go wrong in exactly one of the following ways: by misidentifying *who* is enjoying that state or episode, by misidentifying the *type* of state or episode you are enjoying, or by misidentifying the *content* of your state or episode. This immunity to error through misidentification shows that, when you judge "I M that *p*", your judgment is not based on distinct grounds for telling *who* is M-ing that *p*, for telling that it is M-ing rather than M*-ing that is being done, and for telling that the content is *p* rather than *q*. It doesn't follow that you don't come to judge "I M that *p*" in some particular way. Perhaps there is a way of coming to judge "I M that *p*" that cannot be broken down into such components—just as you can come to judge that your legs are crossed by accepting the content of your proprioceptive experience, even though that way of coming to judge does not involve a distinct basis for telling *whose* legs are crossed (as it would if you saw a cross-legged person in a mirror and identified that person as yourself). Indeed, it seems that immunity to error through misidentification is *explained* in part by the fact that such self-ascriptions are arrived at in a particular way—a way that does not involve distinct bases for identifying a person, a state- or episode-type, or a content. Self-ascriptions arrived at in some other way—by observing your behaviour on a screen, for example—would not exemplify immunity to error because they are not arrived at in the ordinary way.

In sum, the notion of a way of coming to judge is required to capture adequately a number of important features of warranted judgments, including the distinction between there being a warrant for a judgment, and that judgment's being warranted. A judgment constitutes or manifests knowledge only if it is warranted; it is warranted when it has the right epistemic credentials; and in almost every case it has epistemic credentials at least partly in virtue of being arrived at in an appropriate way. Again, I do not claim that a subject knows that *p* only if she makes a judgment that *p* with good epistemic credentials. Rather, I claim that almost every judgment that *does* constitute or manifest knowledge has good epistemic credentials, partly in virtue of the way the subject comes to judge.

I say "almost" because there may be exceptions. Analytic judgments are good candidates. An analytic proposition is such that, if you understand it, you appreciate that it is true. What

20 Classic treatments of this phenomenon include Shoemaker (1968) and Evans (1982).

provides the warrant for a judgment with an analytic content is your understanding of that content, rather than some feature of your situation to which your judging is connected. But my claim still holds for self-ascriptive judgments. Self-ascriptive judgments, unlike analytic judgments, do not have the property of being warranted no matter what; their warrants depend on the circumstances in which they are made.

An apparently more serious (for me) possible exception is this. It has been claimed that there is a range of self-ascriptive judgments that are self-verifying: making a judgment that falls within this range allegedly guarantees that the content of the judgment is true. For example, any judgment of the form, "I am entertaining the proposition that *p*", is allegedly self-verifying, because making the judgment involves entertaining the proposition that *p* (Burge, 1996).[21] It might be argued, further, that the warrant for such judgments is similar to the warrant for analytic judgments—that it depends only on the subject's understanding the content judged. If that is right, then the way that the subject comes to make the judgment is irrelevant to its warrant; it would be warranted however the subject came to make it (other things equal).

But these supposedly self-verifying judgments do not threaten my claim. Even if there are self-verifying self-ascriptions, we also know about those conscious states and episodes which we do not bring into being by taking ourselves to be in them. You are in a position to know about your own perceptual experiences, for example, even though their occurrence is independent of whether you self-ascribe them. Often you make a judgment without doing so *by* self-ascribing it, and yet you are in a position to know about it—you know about it because you make it, rather than the opposite (Chapter 6 will offer an account of this). In Chapter 5 I will discuss self-ascriptions of the form "I am entertaining the proposition that *p*", which correctly self-ascribe a mental state that the subject is in, independently of the self-ascription of it. Much of our self-knowledge is of this non-self-verifying sort; and this is the special self-knowledge I am concerned with.

It might be objected that the foregoing picture of the warrant for self-ascriptive judgments is incompatible with the immediacy of self-knowledge. Immediacy, recall, is the feature that self-ascriptions are not arrived at by acquiring or attending to grounds, and do not express something that the subject finds out about herself. It might be supposed that immediacy entails that the self-ascribing subject does not need to come to make a self-ascriptive judgment by some appropriate procedure, since she is in a position to know her mental state

21 Burge talks about "I am thinking that *p*", rather than "I am entertaining the proposition that *p*". His use of "thinking that" seems to amount to the same as merely entertaining a proposition.

just by being in it.[22]

But this objection misunderstands immediacy. Immediacy reflects the fact that the subject of a conscious state or episode does not have to do anything, beyond enjoying that state or episode, in order to be in a position to self-ascribe it. It doesn't follow that there is no way in which that subject comes to make a self-ascriptive judgment, if she does so. This way of coming to judge does not involve acquiring or attending to grounds, or finding out. But it is a way of coming to judge. (I will argue in Chapters 5 and 6 that the relevant way is a certain sort of transition from a conscious state or episode to the self-ascription of that state or episode.) As we saw above, a subject might, in certain extraordinary circumstances, come to make a self-ascriptive judgment in some other way. In that case, the judgment would not be warranted in the way that our ordinary self-ascriptive judgments usually are, precisely because it is not arrived at in the usual way.

In this section I have claimed that self-ascriptions in thought and speech are acts of judgment (or articulate such acts), and that they manifest knowledge partly in virtue of the way we come to make them. In the next section I will put these claims to use in arguing against the weakly deflationist approach to self-knowledge, as represented by neo-expressivism.

2.3 Against weak deflationism

Section 2.1 ruled out strong deflationism and epistemic deflationism. It follows, given that the Rylean view is not an option (see Chapter 1, section 1.3.2) that the correct approach must be either weak deflationism or to look for an epistemic solution. In this section I will argue against weak deflationism.

A weakly deflationist view is one according to which self-ascriptions constitute knowledge, and have something like features (a)-(e), but at least some of those features are not to be explained by appeal to the epistemic credentials in virtue of which self-ascriptions constitute knowledge. What makes self-knowledge special in certain respects, on this type of view, is not an epistemic matter.[23] Dorit Bar-On's neo-expressivist account of self-knowledge (Bar-On, 2004) represents an attempt to use expressivist resources to explain the specialness of

22 This is a cousin of the immediacy-based objection to my epistemological assumptions, which I considered in section 1.5. There, the objection was that immediacy entails lack of epistemic credentials. Here, it is that immediacy entails that the epistemic credentials are not a matter of a way of coming to judge.

23 Accordingly, the precise characterisations of (a)-(e) would have to be revised somewhat from that I gave in Chapter 1.

self-ascriptions, without denying that self-ascriptions constitute knowledge.[24] That is, neo-expressivism attempts to solve what I called 'the specialness question' (Chapter 1, section 1.3.1), while leaving open the possibility that 'the knowledge question' can be solved, though without entailing any particular solution to the latter. I will argue that, while some of neo-expressivism's claims may be true, we cannot *both* use those claims to explain the specialness of self-knowledge, *and* keep the account compatible with the fact that self-ascriptions are knowledge. Since self-ascriptions *are* knowledge, we must look elsewhere for an explanation of their specialness.

Like the epistemically deflationist account considered above (section 2.1), neo-expressivism appeals to the distinction between the semantics and the pragmatics of an utterance (again, I will characterise the view with reference to utterances rather than thoughts, and assume that it can be extended to thoughts). Consider again the utterance, in ordinary circumstances, of "I think this curry is delicious". The act of utterance, according to Bar-On, should be viewed primarily as an act of expressing the judgment or belief that the curry is delicious, as distinct from an act of asserting that you are making that judgment. Nevertheless, in performing the act, you put forth, by virtue of the sentence you utter, the assertible proposition that you judge that the curry is delicious. What you say is true if and only if you do make that judgment, even though a primary use of your utterance is simply to express the judgment.

For Bar-On, the utterance of "I think this curry is delicious" lies on a continuum of expressive acts (*ibid*., Chapter 7, 8). At one end of the contiuum lie 'natural expressions': non-linguistic behaviours such as facial expressions, groans and bodily movements that by their nature express certain mental states and episodes. In the middle of the continuum lie 'avowals proper'. These are spontaneous expressive utterances that involve the production of self-ascriptive sentences, and thereby the putting forth of self-ascriptive propositions, but that are not brought about by any intention to make a self-ascriptive assertion or to inform interlocutors about your mental states or episodes. Avowals proper are acts performed "*in lieu* of making a gesture, or a face, or grunting, etc." (*ibid*., p. 262), not utterances brought about by an acceptance on the subject's part that some condition obtains with her. Finally, there are 'non-evidential reportive avowals'. These are self-ascriptive utterances that, like avowals proper, involve the putting forth of propositions concerning the speaker's mental states and episodes, but, unlike avowals proper, are performed with the communicative

24 In earlier work Bar-On does deny that self-ascriptions are knowledge (e.g. Bar-On and Long, 2001), and thus takes an epistemically deflationist view. In her (2004) she remains neutral between epistemic and weak deflationism. For present purposes, only the weakly deflationist reading of her account is relevant.

intention of informing interlocutors of the truth of those propositions. Reportive avowals include those self-ascriptive utterances produced in response to questions like "What do you think?", "How do you feel?", and so on. Even such reportive avowals are acts of expressing, of speaking from, mental states, according to Bar-On (*ibid*., pp. 301ff.).

Bar-On claims that a self-ascriptive utterance that, in virtue of its aetiology, expresses the episode or state it self-ascribes, can *also* constitute or manifest knowledge *about* that episode or state. That is, your utterance of "I think this curry is delicious" can both express your judgment or belief that the curry is delicious and articulate your knowledge that you judge that the curry is delicious.[25] This is the claim that distinguishes weak deflationism from epistemic deflationism.[26]

Neo-expressivism, on this reading, agrees with expressivism that the distinctive features of self-knowledge are explained in terms of the expressive role of self-ascriptive utterances and thoughts. On the other hand, neo-expressivism does not deny that such utterances and thoughts manifest knowledge about the subject's mental states and episodes.

Neo-expressivism's explanation of the apparent distinctive features of self-knowledge is vindicatory; it claims that self-ascriptions really do have those features (although it may require a slightly revisionary characterisation of some of them).

Security. The self-ascriptive propositions put forth by sincere expressive acts that involve the production of self-ascriptive sentences are extraordinarily safe. This is explained by the aetiology of those acts. In so far as those acts have the character, not of assertions, arrived at by some epistemically satisfactory route, of a proposition accepted by the subject as true, but rather of acts of expressing mental states and episodes, they will not go wrong in the ways that epistemically satisfactory ways of coming to judge ordinarily can.

In particular, self-ascriptions will be immune to errors of misidentification (see above, section 2.2.2; and Bar-On, 2004, Chapter 4). In expressing your mental state or episode, you do not rely on some way of telling who the subject of the state or episode is. Similarly, you do not rely on some way of telling what type of state or episode it is; nor do you rely on a way of telling what its content is. Thus, in so far as an act is expressive of a state or episode,

25 This claim, which Bar-On calls the 'dual expression thesis' (*ibid*., p. 307), entails the rejection of 'Expressive Exclusivity'—see note 4, above. If Expressive Exclusivity is endorsed, then, given the considerations of section 2.2, neo-expressivism is immediately ruled out. There does not seem to me to be any obvious barrier to rejecting Expressive Exclusivity.

26 This claim is also available to the non-deflationist, as I emphasised in section 2.1. But the non-deflationist does not try to use it in explaining the distinctive features of self-knowledge.

it will not go wrong by misidentifying the subject, type or content of that state or episode.

This explanation of security applies across the continuum of expressive acts, according to Bar-On. Natural expressions are secure indicators of mental states and episodes because they are not the upshot of a procedure of coming to judge. Avowals proper are secure, because they are, in point of how you come to perform them, replacements for natural expressions:

> "when issuing an avowal proper, a subject's *epistemic* position vis-à-vis her own mental condition may be in important respects similar to her position when she expresses her state through smiling, sighing, or cheering. … What matters to the Neo-Expressivist account primarily is not the absence of a self-judgment mediating between the mental state and the avowal, but rather the irrelevance of any such judgment to the treatment of the avowal as a secure performance, protected from epistemic criticism or correction." (*Ibid*., p. 258.)

Reportive avowals are secure precisely to the extent that they resemble avowals proper:

> "insofar as we take [a subject] to be avowing, we take it that she is also expressing a state she is in, and not merely presenting her findings about a state inside her. Furthermore, we would take her response to have a unique security *only insofar as we take it that way*." (*Ibid*., p. 302. Emphasis added.)

Salience. As noted earlier, conscious states and episodes can be expressed in behaviour, in all but the most abnormal circumstances. If a subject possesses, in her repertoire of expressive acts, the ability to utter the appropriate self-ascriptive sentences, she will be in a position correctly to self-ascribe her conscious states and episodes.

First-person privilege. Only you can come to self-ascribe a particular mental state *by* expressing—by giving vent to—that mental state.

Authority. Although self-ascriptive acts do put forth self-ascriptive propositions for which the questions of truth and justification arise, insofar as those acts are expressive of the states and episodes they ascribe, rather than assertions about them, it is inappropriate to gainsay or demand justification for the propositions put forth:

> "if a self-ascription such as "I am annoyed by your reaction" is issued as an expression of one's state, then it is as inappropriate to raise questions about the epistemic grounding of the self-

> ascriber's *judgment that* she is annoyed or of various aspects of it as it would be to ask such questions of a person who utters "Your reaction is so annoying" or "Darn!" (or, for that matter, of a person who stomps her foot, or makes some other non-verbal annoyed gesture or vocal emission)." (*Ibid*., p. 263. Emphasis in original.)

Immediacy. In so far as self-ascriptive utterances are acts of expressing, of giving vent to, mental states and episodes, they of course are not arrived at by acquiring or attending to grounds for judgments about those mental states and episodes; nor is giving vent to a mental state or episode a matter of finding out about it.

I now want to argue that several of the distinctive features of self-ascriptions cannot really be explained, in the manner just outlined, by expressivist resources, consistently with the fact that the self-ascriptions which have those features can constitute knowledge. To respect the fact that self-ascriptions constitute self-knowledge, the neo-expressivist must construe self-ascriptions as acts of judgment that have epistemic credentials in virtue of how the subject comes to judge (see above, section 2.2). If self-ascriptions are construed in this way, the neo-expressivist's explanations of several of the distinctive features of self-knowledge fail.

The claim is not that the neo-expressivist's claims about the role of self-ascriptions, as outlined in the first part of this section, are *incompatible* with the principle that those self-ascriptions can be knowledge. Nor is it that the account entails that the distinctive features of self-ascriptions are *inexplicable*. Rather, the claim is that the putative expressive role of self-ascriptions can't explain their distinctive features, compatibly with their being knowledge. The specialness of self-knowledge can't be explained by the resources specific to neo-expressivism. Thus, as far as the problem of self-knowledge goes, neo-expressivism offers no part of a solution. And that leaves us needing an epistemic solution after all.

I am focusing here on neo-expressivism, as the only worked out weakly deflationist view in the literature,[27] but the criticism applies to any view that attempts to explain the distinctive features of self-knowledge by appeal to expressivist resources, without denying that it is knowledge.

Let me clear that the criticisms I am about to state do not start from a neutral position on the nature of the phenomenon that needs explaining. For one thing, I am taking it that the distinctive features of self-knowledge are features of *acts* of self-ascriptive judgment, not features of the *products* of those acts. It is not merely the sentences that self-ascribers utter

27 Apart from that of Wright (1991, 1998). See note 1 above.

that tend to be true. Those utterances articulate securely true judgments. Bar-On may disagree with me on this point. Secondly, I am taking it that, since self-ascriptions constitute or manifest knowledge, they are acts of judgment with epistemic credentials, which they have in virtue of the way they are arrived at. Bar-On would almost certainly disagree with this. So I am starting from a view of the phenomenon to be explained that would, I suspect, be rejected by Bar-On. However, I have given arguments that this view is correct. What I now want to show is that her account cannot explain the phenomenon, so understood.

Firstly, if the neo-expressivist accepts that self-ascriptions are, or express, self-ascriptive judgments, her explanation of authority fails. The neo-expressivist may be right that a self-ascriptive act is an act of expressing the ascribed state, but if it is *also* an act of judging that you are in the ascribed state, then it seems that the questions of the truth of and justification for that judgment arise. Challenging or gainsaying an utterance is inappropriate if that utterance is *not* an articulation of a judgment. But it doesn't follow that challenging or gainsaying an utterance is inappropriate if that utterance is something else *in addition to* being an articulation of a judgment. If, in saying "I think this curry is delicious", you really are articulating the self-ascriptive judgment that you judge that the curry is delicious (whatever else you are doing), then it seems as though an interlocutor ought to be able to say "No you don't", or "How do you know?". These reactions aren't made inappropriate merely by the fact that you are *also* expressing your first-order judgment that the curry is delicious. (If it *is* a fact, then someone who gainsays your self-ascription will be mistaken. That doesn't mean they are doing something inappropriate.) Even if self-ascriptions are in some respects akin to natural expressions, they are also in many respects *not* akin to them—and in these latter respects it seems as though they ought to be open to challenge and gainsaying.

The point here is not that the authority of self-ascriptions entails that self-ascriptive acts have no role in expressing what they ascribe. It is that their expressive role is not the correct explanation of authority.

Secondly, if the neo-expressivist accepts that self-ascriptions are, or express, self-ascriptive judgments, her explanation of immediacy fails. That explanation is that in so far as self-ascriptions are expressive, they are produced directly from mental states, rather than from some satisfactory way of ascertaining that certain conditions obtain. Again, there is an ambiguity here. If self-ascriptions were *not* acts of judgment, they would of course not involve ascertaining the obtaining of truth-conditions, and so we would be able to see why they are immediate. But self-ascriptions *are* acts of judgment. An act of judgment, I argued in section 2.2, involves the acceptance that the truth-conditions of some grasped content

obtain. But this raises the question of how a subject can come, in some epistemically satisfactory way, to accept that she is enjoying some conscious state or episode, without that acceptance being arrived at by acquiring or attending to grounds for thinking that she is enjoying that state or episode. After all, we need to see self-ascriptions as the upshot of a satisfactory way of coming to accept that some conditions obtain, and not as the mere parroting of a sentence of a certain form. It seems, *prima facie*, that any satisfactory way of coming to accept that a contingent, empirical condition obtains, would have to involve acquiring or attending to grounds for thinking that that condition obtains. This seems like a *general* requirement.[28] The fact that self-ascriptions are something else *in addition* to being judgments (this is the second reading of the neo-expressivist's explanation), does nothing to explain why they are not subject to what *prima facie* appears to be a *general* requirement on knowledgeable judgments. It doesn't explain why the requirement is waived in this case. So the neo-expressivist explanation of immediacy either is incompatible with the phenomenon being explained, or fails to explain it.

Again, my claim here is not that the neo-expressivist account of self-ascriptive utterances is incompatible with immediacy. Rather, my claim is that immediacy is puzzling, and that neo-expressivism doesn't alleviate the puzzlement.

It would not do for the neo-expressivist to appeal here to the fact that self-ascriptions are apparently immune to errors of misidentification (see above, and section 2.2.2). Such immunity shows that: the grounds (if any) for self-ascriptive judgments do not consist, even in part, in grounds for *identifying* a particular person (oneself), from among several, as the person who is M-ing that *p*; or in grounds for identifying a particular type of state or episode, from among several, as the type you are enjoying; or in grounds for identifying a particular content, from among several, as its content. That a judgment is immune to certain errors does not show that the judgment as a whole lacks grounds of any sort, or is not arrived at by acquiring or attending to grounds. When you judge, by proprioception, that your legs are crossed, that judgment is immune to a certain kind of error: you won't be right that someone's legs are crossed, but go wrong about whose legs are crossed. Nonetheless, you have, and presumably attend to, grounds for thinking that *your* legs are crossed. Those grounds—the deliverances of proprioception—are of a sort that you could not have for a judgment about anyone else's legs. They are grounds for making a judgment about yourself, even though they are not grounds for *identifying* yourself as the person who instantiates

28 I am assuming here that a way of coming to judge must be more than merely reliable in order to count as epistemically satisfactory. See Chapter 1, section 1.4.

some property that you have *independent* reason to think is instantiated.

So there is no special connection between immunity to errors of misidentification, and immediacy. The neo-expressivist cannot explain immediacy by appealing to such immunity.

Thirdly, if the neo-expressivist accepts that self-ascriptions are knowledge, then her explanation of security fails. I argued in section 2.2 that, if a judgment constitutes knowledge, it must have epistemic credentials, consisting in a connection to something that provides warrant, sustained by the way in which the subject comes to judge—and this is a *general* requirement (aside from analytic and self-verifying judgments). Self-ascriptions are secure, the neo-expressivist claims, in so far as they are arrived at in the same way that cries, groans, and other natural expressions are arrived at, by giving vent to the state ascribed, rather than being the upshot of an epistemically satisfactory way of coming to judge. Once again, this explanation is ambiguous: on one reading it is incompatible with the phenomenon, and on the other it fails to explain it. On the first reading, the claim is that self-ascriptions are secure because they are *not* the upshot of an epistemically satisfactory way of coming to judge. This is false: I have argued that they are the upshot of a satisfactory way of coming to judge, in virtue of which they have good epistemic credentials. On the second reading, the claim is that self-ascriptions are secure because they are something else *in addition to* being the upshot of an epistemically satisfactory way of coming to judge. This is not yet an explanation. As long as it is agreed that they are the upshot of such a way of coming to judge (whatever else they might be), we face the question of why that way of coming to judge seems to be less fallible than other methods for coming to make judgments.

Bar-On may say that security consists simply in the sort of immunity to errors of misidentification that I have already mentioned. She holds that the way in which we come to make self-ascriptive judgments does not allow for the possibility of error because it does not involve the identification of a particular person, of a particular type of mental state or episode, or of a particular content. It does not involve such identifications because we come to make such judgments by speaking from, by giving vent to, our mental states and episodes. What's more, such immunity to error through misidentification is perfectly compatible with knowledge. For example, you can know by proprioception that your legs are crossed, even though you have no distinct basis for judging that it is *you*, rather than anyone else, whose legs are crossed (Bar-On, *ibid*., e.g. pp. 357ff.).

The first point to make in response to this is that the cases of immunity to error through misidentification that are familiar from the literature do not provide a helpful analogy for the neo-expressivist. Those other familiar cases are explained in part by the fact that the relevant

judgments are based on epistemic access of a certain sort to certain realms of facts. You can judge "My legs are crossed" by proprioception, and your judgment will be immune to errors of misidentification of *whose* legs are crossed, because you have fallible epistemic access to facts about the positions of your limbs—access of a sort you can enjoy only to facts about your own limbs. Similarly, when you judge "It is raining here" by perception, your judgment will be immune to errors of misidentification of *where* it is raining. Nevertheless, your judgment is based on fallible epistemic access, through perception, to facts about your environment. It is only *because* you have such access, and because perception is tied to your immediate environment in a certain way, that you can make a judgment that is immune to such error.

Thus, immunity in each of these familiar cases is explained by the availability of a certain sort of epistemic access, which by its nature is restricted to facts about a certain person or location. This is entirely contrary to Bar-On's explanation of the security of self-ascriptions: she denies that any such epistemic access is involved in the explanation of the security of self-ascriptions. Thus, assimilating self-ascriptions to these other cases does not really explain why subjects, when expressing their conscious states and episodes, produce securely true judgments.

The second point to make is that, in any case, immunity to errors of misidentification does not amount to security. Errors of misidentification are those errors where you have latched onto *some* state of affairs, but misidentified some feature of it or object in it. Security also involves an absence of false positives: you will not easily judge "I M that *p*" when you haven't latched on to *any* state of affairs. A false positive is not an error of misidentification. You can mistakenly judge "It is raining here", on the basis of a misleading visual experience, without there being any misidentification involved; similarly, you can mistakenly judge "I M that *p*" without there being any particular conscious episode whose subject, type or content you have misidentified. To explain the absence of errors of misidentification is not to explain the absence of false positives. Even if subjects come to issue self-ascriptive judgments by expressing their conscious states and episodes, thereby ruling out errors of misidentification, we can ask: why shouldn't the resulting judgments sometimes be false positives?

The fact that expressive acts are tied closely to the states and episodes they express does not seem to explain why the self-ascriptive judgments that those acts sometimes involve almost never go wrong. It does not explain why the content of your judgment, and the state or episode you are expressing, will always go together. The fact that you are expressing your conscious state or episode leaves it open *which* judgment you make; and making the wrong

judgment need not be an error of misidentification.[29]

None of this is to deny Bar-On's claim that self-ascriptions are immune to errors of misidentification. It is merely to say that this claim does not do anything to show that the correct explanation of security is non-epistemic, and to say that Bar-On's non-epistemic explanation of security fails.

The particular problems I have identified all arise from a deep tension within any weakly deflationist account that attempts to explain the specialness of self-knowledge by appeal to expressivist resources. It is a tension between, on the one hand, the expressivist-inspired claims on which such an account relies, and, on the other hand, the sorts of claims we must make in explaining the knowledgeableness of self-knowledge. The expressivist-inspired explanation of security, authority and immediacy depends on emphasising the kinship between self-ascriptive utterances, including what Bar-On calls 'reportive avowals', and 'natural expressions' like frowns, grunts and gestures, and de-emphasising the extent to which such utterances articulate judgments like any other. On the other hand, the conditions for such self-ascriptive utterances to be knowledgeable depend on their being qualitatively different, in several important respects, from natural expressions. What I have argued, in effect, is that there is no stable position that embraces both poles of this tension.

Thus, the weak deflationist cannot simultaneously avoid epistemic deflationism and explain the security, authority and immediacy of self-knowledge. I have argued that epistemic deflationism is unacceptable. And, if the weak deflationist foregoes her claim to explain security, authority and immediacy, there will be little left to recommend her account. We must therefore look elsewhere, to an epistemic account, for an explanation of the distinctiveness of self-knowledge. If we can find an epistemic account that explains all five distinctive features of self-knowledge, we should prefer it to any deflationist approach.

2.4 Conclusion

In this chapter I argued, first, that the phenomenon I have been calling 'self-knowledge' really is first-personal knowledge about conscious mental states and episodes. The argument centred on the role that that phenomenon can play in inference, action and communication. I argued, further, that the weakly deflationist approach, represented by (a certain reading of)

29 Bar-On, I think, would say that an absence of false positives of the form "I M that *p*" is ensured by your being a competent user of the terms or concepts involved in that judgment. I agree with this. The question is what is involved in such competence. It cannot be merely the disposition to think or utter "I M that *p*" in place of a natural expression such as a groan—or so I have argued.

neo-expressivism, fails. I claimed that knowledge in general, and therefore self-knowledge in particular, requires epistemic credentials, and that epistemic credentials attach to acts of judgment when subjects come to make those judgments in epistemically satisfactory ways. I claimed that weak deflationism is incompatible with these facts.

The case for an epistemic approach is complete only when it is made plausible that there can *be* a successful epistemic account. Part of the motivation behind the various deflationary approaches is that no non-deflationist (i.e. epistemic) account can succeed. In Chapters 5 and 6 I will try to remove this motivation by developing my own epistemic account.

Before getting to that task, I must show why one historically influential sort of epistemic account fails. That is the task of the next chapter.

Let me finish by noting that many of the insights of expressivist approaches to self-knowledge are not impugned by the arguments I have offered. It is surely true that self-ascriptions in thought and speech are often expressive of the subject's mental state, and even that the subject's communicative intention is often nothing more than to express her state. There is also something importantly right about the idea that we speak from our mental states. I have argued only that the specialness of the resulting judgments is not explained by their being so produced.

The account I eventually offer will respect these insights. I will claim that self-ascriptions are indeed produced directly from conscious states and episodes, rather than from introspection of those conscious states and episodes. What's more, willingness to issue self-ascriptions is grounded in the nature of the conscious states and episodes they self-ascribe. No wonder, then, that self-ascriptions can be thought of as speaking from the states and episodes they ascribe, that their connections to those states and episodes are not merely causal, and that they are thought to reveal conscious states and episodes in an especially direct way.

As Bar-On rightly argues, we ought not understand the notion of speaking from a mental state in such a way that, in speaking from your state, you can't also be expressing a knowledgeable judgment about it. My conclusion is that her own account, if it is to respect this point, must forego its claim to explain the specialness of self-knowledge.

CHAPTER 3

INTROSPECTION AND CONSCIOUSNESS

In the last chapter I argued that self-ascriptions are often knowledgeable, and so have epistemic credentials, and that we ought to look for an explanation of the specialness of self-knowledge that appeals to those credentials. The rest of the thesis will be devoted to the question of what those epistemic credentials are and how they explain that specialness. In this chapter I want to consider one influential and historically important sort of account.

The sort of account I want to consider is what I will call 'introspectionist'. An introspectionist account is one according to which self-knowledge is based on introspection. I am using the term 'introspection' in a relatively narrow sense. Some philosophers use 'introspection' to refer to the method, *whatever* it is, by means of which we come to know about our own mental states and episodes. In my sense 'introspection' refers to a particular method by means of which, it has been claimed, we come to know about our own mental states and episodes—a method analogous in some respects to perception. In this sense, introspection involves experiential awareness of a mental state or episode, in virtue of which it seems to the subject that she is enjoying that state or episode. A self-ascriptive judgment is an endorsement of how things seem to the subject when she enjoys introspective experience, on the introspectionist view.

The claim I am arguing for in this chapter is that any introspectionist view fails to account for certain ways in which self-knowledge is unlike other kinds of knowledge, such as perceptual knowledge. This claim is not new; it is widely agreed that introspectionist accounts face very serious difficulties. But many of the arguments in the literature are effective against only some versions of introspectionism. I think that a general examination of why introspectionism fails can deepen our understanding of the way in which self-knowledge is unlike other kinds of knowledge.

The claim is not that there is no such thing as introspective experience. It is merely that our characteristic self-knowledge of our own conscious states and episodes isn't generally based on such experience.

The argument of the chapter will go as follows. In section 3.1 I will set out the minimal commitments of any view worthy of the label 'introspectionist'. This will involve distinguishing two different sorts of introspectionist view: the Lockean and the Cartesian. In

3.2 I will give three objections, of increasing depth, that apply to both sorts of introspectionism. First, introspectionism is incompatible with one aspect of the authority of self-knowledge. Second, the procedure for self-ascribing conscious states and episodes often involves directing one's attention out at the intentional objects of those states and episodes, not inwardly at the states and episodes themselves. Third, introspectionism cannot capture the essential rational connections between self-ascriptions of certain mental states and episodes, and the commitments of, and responsibility for, those states and episodes themselves. In 3.3 I will argue that while Cartesian introspectionism is correct to emphasise the connection between self-knowledge and consciousness, its failure suggests that we should embrace a rather different picture of the connection. I will conclude by suggesting that these difficulties for introspectionism can furnish us with a refined understanding of the distinctiveness of self-knowledge.

3.1 The commitments of introspectionism

Introspectionism holds that self-knowledge is based on introspection. What does this mean? Introspection, on this view, involves a type of experiential awareness of your own conscious mental states and episodes. This introspective experience takes mental states and episodes as its objects. In enjoying such an experience of a mental state or episode of yours, you experience certain features of that state or episode: its type and its content, for example. In virtue of this experience, it seems to you that you are enjoying a state or episode of that type, with that content. You can achieve self-knowledge by basing a self-ascriptive judgment on this experience—by endorsing in judgment how things seem to you in the experience. Introspective experience thus mediates epistemically between, on the one hand, a subject's mental states and episodes, and, on the other hand, her self-knowledge of those states and episodes: it epistemically grounds self-knowledge by providing access to those states and episodes.

To say that introspection involves an experience is just to say that it involves a conscious state with representational content, in virtue of which it seems to the subject that the content is true, and that this state is not a judgment or belief. In other words, introspective experience is not only (like conscious thought) a modification of consciousness, but also (like perceptual experience) a seeming.[1] However, introspectionism need not claim that introspective experience is sensory or that it is a type of perception. There are experiential seemings that

1 I will discuss this distinction further in section 3.3.

are neither sensory nor perceptual, such as occurrent propositional memories. Perhaps introspection is like those.

That is a rough characterisation of introspectionism. To achieve a more precise characterisation we must address the question of what it is (or what it would be) to experience introspectively a mental state or episode. Introspectionism is often associated with the metaphor of inner perception, but introspectionist views vary greatly in how closely they base their conception of introspective experience on the model of inner perception. There are, however, certain minimal commitments that any introspectionist view will share. To bring these out, I want to consider two different introspectionist views.

The first type of introspectionist view we should consider takes the analogy with perception very seriously. It conceives of introspection as subserved by something like a perceptual *modality*, or a dedicated cognitive *faculty*, by means of which we come to be experientially aware of our own mental states and episodes. The modality or faculty is individuated by the way it functions to yield awareness of a particular range of objects, namely our mental states and episodes. We achieve this experiential awareness when we direct our attention to our mental states and episodes, via the introspective modality, as we can become aware of objects in the environment when we direct attention to them by means of vision, or of the positions of our limbs when we attend to them by proprioception. Knowledge based on introspection can thus be compared to knowledge based on visual observation or on proprioception.

This type of introspectionism receives a classic statement in Locke's *Essay Concerning Human Understanding*:

> "the other fountain from which experience furnisheth the understanding with ideas is,- the perception of the operations of our own mind within us, as it is employed about the ideas it has got[…]. And such are perception, thinking, doubting, believing, reasoning, knowing, willing, and all the different actings of our own minds;- which we being conscious of, and observing in ourselves, do from these receive into our understandings as distinct ideas as we do from bodies affecting our senses. This source of ideas every man has wholly in himself; and though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be called internal sense." (Locke, 1700, II.1.v.)

A similar sort of view is arguably taken by Berkeley and Hume. Descendants of Locke's view include Lycan's (1996) 'higher-order perception' theory of consciousness. Lycan says: "introspection is the operation of an internal attention mechanism that monitors experiences

and produces second-order representations of their properties." (Lycan, 2003, p. 26.)[2]

Whatever the perceptual analogy at the heart of this Lockean view precisely amounts to, it entails a couple of commitments that are not essential to the general introspectionist picture. Pointing up these commitments will enable me to introduce the second type of introspectionism, which avoids those commitments, and then to identify what is really essential to the picture.

Firstly, the Lockean view conceives of introspection as involving two metaphysically distinct states or episodes. There is a mental state or episode, and a distinct introspective experience of that state or episode.

Secondly, this type of introspectionism has it that when there is an introspective experience of a mental state, the mental state and the introspective experience are related *causally*. Veridical perception of an object, at least where the perception and the object are metaphysically distinct, involves a causal dependence of the perception on the object. If one took introspective experience to be distinct from its object, and yet not to depend causally on it, then the analogy with perception would disintegrate.

Although both of these commitments flow from the analogy between introspection and perception, neither of them is essential to the general introspectionist picture. One can claim that self-knowledge is based on experiences directed on mental states and episodes, without claiming that those experiences are analogous to perceptual experiences in every respect. One can deny that introspective experience is a *distinct* state or episode, occurring between the introspected state and the self-ascription of it. One can hold that the introspected state and the introspective experience of it are not metaphysically distinct existences, and, accordingly, that the relation between them is not causal, but constitutive.

This leads us to the second type of introspectionism: Cartesian introspectionism. On this view, introspective experience is constitutive of the occurrence of a conscious episode or state. A state or episode's occurring *is* the subject's being experientially aware *of* it, or in a position to experience it, in a certain way. Introspective experience is not, as the Lockean would have it, a matter of an inner spotlight that illuminates goings-on that can occur quite happily in the dark; rather, those goings-on have a certain intrinsic luminosity, and so by

2 Higher-order *thought* theories of consciousness—those that conceive of the higher-order state as a thought, belief or judgment rather than as a seeming—are not introspectionist. They posit an epistemically unmediated connection between any conscious episode and the subject's judgment about that episode. There is nothing playing a role analogous to perceptual experience, on these views. By contrast, introspectionism holds that self-ascriptive judgments are based on experiences *of* conscious states and episodes.

their very nature present themselves to awareness.[3] Consciousness, on this view, has a *self-intimating* character: a conscious state with a particular object makes its subject experientially aware not only of that object but also of the state itself.

The Cartesian view, then, is that whenever you enjoy a conscious state or episode, you *ipso facto* enjoy an introspective experience of that state or episode. The experience represents its type and its content. In virtue of this introspective experience, it seems to you that you are enjoying a state or episode of that type, with that content. A self-ascriptive judgment arrived at by accepting how things seem to you in introspective awareness, ordinarily is knowledge.

The view that there is a metaphysical connection between at least some types of mental states and the self-awareness of those states has been attributed to Descartes (see Ryle, 1949) and can be found in Chisholm (1981), Gertler (2001) and Chalmers (2002).

Cartesian introspectionism is thus based on a very different conception of introspection to that of Lockean introspectionism. In particular, introspection is not inner perception, on the Cartesian view, if this suggests an object that is independent of the perceiving of it. Nevertheless, there are certain commitments that the Cartesian view shares with the Lockean view. These are the commitments that any introspectionist view must have, and that introspectionism must be assessed on.

Firstly, self-ascriptive judgments are epistemically grounded by experiences that constitute access to the states and episodes that those judgments are about. Thus, introspective experience still mediates *epistemically* between conscious states and episodes, and self-ascriptive judgments about those states and episodes—even though, on the Cartesian view, nothing mediates *metaphysically* between conscious states and episodes, and judgments about them. Introspective experience can play this epistemic role because it is a seeming which represents a certain state of affairs as obtaining. Thus, introspective experience plays a similar epistemic role to perceptual experience, even if it is metaphysically rather different. This is the minimum analogy between introspection and perception to which any introspectionist view is committed.

To say that an introspective experience mediates epistemically between the subject-matter of your self-ascriptive judgment and the judgment itself is *not* to say that the judgment is based on the *fact that* you are having the introspective experience. The judgment is based on the fact to which the introspective experience constitutes a mode of access—the fact that you are in a certain mental state, say. The experience epistemically grounds the judgment, but the

3 The metaphor is inspired by Ryle (1949, pp. 152 ff.), who finds this view in Descartes.

occurrence of the experience is not itself the grounds for the judgment, any more than the occurrence of a perceptual experience is the grounds for a perceptual judgment.

Secondly, conscious states and episodes are *objects* of awareness, in a certain sense, on an introspectionist view—they are represented as occurring in introspective *experiences*. It is important, here, to distinguish between, on the one hand, awareness *of* something, and, on the other hand, awareness *that* something is the case.[4] Introspectionism does not claim merely that you enjoy awareness *that* you are in certain mental states, when you are. It claims that you enjoy experiential awareness *of* the conscious states and episodes you have. It is in that sense that they are objects of awareness.[5]

On the Cartesian version of introspectionism, introspective experience has the feature that the experienced object and the awareness of it are not metaphysically distinct: there is one complex state that involves an experience of itself. It would be misleading to say that on this view there is no distinction of awareness and object, for we can still conceptually distinguish one's awareness, on the one hand, and that of which one is aware, on the other, and there is no doubt that the latter is the experienced object of the former. What is true is that there is no *metaphysical* distinction between them: the awareness is its own experienced object.

When something is an object of experiential awareness, you are typically aware of features of it that can be used in recognising and identifying it.[6] In the visual case, these would be features like colour, shape, and so on. In introspection, these features would include the intentional content of the state or episode, and its type, or introspectible features that indicate its type.[7]

Thirdly, an introspectionist account must give some role to attention. Judgments based on experience typically involve attending to the deliverances of experience; for example, you judge that the foliage is yellow by attending to the apparent colour of the foliage, as presented in vision. Making a self-ascriptive judgment, on the introspectionist view, will involve attending not to the intentional object of the self-ascribed state or episode, but to the deliverances of introspective experience. If you are seeing the yellow foliage and someone

4 As emphasised by Dretske (1999).

5 This may not mean that they are objects of attention, even though attention clearly has a role to play in the introspectionist account. See below.

6 This is Dretske's "property-awareness" (Dretske, 1999). Dretske emphasises ways in which property-awareness, object-awareness and fact-awareness can come apart. Nevertheless, if we know about our own mental states and episodes through introspective experience, then that experience must involve awareness of the properties of those states and episodes.

7 William James defends a view on which certain mental activities possess introspectible features that enable the subject to tell what activity she is engaged in. See James (1976).

asks you about your present state of consciousness, you will shift your attention from the yellow foliage to the visual experience itself.

On the Lockean view, this shift will be like a shift of perceptual attention in which attention switches between the deliverances of different sensory modalities. You switch attention from the deliverances of vision, say, to the deliverances of introspection—just as, for example, if someone asked you about the positions of your limbs, you would switch your attention to the deliverances of proprioception.

On the Cartesian view, there is no switching between modalities, since introspection is not seen as a distinct modality. The intentional object of your experience and the experience itself are apprehended in the same act of awareness. But there is a shift of attention within that act of awareness. This is a familiar phenomenon. Consider aspect shifts. When you look at Wittgenstein's duck-rabbit, you can bring it about that you see it now in one aspect, now in the other, by a sort of shifting of attention. But you don't do this by moving your attentional spotlight to a new object. At some level, your experience remains the same throughout. Similarly, when making a judgment about your mental state, you shift your attention to your mental state, rather than to its intentional object, even though (on the Cartesian view) you will have been experientially aware of both the state and its object all along. The shift of attention is not selection of a new object of experience; it is, like an aspect shift, a shift of attention within an experience.

The Cartesian may wish to appeal here to Peacocke's distinction between two ways in which something can determine attention (Peacocke, 1998). Something can be an *object of* attention, or it can *occupy* attention. The intentional object of a perception can be an object of attention. For example, when you visually examine a tree, that tree is an object of attention for you. Conscious episodes themselves, according to Peacocke, *occupy* the subject's attention, without being objects of attention for the subject. Thus, your experience of the tree, and your activity of examining it, both occupy your attention, but are not objects of attention in the same sense that the tree is. The Cartesian may wish to say that when a conscious episode is directed on an object of attention, the subject is experientially aware of the episode itself (not just that object) because the episode occupies attention.

The fourth commitment of introspectionism is one I have already noted. In enjoying an introspective experience *of* a mental state or episode, which represents that state or episode as of a particular type and content, it seems to the subject *that* she is enjoying a state or episode of that type with that content.

Fifthly, introspective experience is not belief or judgment. In enjoying an introspective experience of your mental state, you are not yet taking it that you are in that state. The introspectionist can allow that it is possible to withhold endorsement of the deliverances of introspection (even if only in the throes of misguided philosophising). Even when you withhold endorsement of the content of your experience, it still seems to you introspectively that you are in the mental state that experience represents—just as it seems to you in perceptual experience that things are as your experience represents them, even when you don't believe that they are.8

I take these five commitments to be essential to introspectionism. Before going on to the criticisms of the view, let us see how introspectionism purports to give an epistemic solution to the problem of self-knowledge.

The first part of a solution is to answer the *knowledge* question: why are self-ascriptions knowledge? According to introspectionism, self-ascriptions are knowledge when they are based on the deliverances of introspective experience. These experiences provide warrant, in part because they provide reliable access to mental states and episodes. Self-ascriptions thus have warrants similar to those of perceptual judgments based on the deliverances of perceptual experience.

The second part of a solution to the problem of self-knowledge is to answer the *specialness* question: why does self-knowledge appear to have the distinctive features of security, salience, first-person privilege, authority and immediacy? The introspectionist again appeals to the nature of introspective experience. The security and salience of self-knowledge are inherited from the security and comprehensiveness of our introspective access to our conscious mental states and episodes: introspective experience is extraordinarily reliable, and mental states and episodes do not easily escape its grasp. Self-knowledge is first-person privileged because introspection provides access only to your own mental states and episodes. Authority and immediacy may be more difficult to explain, given the parallels drawn by the introspectionist between introspection and perception, and the absence of those features in the perceptual case. I will discuss authority below. Although the success or otherwise of these explanations—whose details would vary from one introspectionist view to another—is important to the plausibility of introspectionism, it is not the part that I want to

8 In characterising the view as involving this commitment, I exclude Armstrong's so-called 'inner sense' view (Armstrong, 1981) from the category of introspectionist views. Armstrong's inner sense does not involve a state of awareness mediating between the first-order state and the self-ascriptive judgment. The first-order state directly causes the judgment that one is in it. This seems to me to be a fundamentally different picture, cognitively and epistemologically, to the Lockean picture, which is also sometimes called 'inner sense'.

focus on. There are, I think, deeper objections to any account that tries to explain self-knowledge, and its distinctive features, by appeal to introspective experience.

3.2 Self-knowledge is not introspective

Introspectionism claims that self-knowledge of conscious states and episodes is based on experiential seemings; that is, it is acquired by accepting what seems to be the case in a certain class of experiences. With respect to your conscious states and episodes, you are, epistemically, like a perceiver with a privileged view.

This claim fundamentally mischaracterises the nature of our self-knowledge—the way we acquire it, the epistemology of it, and its role in our rational cognition—and it mischaracterises the subject's first-person perspective on her own consciousness. That is the claim of this section and the next. There have been many critiques of introspectionism (see e.g. Ryle, 1949; Shoemaker, 1994 Lecture I; Burge, 1996), but focusing on a few important general objections can bring out these points, and give us a deeper understanding of the phenomenon we are trying to account for. In this section I will offer three objections that make up a cumulative case against the idea that self-ascriptions are based on experiential seemings.

3.2.1 Authority and seeming

We saw in Chapter 1 that self-knowledge of conscious states and episodes is distinctively authoritative. I understand authority as a feature of our practice for reacting to self-ascriptions, rather than as an epistemic phenomenon. One aspect of authority is that a demand for justification, in response to a self-ascription, is inappropriate. If, in ordinary circumstances, you say, “It perceptually appears to me that the leaves are yellow”, or “I judge that it is windy”, it is inappropriate for an interlocutor to ask, “How do you know?” (or “How can you tell?”), where this is a challenge to the self-ascription itself rather than to the judgment self-ascribed.

In this respect, self-ascriptions stand in contrast to perceptual judgments. It can be appropriate to respond to a perceptual judgment by asking, “How do you know the foliage is yellow, not green?”; your answer might simply be, “I can see clearly, and that's how it looks to me.” This sort of exchange makes sense even in a scenario in which, for whatever reason, you are the only one in a position to see the foliage.

This contrast between self-ascriptions and perceptual judgments stands in need of explanation. The obvious explanation is that we don't demand justifications for self-ascriptions because, unlike for perceptual judgments, we don't expect subjects to be able to offer justifications. And this has implications for *the way in which we come to make self-ascriptive judgments.* It implies that the way in which you come to self-ascribe conscious states and episodes does *not* involve reliance on a seeming, which you could then articulate and offer as a justification. In point of how you come to judge, you are not in an analogous situation to a perceiver who has a privileged view. Whatever you rely on in coming to self-ascribe is not something you can offer as a justification.

Support for this explanation comes from the observation that we are not disposed, in remotely ordinary circumstances, to say things like, "It seems to me that I am having an apparent perception that the leaves are yellow, and not just imagining that they are", or "It seems to me that I am judging that it is windy, not that it is sunny".

If introspectionism were correct, these would be unremarkable things to say. They would accurately express one's reasons for certain self-ascriptive judgments. After all, introspectionism claims that you know what mental episode is occurring when it introspectively seems to you that that episode is occurring. The question "How do you know?", asked in response to a self-ascription, would admit of an answer. It should therefore be legitimate, on the introspectionist view, to ask it. The two cases—self-ascriptive judgment and perceptual judgment—should be parallel. But in fact it is not legitimate to ask "How do you know?" in the self-ascriptive case. There is a contrast rather than a parallel.

In sum, our discourse surrounding self-ascriptions suggests that, unlike perceptual judgments, they are not based on seemings at all.

I have, in effect, used inference to the best explanation to argue from the claim that we do not treat self-ascriptions as though they were based on seemings to the conclusion that they are not based on seemings. Might there be an alternative explanation of this aspect of authority—this failure to treat self-ascriptions as though they were based on seemings? It might be suggested that it is an artefact of some conversational principle or implicature. But I cannot imagine any principle or implicature that would forbid a demand for justification of a judgment based on a seeming. Of course, the introspectionist can point out that it would violate implicature, in certain circumstances, to assert "It seems to me that I am judging that *p*" when you are also prepared to assert, unqualified, "I am judging that *p*".[9] But it would

9 Specifically, it would violate the maxim of quantity—be as informative as required for present

equally violate implicature, in certain circumstances, to assert "It seems to me that the foliage is yellow" when you are also prepared to assert, unqualified, "The foliage is yellow". The point is that there are certain circumstances in which this would *not* violate implicature—such as when your justification for the unqualified claim is called into question. If introspectionism is correct, there should also be such circumstances in the self-ascriptive case. The introspectionist has not captured any contrast between the self-ascriptive and the perceptual cases.

So, introspectionism is in tension with an aspect of the authority of self-knowledge. I do not take this to be a decisive objection to introspectionism. As I said in Chapter 1, explaining every apparent feature of self-knowledge is not a *sine qua non* of an acceptable account. The next two objections to introspectionism go deeper.

3.2.2. Transparency

Evans writes, of a subject in a conscious mental state: "His internal state cannot in any sense become an *object* to him. (He is *in* it.)" (Evans 1982, p. 227.) If Evans is right, then it looks as though the introspectionist is guilty of a radical mischaracterisation of the relation between the subject and his conscious states and episodes (see the second of the commitments outlined above, section 3.1). But just what contrast is Evans drawing here?

Part of what Evans is getting at is brought out in the following famous passage:

> "in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me 'Do you think there is going to be a third world war?', I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*." (Evans, *ibid*., p. 225.)

In this passage Evans notes the *transparency of the self-ascribing procedure*.[10] He is talking in particular about self-knowledge of the content of a belief. The point is that you come to

---

purposes.

10 Some philosophers object to introspectionism on the basis of a different sort of transparency, namely the (alleged) transparency, or diaphanousness, of perceptual experience—i.e. the property that the only features of experience that are accessible to the experiencing subject are those to do with how it represents the objects of experience as being. I hope it is clear that, by "transparency", I mean something quite different. I am referring to a property of the procedure for coming to make self-ascriptive judgments.

know what the content of your belief is—whether it is *that there will be a third world war*, or *that there won't be a third world war*—not by shifting your attention *away* from the prospects of a third world war and onto the apparent features of the belief itself, but by attentive consideration of the prospects of a third world war. There is no role, in this procedure, for an introspective seeming that represents you as believing this or that.

Evans's insight is not limited to making-up-your-mind cases—to cases where you must first form a belief, in order to then know what you *do* believe.[11] You may have a long-held conviction about whether there will be a third world war. Still, if I ask you whether you think there will be a third world war, you do not perform an internal inspection to see if you have such a conviction. You simply consider whether there will be a third world war. The answer that comes to mind will of course be the content of your belief.

Evans's insight also has broad application to self-ascriptions of conscious states and episodes. Consider, first, the *content* component of self-ascriptions. If I ask you whether you are seeing that the foliage is yellow or that it is green, you will not consult the apparent features of your perceptual experience itself. You will attend to the foliage, in order to determine whether it is yellow or green. To know how you are experiencing the world, you attend to how the world seems in experience. If I ask you whether you judge that it is windy or that it is sunny, you consult the weather, as you take it to be. To know the contents of your judgments, you check how the world is, for you. You do not turn away from the world and check your mental goings-on.[12] If I ask you whether you are hoping for this or that outcome, you report on which of those outcomes is more attractive.

The point also applies to the *type* component of at least some self-ascriptions. If I ask you whether you are seeing or merely imagining that the foliage is yellow, you will look again at the world, as it seems to you in that experience: does the foliage really seem to be yellow? If I ask you whether you *judge*, or merely *suppose*, that it is windy, you do not attend to the introspectible features of your conscious act, in order to tell whether it is a judgment or a supposing. Again, you attend to the weather, as you take it to be: *is* it windy?

Thus, in many cases, self-ascriptions are arrived at without *attending to* the features of the self-ascribed state or episode, as allegedly represented in an introspective experiential seeming. The procedure is transparent. These transparent cases count as self-knowledge, and display all the usual distinctive properties thereof. This is incompatible with

11 Mike Martin emphasises this (see Martin, 1998).

12 It is worth noting again here that I am only talking about present-tense self-ascriptions. You may not be able to tell what you judged five minutes ago by checking how the world is (for you) now.

introspectionism, which holds that self-knowledge *generally* is based on experiences of conscious states and episodes, seemings that represent the type and content of the state or episode in question.[13]

How might the introspectionist reply to this argument against his view?

The introspectionist might appeal to the distinction between *occupying* and being an *object* of attention (see above, section 3.1). He might claim that, although Evans's insight (appropriately extended) shows that conscious states and episodes do not become objects of attention when we come to self-ascribe them, it does not show that they do not occupy attention. And, he might add, in occupying attention they are experienced.

But this reply does not really engage with the transparency objection. The insight on which the objection rests is that many self-ascriptions are not based on *any* experiential seeming that represents the conscious state or episode self-ascribed, or any of its features. Such a seeming has no role. This insight is indifferent as to whether we construe the experienced state or episode as an *object* of attention, or merely as an *occupier* of attention.

Another reply would be to point out that the transparency claim is not obviously correct for the type-component of *every* self-ascription of a conscious state and episode. In particular, it is arguably more compelling for those types of state and episode that in some sense aim at the truth, such as perceptual experience and judgment, than for those that don't, such as conscious hopes. If I ask, "Do you really *hope* for this outcome?", you must perhaps do more, in answering, than merely considering how attractive the outcome is.

Even if this is right, however, it does not rescue introspectionism, which purports to be a general account of self-knowledge of conscious states and episodes. If introspectionism fails for the content-component in all cases, and for the type-component in very many central cases, as I have argued, then it certainly fails as a general account, and we have reason to look for an alternative. The account that I will offer in Chapters 5-6 acknowledges that different types of state and episode may be known about in somewhat different ways.

I conclude that the transparency objection constitutes a serious and deep problem for introspectionism.

This objection captures something that is right about the expressivist and neo-expressivist views considered in Chapter 2. The transparency of the self-ascribing procedure yields a

13 Nothing I have said is meant to show that there is no such thing as introspection. It is meant to show that self-knowledge of conscious states and episodes is not generally based on introspective experiences.

sense in which self-ascribing can also be expressing, or speaking from, your mental state. You come to self-ascribe a judgment, say, by doing some of the things characteristic of forming or reaffirming a judgment; thus, the self-ascription will be associated with the commitment that the judgment involves. The next objection centres on a similar point.

3.2.3 Rational commitment and non-alienation

The third objection to introspectionism is that it fails to account satisfactorily for a rationally committing aspect that certain self-ascriptions have. It is symptomatic of this failure that the view allows for the possibility that a self-knowing subject would suffer from a certain kind of alienation—a kind of alienation that in fact does not occur, and if it did would seriously undermine the unity and rationality of the subject.

Certain conscious episodes, such as judgments and decisions, involve commitment on the part of the subject—to something's being the case, to doing something, or whatever. I will focus on the case of self-ascriptions of judgment, but the points I make apply to other cases too.

Consider the following errors of alienation:

(i) Making a judgment of the Moore-paradoxical form "I judge that *p*, but ~*p*".

(ii) Being indifferent, when endorsing a content "I judge that *p"*, to evidence that supports ~*p*, on the grounds that the evidence is not directly relevant to the psychological question of what your judgment in fact is.[14]

These errors are usually formulated with reference to the attitude of belief, but they seem equally to be errors when formulated with reference to of the act of judgment. The oddness of Moore-paradoxical sentences remains when we replace "I believe that *p*, but ~*p*" with "I judge that *p*, but ~*p*". Likewise, indifference to evidence relevant to whether *p* seems just as odd when self-ascribing the judgment that *p* as when self-ascribing the belief that *p*.

Let me also make clear that these are not merely errors of *assertion*. It is an error to *judge* a

14 Analogous errors to (i) and (ii) would include, respectively, judging "I am deciding to Φ, but I will not Φ", and being indifferent, when self-ascribing the decision to Φ, to considerations that count against Φing.

Moore-paradoxical content (the content expressed by a Moore-paradoxical sentence), just as much as it is an error to assert it. Equally, evidence pertinent to whether *p* should impinge on your *judgment* that you judge that *p*, just as much as on your assertion to that effect.

There are two important things to note about (i) and (ii). The first thing to note is that they are indeed errors, even though neither involves any contradiction or surface incoherence. The Moore-paradoxical content in (i) is not contradictory, and contents of that form are often true; and the evidence ignored in (ii) is indeed not directly relevant to the self-ascriptive judgment. The second thing to note is that we almost never make these errors.

The fact that (i) and (ii) are errors, and the fact that we don't make them, are evidence of the rationally committing aspect of self-ascriptions of the form "I judge that *p*". Ordinarily, when a subject self-ascribes the judgment that *p*, she rationally commits herself to that first-order judgment—that is, she commits herself to the truth of *p*. To make an error like (i) above is in effect to commit oneself a contradictory pair of propositions, the proposition that *p* and the proposition that ~*p*. To make an error like (ii) is in effect to fail to acknowledge one's commitment to *p*, or to fail to take responsibility for that commitment.

It is not just that self-ascriptions of judgment are typically *accompanied by* the commitment of the judgment self-ascribed. That would be guaranteed simply by self-ascriptions' being true. If that were the whole story, Moore-paradoxical judgments would merely be instances of factual errors about what one's judgment in fact is. What makes Moore-paradoxical judgments interesting is that they involve more than factual error: they involve a certain kind of incoherence of commitments. Making such a self-ascription, or coming to make it, is itself committing. Self-ascriptions of judgment *inherit*, and are not merely accompanied by, commitment.

This feature of self-ascriptions of judgment, their rationally committing aspect, is tied to their first-personal character. It marks an asymmetry between first-person ascriptions of judgment and third-person ascriptions of judgment. To ascribe to some other person the judgment that *p* is not to commit oneself to *p*; and it would be legitimate to ignore evidence against *p* in making such an ascription to another.

How is the rationally committing aspect of self-ascriptions of judgment to be explained? Why, when self-ascribing a judgment, do you commit yourself to that judgment, and how do we avoid errors of alienation so successfully?

Often, we explain the commitments incurred by an assertion or judgment, self-ascriptive or otherwise, by appealing to the content of the assertion or judgment, and the logical

entailments of that content. But the committing aspect of a self-ascriptive judgment "I judge that *p*" is not explained by the content of that self-ascriptive judgment. For one thing, the content of that judgment does nothing to imply the truth of *p*; it could be true while *p* was false. For another thing, there are circumstances in which one can make a judgment with the very same content, and yet *not* incur a commitment to *p*. Lying on the psychoanalyst's couch, you could be persuaded to self-ascribe the judgment that your parents mean to harm you, and at the same time sincerely judge that your parents don't mean to harm you. This would be a deviant case, in which your supposed first-order judgment that your parents mean to harm you was inaccessible to you by ordinary means, but there is no irrationality in conjoining the judgment that you judge that your parents mean to harm you, with the judgment that they do not, when the self-ascriptive judgment is arrived at in this deviant way. Similar remarks apply to cases of the same sort as (ii): your judgment, reached in this deviant way, that you judge that your parents mean to harm you, will be and ought to be insensitive to evidence that your parents mean no harm to you.

To find the correct explanation of the rationally committing aspect of self-ascriptions of judgment, we must first say what distinguishes the cases in which such self-ascriptions involve that aspect, from those cases in which they do not. What seems to distinguish them is the way in which the subject comes to make the self-ascriptive judgment. When you come to make a self-ascription in the ordinary, first-personal way—the way that leads to secure, authoritative self-knowledge—it will have the rationally committing aspect. When you come to make it in some other way, such as by considering behavioural evidence or by accepting the testimony of your psychoanalyst, it will not have the rationally committing aspect.

Thus, the rationally committing aspect of judgments that self-ascribe judgment must be explained, in part at least, by the way in which subjects come to make those self-ascriptive judgments—for that is what determines whether a particular self-ascription has a rationally committing aspect. According to introspectionism, subjects come to make such judgments by endorsing the contents of seemings. The problem is that coming to self-ascribe a judgment in this way will not involve you in any commitment to the first-order content of the judgment. A self-ascription made by an introspective method will not have a rationally committing aspect. No matter how forcefully it seems to you that you are judging that *p*, that will not amount to its seeming to you that *p*; and no matter how decisively you endorse the content of an introspective seeming, thus committing yourself to the content "I judge that *p*", that *in itself* will not in the slightest commit you to *p*. The procedure described by the introspectionist involves only an experience as of making a judgment, and the endorsement

of the content of that experience. These elements cannot by themselves suffice for a commitment to a content that is independent of, and known by the subject to be independent of, the content of that experience.[15] The introspectionist procedure doesn't seem to be different, in the right way, from procedures that do not involve rational commitment, such as relying on a psychoanalyst's testimony, or on behavioural evidence.

It is symptomatic of this problem that introspectionism leaves open that a subject might be alienated from the judgments she self-ascribes. A subject could self-ascribe those judgments that she seems to herself to be making, and yet in the same breath eschew commitment to those judgments, and responsibility for those commitments. Such a subject would, in effect, treat all of her self-ascriptions as being like the psychoanalytic case described above—she might say, 'It certainly seems introspectively as though I am judging that *p*, and I have no reason to doubt that I am', and she might thus accept that she is making the judgment, and at the same time she might refuse to assent to *p*.[16] This might happen because the subject's introspective experience of judging that *p* is illusory.

Our ordinary self-ascriptions of judgment in fact do not leave open the possibility of alienation. So they are not made in the way the introspectionist claims.

The introspectionist might claim, in response, that introspective experience is infallible, making such alienation impossible: a subject would never have an apparent introspective experience of judging that *p* without also in fact committing to *p*. But this response doesn't really make self-ascriptions rationally committing. It fails to capture the distinction I made earlier between a self-ascription's *inheriting* the commitment of the judgment it self-ascribes, and its merely being *accompanied by* that commitment. Even if a self-ascription of the judgment that *p*, based on introspection, was necessarily accompanied by a commitment to *p*, the self-ascription and the procedure for coming to make it would in themselves leave it open whether *p*—they would not inherit the commitment. A self-ascription based on introspection would not itself be committing. In this sense, it would still exemplify a certain sort of alienation. It would be like a self-ascription arrived at a perfectly reliable but third-personal

15 Of course, it may be that a self-ascription of a judgment (that *p*, say) does not count as *knowledge* unless the subject is also committed to the truth of *p*. If that is so, a subject who makes an error of alienation will not be knowledgeable about her judgment. But that is not to the point. The question is why coming to self-ascribe a judgment ordinarily involves a commitment to that judgment, in such a way that a subject who comes to self-ascribe in that way will not make an error of alienation. It's one thing to say that such errors would impugn self-knowledge; it's another thing to say why those errors don't occur.

16 The argument here is *not* the same as Shoemaker's argument from the impossibility of self-blindness (Shoemaker, 1994). A self-blind subject would be one who made judgments but did not know about them. The impossible case I am considering is one in which the subject self-ascribes judgments that she is not prepared to commit to.

route.

To clarify: my claim is *not* that introspectionism entails that alienation is not irrational. It is that an account of the ordinary procedure for coming to self-ascribe must contribute to the explanation of why we avoid alienation when we self-ascribe in that way. It must do so because the avoidance of alienation is peculiar to that way of coming to self-ascribe. And introspectionism fails to do so.

The introspectionist might step in here. He might claim that my demand—that introspectionism should explain why we avoid alienation—is too strong. For, there is another resource he can appeal to in such an explanation, besides his account of the procedure for coming to self-ascribe. He can claim that commitment and non-alienation are explained by the self-ascribing subject's *conceptual capacity*. According to this line of thought, it is a condition on a subject's possessing the concept of judgment that her self-ascriptions of judgment inherit the commitments of the ascribed judgment. So no subject that is conceptually capable of self-ascribing judgments will be alienated from the judgments she self-ascribes. And so, even if introspectionism *alone* doesn't explain commitment and non-alienation, it can do so when combined with a certain account of the concept of judgment.

This reply, on the part of the introspectionist, could come in either of two versions. The *simple concept reply* would say that it is simply a necessary condition on possession of the concept of judgment that a subject does not make errors like (i) and (ii) above. Thus, any subject capable of self-ascribing judgment will, in doing so, commit to the judgment she self-ascribes, and not be alienated from it. This leaves open that such self-ascriptions might be made by an introspective procedure.

But the simple concept reply is not correct. As we saw, there are circumstances under which judgments like (i) and (ii) above are not errors. Such judgments may not be errors when made on the psychoanalyst's couch. So avoidance of such judgments is not simply written into the possession-conditions of the concept of judgment. One could fully possess the concept of judgment, and yet be prepared to make judgments of the form "I judge that p, but $\sim p$", when one comes to make the self-ascription by a deviant route.

The *sophisticated concept reply* says that it is a necessary condition on possession of the concept of judgment that a subject does not make errors like (i) and (ii) *when the self-ascriptive judgment is made introspectively, i.e. by taking the content of introspective experience at face value*. On this view, the possession-conditions for the concept of judgment give a special role to the introspective procedure for coming to self-ascribe judgment,

ensuring that self-ascriptions made in this way are rationally committing and non-alienated. Commitment and non-alienation are thus explained for self-ascriptions made in this way, but not mistakenly entailed for the deviant cases, as they were by the simple concept reply. They are explained by possession of the concept of judgment, but only because possession of the concept is tied to introspection—thus, both the account of the concept and the account of the self-ascribing procedure are crucial.

My response is that the view envisaged in this reply does not offer any genuine explanatory power—no real explanation of commitment and non-alienation is being offered here. It is mere *ad hoc*ery. I do not deny that it is a necessary condition on possession of the concept of judgment that one will not make errors of alienation when one comes to self-ascribe a judgment in the ordinary first-personal way. But the question is what that way is, and why it has that distinctive property. My point has been that there must be an explanation, concerning that way of coming to make a self-ascriptive judgment, of why self-ascriptive judgments made in that way will incur first-order commitments and will thus not be alienated. Why *that* way, and not others? Without such an explanation, it will be quite mysterious why the concept of judgment should give a special role that particular way of coming to judge; the relevant possession-condition would seem arbitrary. No such explanation is in the offing for introspectionism.

We can see now that my demand, that introspectionism explain the rationally committing aspect of self-ascriptions, was not too strong after all. For, that aspect is present precisely when a self-ascription is made by the ordinary first-personal procedure, of which introspectionism is an account. And, no matter what connections we draw between the rational commitment of self-ascriptions and other cognitive capacities of a subject, we will still want some non *ad hoc* explanation of why self-ascriptions made in just *that* way have the rationally committing aspect.

That concludes the objection. It is, I think, the trickiest of the objections to formulate, but also the deepest. It appeals to a fundamental first-/third-person asymmetry in ascriptions of judgment, and of other committing conscious episodes. The problem for introspectionism, which underlies all of my objections, is that the procedure it posits for coming to make self-ascriptive judgments is not deeply first-personal enough to account for such first-/third-person asymmetries. Together these objections constitute a strong cumulative case against introspectionism. The case will be complete when I can show that there is a non-introspectionist alternative that can account for these asymmetries (Chapters 5-6).

3.3 Consciousness, experience and introspection

We seem to have ordinary first-person knowledge of precisely those mental states and episodes that are conscious; the conditions under which a state or episode is conscious are also the conditions under which its subject is in a position to have first-person knowledge of it. In this brief section I want to address the relation between consciousness and self-knowledge, and argue that it is rather different to how the introspectionist conceives it.

According to Cartesian introspectionism, and some versions of Lockean introspectionism (e.g. Lycan, 1996) you know about your conscious states and episodes because a state's or episode's being conscious consists in your having an experience of it—an experience on which you can base a self-ascriptive judgment. This is a temptingly neat account of the connection between consciousness and self-knowledge, which I have argued, on broadly epistemic grounds, must be rejected.

But this might give rise to a worry. Isn't it platitudinous that we are experientially aware of the states and episodes that make up our stream of consciousness? Isn't that just the nature of consciousness? And isn't it therefore overwhelmingly likely that the epistemology of self-knowledge will appeal to this experiential awareness?

In fact, this reasoning is compelling only when we conflate two different claims, involving two different notions of experience. There is a notion of experience that simply means *modification of consciousness*. In this sense, *any* conscious episode or state, be it a thought, a perception, or whatever, *is* an experience. It is a determination of the subject's consciousness. This is the notion of experience that Husserl expressed with the term 'Erlebnis'.[17] But there is a narrower notion of experience, referring to a type of episode or state, not a judgment or belief, that represents its content as true—a seeming. This is the notion of experience that is appealed to by the introspectionist account of self-knowledge (see section 3.1). To say that a thought or perception is a modification of consciousness is not yet to say that the thought or perception is itself experien*ced*, in the sense of something whose occurrence is represented in an experience, as understood in this latter sense.[18] It seems to be one thing to be a modification of consciousness, and another thing to be an object of experience.[19] So there is

17 See, for example, his *Ideen* (Husserl, 1982).

18 One might worry about whether it makes sense to talk of a perception being experienced. Certainly, you don't perceive your perceptions in some sensory modality. But recall (section 3.1) that the notion of experience employed by the introspectionist need not be restricted to sensory or perceptual experiences. He can hold that there are introspective experiences—conscious, occurrent seemings—that play a similar *epistemic* role to perceptual experiences, even though they are not perceptual or sensory.

19 This is related to Dretske's (1999) distinction between being aware *with* an experience, and being

room for the claim that the states and episodes that make up our stream of consciousness are not *ipso facto* themselves experienced, even though they are themselves, in the broad sense, experiences.

According to the Cartesian view, we should identify these two things: for a state or episode to be conscious—to be an Erlebnis—is for it to be experienced.[20] This is implausible for at least two quite general reasons.

Firstly, the view faces a dilemma. Suppose the view is correct: a particular mental episode is conscious in virtue of the occurrence of an experience of that episode. We can ask: must that experience itself be conscious? Both answers lead to ruin. If the experience need not be conscious, then it is wholly mysterious how it could confer consciousness on anything else.[21] It doesn't seem plausible that a non-conscious state or episode becomes conscious in virtue of having another non-conscious state or episode directed on it. If there can be non-conscious experiences, and if those experiences can be directed on other mental states and episodes, then surely they can be directed on states and episodes that are themselves non-conscious, as well as on states and episodes that are conscious. So this horn of the dilemma is hopeless. On the other hand, if the experience must itself be conscious in order to confer consciousness on the episode on which it is directed, then a vicious regress ensues. For, on the view under consideration, the experience is conscious only if there is a further experience of it. And now that further experience must be conscious, on pain of falling back onto the first horn of the dilemma. And so there must be yet another, even higher-order experience. And so on. So, on this horn of the dilemma, any subject who enjoys a conscious episode simultaneously enjoys an infinity of increasingly higher-order conscious experiences. That is absurd.

The Cartesian might reply that there is an independent account of what it is for an experience to be conscious, one that does not appeal to further experiences. So the regress does not get started. But then it seems that this independent account should apply directly to the episode whose consciousness is being explained in the first place. If we have a handle on what it is for a state or episode to be conscious, independent of the occurrence of further experiences, then there is no reason to embrace the Cartesian account of consciousness.

The second problem for the Cartesian view of consciousness is that, like the view of self-

---

aware *of* an experience.

20 This claim is also defended by Lycan (1996). On his version, the experience of a state (say) is constitutive of the state's being conscious, but not (as on the Cartesian version) of the state's existence. This difference is not relevant to anything I have to say in this section.

21 It might be claimed that the notion of a non-conscious experience makes no sense. If that's correct, all the better for my argument: it simply closes off this horn of the dilemma.

knowledge that goes with it, it misdescribes the structure of the first-person perspective. The Cartesian view explains state-consciousness in terms of a certain sort of transitive consciousness: a state or episode is conscious when its subject is conscious *of* it. But if state-consciousness is constitutively tied to a certain sort of transitive consciousness, then this is the wrong sort. What you are conscious *of*, in having a conscious experience or thought, is the intentional object of the experience or thought, not the experience or thought itself. And consciousness of an intentional object is not constitutively explained in terms of consciousness of one's intentional state or episode. What's more, a subject's perspective on her own conscious states and episodes is deeply different from her perspective on the intentional objects of those thoughts and episodes. A conscious experience or thought with an intentional object in a sense *constitutes* the subject's perspective on that object. To take an analogous perspective on the experience or thought itself would be to have a further and quite different experience.

I have argued that it is a mistake to conceive of state-consciousness in terms of consciousness of states, as the Cartesian does. But if conscious states and episodes are not those that subjects are conscious *of*, how can subjects be in a position to know about them? Our conclusion should be that the mere occurrence in consciousness of a state or episode puts its subject in a position to self-ascribe it, without any need for a mediating experience. A conscious state or episode can be the basis for its own self-ascription, simply in virtue of being conscious.[22] Self-ascriptions are based on experiences, in a certain sense, but not experiences *of* mental states and episodes; rather, those states and episodes function as epistemic bases because they are themselves experiences, in the Erlebnis sense of being modifications of consciousness.

How can *this* be? That is the question for the rest of the thesis.

3.4 Conclusion

I have argued that we should reject introspectionism as an account of self-knowledge, and

22 Peacocke makes this claim, for certain types of conscious mental episodes, including perceptual experiences (Peacocke, 2005), but rejects it for those mental episodes that are actions, such as judgments (Peacocke, forthcoming). He offers a different account of what it is for a mental action to be conscious from his account of what it is for an experience to be conscious (the account for mental actions would fall under what I have called 'introspectionism'). I differ with Peacocke in two respects. I think that there is something in common to all conscious episodes and states, including mental acts of judgment, in virtue of which they are conscious, and in virtue of which they can function as bases for self-ascriptions. And I offer a different account of why a conscious episode can be a rational basis for a self-ascription (see Chapter 5).

the associated picture of the relation between self-knowledge and consciousness. I identified two different conceptions of introspection: the Lockean conception, on which it is metaphysically and epistemically analogous to perception, and the Cartesian conception, on which it is epistemically but not metaphysically analogous. I claimed that introspectionism of either sort falls foul of three objections. It fails to capture the authority of self-knowledge. It is at odds with the transparency, in many cases, of the self-ascribing procedure. It construes the first-personal character of self-knowledge too superficially, with the result that it makes no contribution to an explanation of the rationally committing aspect of certain self-ascriptions, when it ought to do so. In the last section I argued that the failure of introspectionism suggests a rather different picture of the relation between self-knowledge and consciousness.

It is no surprise that introspectionism fails as an account of self-knowledge. But what do we learn from its failure?

Firstly, this chapter has yielded two more distinctive features of self-knowledge, to add to our list from Chapter 1. There is, firstly, what I have called the *transparency of the self-ascribing procedure*. Secondly, there is the *rational commitment* of certain self-ascriptions. An adequate account of self-knowledge must aim to respect these features.

Another lesson we can learn from the rationally committing aspect of certain self-ascriptions is that the security of self-knowledge is deeply rooted in our nature as subjects. Security, as I characterised it in Chapter 1, is a modal feature of particular self-ascriptive judgments. We can now see that it is necessary that the self-ascriptions of subjects like us will be, *in general*, correct. It is not only that particular errors are, as Burge (1996) puts it, never 'brute', but always the result of a cognitive or rational malfunction. It is that systematic or widespread errors, on the part of some subject, would undermine our sense of that subject as a unified subject with determinate commitments, because willingness and unwillingness to make certain self-ascriptions itself contributes to the rational commitments of a subject.

Relatedly, first-person privilege is an aspect of the deep asymmetries between first-person and third-person perspectives on conscious states and episodes. Your conscious states and episodes are not like objects which, due to your privileged vantage-point, only you can see. They are not, to you, objects among others. They constitute your perspective on the world.

I have suggested that conscious states and episodes can function as bases for their own self-ascription, without any need for experiences *of* them. In the next chapter I will pave the way for a defence of this view, by presenting an epistemological framework in which we can

make sense of it.

CHAPTER 4

AN EPISTEMOLOGICAL FRAMEWORK

I have argued that we need an epistemic but non-introspectionist account of self-knowledge. I suggested in the last chapter that finding such an account will involve making sense of the thought that the conscious occurrence of an episode or state can rationally warrant the self-ascription of that episode or state, without that warrant requiring a further experience *of* the episode or state. The claim I will eventually make is that: when a subject enjoys a conscious state or episode of type M with content C, she thereby has a *reason* to judge that she Ms that C; and she can rationally exploit that reason to achieve self-knowledge.

I do not claim that rational warrants for judgments *always* involve reasons; perhaps you can make a rationally warranted judgment without judging for any reason. My claim is merely that knowledgeable self-ascriptive judgments, in particular, are made for reasons. This claim is stronger than the mere claim that they are rational. Nevertheless, I will offer a plausible reasons-based account, which meets the *desiderata* for a solution to the problem of self-knowledge better than any of its less ambitious rivals. The account will be consistent both with epistemological views according to which epistemic warrant always involves reasons, and with views according to which only some warrants involve reasons.

In this chapter I want to present an epistemological framework, within which I can subsequently go on to present and defend my view. In particular, I want to offer a partial account of what it is for a subject to have a reason to judge, and to judge for that reason.

In section 4.1 I will distinguish between there *being* a reason for you to Φ, your *having* a reason to Φ, and your Φing *for* a reason. I will argue that a non-Humean understanding of these notions is preferable to a Humean one. In section 4.2 I will argue that there *is* a reason for you to make a judgment with a given content when some consideration makes that content likely to be true. Section 4.3 will deal with the notion of *having* a reason to judge. You have such a reason, I will argue, when a reason-giving consideration is accessible from your point of view in such a way that, from your point of view, the content thereby seems, or could on reflection come to seem, likely to be true (relative to that consideration, at least) (4.3.1). I will argue that we should reject the traditional epistemologically internalist account of what this involves (4.3.2), and offer an alternative, hybrid account (4.3.3). Section 4.4 will deal with judging *for* a reason. You judge for a reason, I will claim, if your having that reason (or apparent reason) plays the right role in your coming to judge. This involves the

reason-giving consideration's showing up in the experience or thought by which you come to judge.

4.1 Reasons

Self-ascriptive judgments are knowledge,[1] I will claim, because they are made for good reasons. An act's being rationalised by a reason requires more than that there merely exist a reason for it.[2] We can distinguish between the following sorts of statements about reasons:

- R is a reason for you to Φ.
- You have the reason R to Φ
- You Φ for the reason R.

These statements can occur in both theoretical (epistemic) and practical contexts. In theoretical contexts Φ is judgment or belief; in practical contexts it can be any type of act.

We can gloss (a)-(c) with reference to the distinction between 'normative' (or 'good') and 'motivating' reasons. I will say much more to elucidate (a)-(c) as the chapter progresses, but for now a rough sketch will suffice.

Statement (a) says that some consideration R in fact counts in favour of your performing an action or making a judgment of type Φ.[3] R has some positive normative force with respect to the prospect of your Φing. That force is typically independent of what you believe, or might come to believe, about R and Φing. I will call such a reason a 'good reason', to indicate its normative force (as it is sometimes put, it is a 'normative reason'). A good reason may not be

1 I will talk about the judgments that constitute or manifest knowledge, rather than the beliefs that those judgments manifest. See section 1.4.
When I talk about self-ascriptive judgments, I am referring to those that *are* knowledge and that fall within the range I demarcated in Chapter 1 (section 1.1).

2 I assume that judging is an act. I will discuss the nature of judgment, and defend this assumption, in Chapter 6.

3 Dancy says: "When we think of such reasons, we think of features that speak in favour of the action (or against it)." (Dancy, 2002, p. 1). Similarly, Scanlon says: "Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favour of it." (Scanlon, 1998, p. 1.)

a sufficient reason, or even a remotely powerful one. To say it is a good reason is to say it genuinely favours Φing; but it may not favour Φing very strongly, and it may be outweighed by countervailing considerations.[4] Something that is not a good reason, in this sense, is not a genuine reason at all. It has no normative force.

Statement (c) says something about the *explanation* of why you perform a particular token-act of Φing. It says that R is the reason for which you Φ. This does not mean that R is a good reason for Φing. You can Φ for a reason that does not favour Φing at all—that is not a reason for you to Φ—even though it appears to you to favour it. In Dancy's terminology, (c) is a *motivating reason* statement.[5]

The notion of a motivating *reason* is narrower than the ordinary notion of a *motivation*. Intuitively, we can be motivated to act by things that are not reasons for which we act. Motivating factors feature in psychological explanations of action, but not necessarily in reason-attributing explanations. Your motivations for going to a party might include a subconscious yearning for companionship, but the reasons for which you go to the party might be quite different—that so-and-so will be there, that you have some obligation to go, or whatever. This is partly because doing something for a reason involves recognising or responding to the putative normative force of the reason, whereas that is not true of motivations. Your reasons for going to the party are considerations that you see as favouring going to the party; and your seeing them as favouring it explains your going. Generally, when you Φ for the reason R, part of the explanation of your Φing is that, rightly or wrongly, you in some sense see R as favouring that course of action (I will say more later about what this might involve). Mere motivations need not enter into your thinking in this way.

Statement (b) connects statements (a) and (c) by saying that the reason identified in (a) is potentially a reason for which you Φ (or could Φ), in the sense that it is available to play a

4 Some philosophers distinguish between something's being a reason (*simpliciter*) to do something, and its being a good reason to do that thing (e.g. Stampe, 1987). I reject that distinction. The notion of something's being a reason to do something is already normative: it attributes some value to doing that thing. And that is what is signified by 'good'. On the other hand, if those philosophers mean, by 'good', that a reason is sufficient, or powerful, or undefeated, then I have no quarrel with them—as long as they acknowledge that something can normatively favour Φing without being a sufficient or undefeated reason to Φ, and as long as they acknowledge the distinction (see below) between something's being a reason for which you Φ, and its having genuine normative force that favours your Φing.

5 The terminology is not very well suited to the case of reasons for judgment; the ordinary notion of motivation does not sit well with the non-voluntary nature of judgment. But the distinction marked by the term still applies. We can talk about the reason for which you judge, without committing to the claim that that reason is a normative reason—a good reason—for you to make that judgment.

role in a reason-attributing explanation of your behaviour.[6] That is, (b) says that R, a reason for you to Φ, stands to you in whatever relation is necessary for it to be potentially a reason for which you Φ—for it to be such that you could Φ for that reason (provided you *could* Φ).[7]

Some philosophers treat statements like (a) and (b) interchangeably: they make no distinction between there *being* a reason for you to do something, and your *having* that reason. Although our ordinary talk is sometimes consistent with this treatment, I think it obscures an important distinction. There are lots of considerations that favour your acting in various ways, but many of those considerations are wholly inaccessible to you. As such, you *could not* act on them, and they are irrelevant to what you rationally ought to do, what you can be blamed for not doing, and so on. Only reasons that are in some sense accessible to you—that you *have*—can rationalise your actions, can justify praise or blame, and so on. In so far as reasons talk is intimately connected to talk of rationality, justification, blame, etc., it is important to distinguish reasons *simpliciter* from reasons that you have.

Thus, a reason you have is a reason you are in a position to act on.[8] This does not mean that you always recognise the normative force of reasons you have. A consideration could be available for you to act on, but you might fail to act on it, either because you fail to notice it or because you fail to recognise its force. For example, if you are standing on a railway track and can plainly see a train approaching, you have a reason to get off the track—even if you are so out-of-sorts that you fail to recognise the force of that reason, or you refuse to believe that the train is there. Thus, having a reason is *not* the same as having a belief that you have a reason, nor is it the same as having a belief whose content constitutes a reason.

In sum: (a) says that R is a good (normative) reason for you to Φ; (b) says that a good reason for you to Φ, R, is available to feature in a reason-attributing explanation of your action; and (c) says that a consideration R features in such an explanation of an act of Φing, and is thus a motivating reason for your Φing, regardless of whether it is a good reason.

---

6 A complication: you can have a reason and yet be rationally barred from exploiting it. For example, a sceptic may have reasons for lots of beliefs, and yet be unable to exploit those reasons because of his (mistaken) sceptical beliefs about reasons for belief. A stricter gloss on statement (b) would include a *ceteris paribus* clause to take these cases into account. I do not think the point affects the fuller account of having a reason I will offer in section 4.3.

7 I assume here that the motivating reason for which you Φ can be identical to the normative reason for you to Φ. I assume, that is, that the reasons for which we act are sometimes the reasons that there are for us to act—the reasons that favour our so acting. Smith (1987) appears to doubt this. Smith's view entails that the rationalisation of actions is quite independent of the *counting-in-favour-of* relation characteristic of normative reasons. But surely when your Φing is rationalised by a reason, that reason must be both a reason that favours Φing *and* the reason for which you Φ. See Dancy (2002) for more on this.

8 Provided you are not barred from doing so. See note 6.

Statement (b) entails (a): if you *have* a reason to Φ, then that reason *is* a reason for you to Φ (though perhaps not a sufficient one). However, (c) does not entail (a), because you can Φ for a reason that is not a good reason for you to Φ. When you Φ for a good reason, your Φing is explained in a certain way by something that counts in favour of it; when you Φ for a bad reason, there is a similar form of explanation, but what plays the explanatory role of the reason does not count in favour of your Φing. Since (c) does not entail (a), and (b) entails (a), (c) does not entail (b): you can Φ for a reason that does not in fact favour Φing, and *a fortiori* is not a reason that you have to Φ. The *conjunction* of (c) and (a), however, does entail (b). A good reason for which you act or judge must be a reason you have, for otherwise the reason could not explain your action or judgment. Your having the reason just is its being potentially a reason for which you act or judge.

I have claimed that (good) reasons are considerations that count in favour of actions, regardless of your beliefs about them; and that having a reason is a matter of having access, of some sort, to a consideration that is in fact a reason. These considerations are facts, or states of affairs, not your beliefs about them Although I have arrived at these claims merely by spelling out an intuitive understanding of various types of statement about reasons, they are incompatible with one theory of reasons—what I will call the Humean theory.[9] According to the Humean theory, (good) reasons have their sources in psychological states, rather than, as I have claimed, in the considerations to which our psychological states sometimes give us access.[10] Thus, on the Humean view, merely being in certain psychological states will suffice for having a reason, whereas on my non-Humean view that will not be so.[11] For example, on (a version of) the Humean view, the reason for you to Φ is (say) that you desire that *p*, and you believe that by Φing you will bring it about that *p*. You *have* that reason as long as you have the relevant belief and desire—regardless of what the

9 I do not claim Hume was a Humean theorist of reasons, in this sense. Note also that this is not the same as the Humean theory of motivation, as defended by Smith (1987). The Humean theory of motivation says that motivating reasons are psychological states, not that normative reasons are psychological states. If this theory is combined with the claim (eschewed by Smith) that motivating reasons are in some cases the very same things as normative reasons (see note 5), then the 'Humean view' of (normative) reasons results.
What I am calling the Humean view of reasons is called 'the attitudinal view' by Broome (forthcoming). Wedgwood (2002) seems to endorse it, or something close.

10 The Humean shares my assumption that normative reasons and motivating reasons are the same sorts of things. He holds that they are psychological states, whereas I hold that they are considerations (facts, states of affairs) to which those states sometimes give us access. Both of these views are to be distinguished from the view that motivating reasons are psychological states and normative reasons are facts or states of affairs.

11 Since the Humean view, in this sense, is a *sufficiency* claim, not merely a necessity claim, some views of reasons that are usually seen as Humean in spirit, such as that of Bernard Williams (1981), may count as non-Humean in my sense.

facts are with respect to Φing and *p*.

The Humean claims, contrary to what I have said, that if you Φ for a reason (and you are responding appropriately to that reason) then it *is* a reason for you to Φ, and it is a reason you *have*. To say that you Φ for the reason R, on his view, is to say that certain psychological states—a belief-desire pair, say—play a certain role in your coming to Φ. Your Φing for a reason ensures that you are in the psychological states that constitute your reason—otherwise they couldn't play the role of the reason for which you Φ. The existence of a (good) reason consists in your being in the appropriate psychological states. So if you Φ for the reason R, and you are responding appropriately to R, then there is a reason for you to Φ, namely R. What's more, it's a reason you *have*—for psychological states are always available to play a role in explaining action. On my view, by contrast, there can be some putative consideration that is your reason for Φing, and to which you respond appropriately, but which is not in fact a reason for you to Φ—because, for example, that consideration is a falsehood that you mistakenly believe.

Why should we prefer the non-Humean account I have offered over the Humean view? There is a straightforward and very powerful objection to the Humean view. I will state that objection, before refuting the main objection that is given against the non-Humean view.

The objection to the Humean view is that, if it were correct, you could conjure up a reason for yourself to perform some action that in fact has nothing whatsoever to be said for it, and that consequence is highly implausible. For example, suppose John desires to get rich. Suppose he adopts, for no good reason and against all the evidence, the belief that by dropping his trousers he will get rich. On the Humean view, there is now a reason for John to drop his trousers, namely his belief-desire pair. But in fact there is *no* reason for him to drop his trousers. To say that there is a reason for John to drop his trousers is to say that something counts in favour, normatively, of his doing so. In the scenario given, nothing counts in favour of his doing so; there is nothing good, even defeasibly, about his doing so. He merely believes that there is.[12]

Another type of example comes from Broome (2007). If beliefs are sources of reasons, then presumably your existing beliefs give you reasons to believe what logically follows from their contents. Suppose Helen believes the following conditional: if the earth is flat, then the

12 Smith (1987) would say that John *has* a *motivating* reason to drop his trousers, even if he has no *normative* reason to do so, and even if there *is* no normative reason for him to do so. I think that what John has is not a reason to drop his trousers, even if it could explain his dropping his trousers. For what John has cannot rationalise his dropping his trousers.

earth is flat. Presumably, that belief is one for which she has excellent reasons. Suppose, now, that Helen adopts the belief that the earth is flat for no good reason and against all the evidence. From the contents of these two beliefs, it follows that the earth is flat. On the Humean view, then, there is now a reason for Helen to believe that the earth is flat.[13] She bootstrapped into existence a reason to believe that the earth is flat, by adopting that very belief. That is absurd.

This objection turns, in part, on the point that you can act or believe for bad reasons, even though what you do is appropriate relative to certain of your psychological states. But there is an objection to the *non*-Humean view that appeals to a similar point, which I will now consider.

Actions performed on the basis of mistaken beliefs need not be irrational. The Humean claims that only his theory can account for this fact. Suppose you Φ because you mistakenly believe that Φing would serve some goal of yours. According to the Humean, your belief, though mistaken, constitutes a reason for you to Φ; thus, if you Φ for that reason, we can explain your action by reference to a reason you have. On the non-Humean view, however, your belief is not a reason, and nor does it constitute access to one—for Φing would *not* serve your goal, so there *is* no consideration favouring Φing, to which your belief constitutes access. But now it looks as though, when you Φ on the basis of the belief, you are doing something for which there is *no* reason, or at least no reason connected to that belief. So it looks as though your action can't be explained by reference to a reason. But then it is irrational. This, says the Humean, is an implausible consequence of the non-Humean view.

The Humean's objection assumes that acting for no genuine reason is sufficient for acting irrationally. This assumption can be rejected. There may be ways of acting rationally that do not depend on having a reason. It may be that there is a normative requirement on you to satisfy the conditional, 'If you believe Φing would serve some goal (and that it would be the best way to do so, etc.), then you Φ', even though believing that Φing would serve some goal is not sufficient for having a *reason* to Φ (see Broome, 1999).

The Humean may press the objection, however. He may say: even if the non-Humean view does not entail that actions performed on the basis of mistaken beliefs are *irrational*, it *does* entail that they cannot be explained by reference to good *reasons*—and this is implausible.

But the Humean is wrong even to claim that actions performed on the basis of mistaken

13 I assume here that having reasons to believe is closed under entailment. We could add to the example that Helen *knows* the entailment, and accordingly make the closure assumption weaker. Note that these reasons can be defeasible.

beliefs are always, on the non-Humean view, actions that cannot be explained by good reasons. If you go to Dave's house on the basis of a mistaken belief that the party is happening there, that belief does not constitute access to a consideration that counts in favour of going to Dave's house—for the party is *not* happening there. But unless that belief is itself arbitrary, your action can be explained in terms of good reasons. Suppose Chris told you that the party was happening at Dave's house. In that case, that Chris said so gave you a good reason to go to Dave's house, since going to where you are told the party is happening is generally a good way of going to parties: your going to Dave's house is connected to your goal of going to the party by the fact that Chris said the party was happening there. (Chris's saying so also gives you a reason to *believe* that the party is happening at Dave's house. But it doesn't follow that the belief is itself a reason to go, nor that it is the reason for which you go.) So, in this case, there is a good reason for you to go to Dave's house, a reason that contributes to the explanation of your going there. Unfortunately, the reason is defeated by other, countervailing reasons (that Chris was lying) of which you are unaware. Here is a case, then, in which you act unsuccessfully because of a mistaken belief, but the action can be explained by a reason, without supposing that the belief itself is the reason for the action.

Of course, on my account, if your belief that the party is happening at Dave's house is one for which there is no good reason, then your action of going to Dave's house is one for which there is no good reason either. But that seems correct: we can't give ourselves reasons for actions by adopting arbitrary beliefs.

I conclude that the objection to the non-Humean view fails, while the objection to the Humean view succeeds. I will therefore proceed on the basis of the non-Humean view I began outlining above.

In the rest of the chapter I will consider in more detail the three forms of reason-statement, (a)-(c); and in particular what is involved in their being true when 'Φ' refers to judgment. I will be constructing a framework in which, in subsequent chapters, I can present and defend my account of self-knowledge.

4.2 Reasons to judge

First, I consider the notion of some consideration *being* a reason for you to Φ.

We saw that something is a reason for you to Φ when it counts in favour of your Φing. If something counts in favour of your Φing, then there is at least some respect in which Φing would be a good thing for you to do. Either there is some intrinsic worth to Φing, or Φing is

connected to some further right, good or goal. Something that counts in favour of Φing is something that connects it to a right, good or goal.

So what is it for something to be a reason for you to *judge* a particular content?

In judging, it is good (or right) to judge truly, and bad (or wrong) to judge falsely. Indeed, it is a constitutive fact about judgment that in doing so you aim, in some sense, to judge what is true. Therefore, reasons to judge—epistemic reasons, at least[14]—are considerations that count in favour of judgments by connecting them to truth. A reason to judge that *p* is a consideration that in some sense contributes to the likelihood that *p* is true, and thus that judging accordingly will serve the goal of truth.

A reason to judge is always something that supports the truth of the content of the judgment for which it is a reason. Support for the truth of a content can always be captured or displayed in an *argument*. What does the supporting can be laid out in the premises of an argument; the relation of support is captured (or constituted) by the validity of the argument; the content supported is the conclusion. 'Valid argument' here is to be understood in a broad sense, as including abductive and inductive support for contents, not just deductive validity. If the considerations that putatively support a content cannot be connected to the content in a deductively, inductively or abductively valid argument, they do not support it at all. The epistemic reason-giving relation, on any conception, is a truth-conducive one, and truth-conducive relations are precisely what is captured in valid arguments.[15]

All of this has to do with what it is for something to *be* a reason for judgment. I take it that it is consistent with many different conceptions of what it is to *have* a reason, and what it is to judge *for* a reason. I also take it to be consistent with many different conceptions of warrant, and of the warrant that attaches to judgments made for good reasons. Making a judgment for a good reason may be a way of fulfilling your epistemic duty, an exercise of an intellectual virtue, or whatever.

Next I turn to the notion of having a reason.

4.3 Having a reason to judge

In this section I will offer my account of what it is to have a reason to judge. I will begin

14 When I talk of 'reasons for judgment' I will be referring to epistemic reasons throughout.

15 The intimate relation between reasons and arguments is emphasised by Brewer (1999, p. 151). Brewer puts this point to use in the service of an internalist theory of the sort I will reject in section 4.3.2.

with some consideration of the general notion of having a reason. The question will then arise of how exactly that notion applies in the case of judgment. I will argue against an internalist account, and offer my alternative hybrid view.

4.3.1 Having a reason and points of view

What needs to be added to (a) ('R is a reason for you to Φ') in order to get (b) ('You have reason R to Φ')? That is, what relation must you stand in, to a reason, in order for it to be a reason you have? I have talked of accessibility. But what does this amount to?

A reason that you have is a reason that is available to play a role in a reason-attributing explanation of your action, judgment or belief. So we can begin to address our question by asking: what sort of explanation do we offer when we offer a reason-attributing explanation?

In offering a reason-attributing explanation of an action, we make the action *intelligible*; we *make sense* of it. But we don't simply make sense of it to *ourselves*—we could do that simply by understanding its causes or seeing that it in fact served some goal. We show that it makes sense *to the agent*. Doing this involves adopting the agent's *point of view* in a certain way. We show that Φing was, or seemed, a good thing to do from the agent's point of view. A reason explains an action *via* its making that action in some sense seem appropriate from the agent's point of view.

It seems, then, that R is a reason you *have* to Φ when it contributes to your point of view in such a way as potentially to make Φing seem appropriate from your point of view.

An action seems appropriate when it can be seen as serving a goal or good. That is how actions or judgments inherit normativity (or apparent normativity) from goals or goods. Thus, Φing could seem appropriate from your point of view when you could see it as serving a goal or good.

There are two things to say about the second 'could' in that last sentence. First, it means 'could on reflection alone': Φing will not count as appropriate from your point of view if you can see it as serving a goal only by ascertaining new facts that do not follow from those already known, or by adopting new non-instrumental goals. Second, it is not restricted to subjects who have the conceptual sophistication to think of goals as such, and to think reflectively of actions and judgments as serving those goals. Φing will count as appropriate from your point of view if you have a goal G that you could be aiming at (not merely furthering) in Φing; it is not necessary that you have higher-order theoretical knowledge of

that fact.

For example, consider a small child who has goals, but is not yet capable of thinking reflectively about goals as such. Suppose the child has the immediate goal of getting a particular toy, and she sees the toy as being in a demonstratively picked out location, *over there*. That the toy is over there is thereby a reason that the child *has* to go over there (wherever it is); she can go over there *for* the reason that the toy is over there. Going over there is appropriate from the child's point of view; it makes sense to her to do so, because she can intentionally pursue her goal by doing so. This doesn't require the child to have higher-order knowledge that she has a certain goal, and that going over there would serve that goal.

We saw earlier that a (good) reason for Φing is a consideration—a feature of how the world is—in virtue of which Φing in fact serves some goal or good. What turns such a consideration into a reason you *have* is that it is a feature of how the world is *for you*. Only if a consideration is a feature of the world as it is for you, can it be recruited as a reason by you.[16]

Having a point of view, then, involves representing the world, or having it present itself, as being a certain way. If a consideration contributes to how the world is from your point of view, this can only be because it is a feature of the world, as the world is represented in one of your representational states. That is, a consideration constitutes a reason you have only if it is represented by a representational state of which you are or have been the subject, or it is somehow implicit in, or a consequence of, what is represented by one or more such states. And that is because how the world is for you cannot outrun how it is represented to you (how you represent it) as being. What else could it be for a consideration to contribute to how the world is for you? What a *sub*personal state represents, if anything, cannot as such be anything to *you*, since by definition the person is not the subject of such a state. *A fortiori*, a representational state that is not yours at all could not contribute to how the world is for you. For, again, what difference could such a state make to you, other than by making a difference to some state that *is* yours? And of course the same is true of anything that is not a representational state.

The claim in this section has *not* been that only features of the world as it is for you can contribute to *what* you have a reason to *do*. It has been, rather, that only features of the world

16 A consideration might be a reason for you to Φ, not in itself, but only given certain other considerations. In that case, you must have the relevant access to all of those considerations in order to have that reason to Φ. For otherwise you could not see Φing as appropriate. For simplicity, I will talk as though a particular consideration is in itself a reason for you to Φ.

as it is for you can be reasons you *have*. What you have reasons to do depends on more than the identity of the reasons you have. Other factors, such as your goals, make a difference to what the various reason-giving considerations give you reasons to do.[17] Your goal is not a feature of the world as it is for you; nor need the fact that you have some goal be a feature of the world as it is for you, in the relevant sense. As we saw, you need not have higher-order knowledge about your goals. Your goals make a difference to what you have reasons to do, but they are not themselves reasons you have.[18]

Thus, you have a reason to Φ when some consideration R, accessible to you in the sense I have set out, counts in favour of your Φing. The reason you have is that consideration R. Other factors, besides the accessibility of R, are necessary for your having that reason to Φ. Those factors need not be reasons you have.

It might be objected to the account I have given that it makes ordinary subjects too authoritative about whether they have a reason. After all, ordinary subjects typically know, by and large, what is the case from their point of view; and yet subjects often have reasons that they fail to recognise. But recall that we distinguished between having a reason and recognising a reason's force (section 4.1). It is one thing for the world, as it is for you, to provide you with a reason to Φ, and quite another thing for you to recognise that it does so. You may be aware of a consideration but fail to appreciate its force. Even a consideration that you fail to believe in—a fact that you do not believe to obtain, say—can be a reason you have to Φ, if you have good reasons to believe in that consideration. If you could come to know that *p* by deduction from contents you know, and *p* counts in favour of your Φing, then you have a reason to Φ, even if you fail to make the deduction; in that scenario, nothing further is required, except some clear thinking, in order for you to come to see Φing as appropriate. (The consideration would thus count as a feature of how the world is for you, even though you hadn't realised it.) Having a reason to Φ does not entail that you actually see Φing as good in any respect; it entails that you *could* come to see it as such by reflection alone.

---

17 I will argue in Chapter 6 that the *way* in which the world is given to you as being some way can make a difference to what you have reason to do. In such a case, it is the way the world is for you that is the reason you have; but the way it is given to you as such makes a difference to what that reason rationalises for you.

18 *That* you have a certain goal G may sometimes be a reason you have. But if you act *for* that reason, G does not directly guide your action; rather, the fact that you have G guides your action, via some state of yours, such as the belief that you have G. A belief of this sort won't alone drive you to action; there must be some other goal, G', that makes the content of the belief a reason and that drives you to action. Generally, goals guide action directly, not via beliefs about goals (or some other sort of epistemic access to those goals). Thus, goals make a difference to what you have reason to do, without themselves being reasons you have.

So much for the general notion of having a reason to Φ. What of the case where Φ is judging?

The goal of judging is to judge truly. The act of judging some content will be appropriate from your point of view if you see it as serving that goal. Thus, it will be appropriate from your point of view if that content strikes you as true, or likely to be true—for, in that case, you can be aiming at truth in judging that content. You have a reason to judge a content when some consideration makes that content seem true to you, or would do so on competent reflection.

The next two subsections will ask what this involves.

4.3.2 The internalist conception of having a reason to judge

The question under discussion is: what is it for you to have a reason to judge a given content? We have seen that it will involve there being some consideration, which is a feature of how the world is represented as being by you, in virtue of which you could come to see that content as true, or likely to be true. There are several ways in which this rough sketch could be filled out. In this section I want to argue against one way of filling it out. It is an account associated with traditional internalist epistemology.

Traditional epistemological internalism is an account of warrant rather than of reasons *per se*, but it gives reasons an essential role in that account. Importantly, however, internalism relies on a particular and non-obligatory conception of what it is to have a reason.[19]

Internalism is the doctrine that a subject's being warranted in making a judgment supervenes on facts that are epistemically accessible[20] to the subject. We are interested particularly in the case where a subject is warranted in virtue of having a *reason*. Recall that the truth-conducive relation between a reason and the content of the judgment for which it is a reason can always be articulated in a valid argument, deductive, inductive or abductive (section 4.2). The premises of this argument form part of the base on which the warrant provided by the reason supervenes—for it is the support provided by these premises that constitutes the

19 By 'internalism' I mean epistemological internalism (in particular, access internalism), *not* the view of reasons known as 'reasons internalism'.

20 Internalism is sometimes characterised in such a way that this access must be through certain privileged channels, and in particular that it must be the sort of access one has to one's own mental states (Pryor, 2001). Restricted in this way, internalism would naturally go with a Humean view of reasons, according to which one's reasons are psychological states to which one has privileged access. However, there are epistemological approaches that share the features of internalism that I am interested in, but do not endorse Humeanism about reasons (e.g. Brewer, 1999).

truth-conduciveness of the relation between the warrant and the judgment, and this truth-conduciveness is part of what makes the reason a reason. Thus, the internalist doctrine, applied to reasons, is that a subject who has a reason must have epistemic access to a deductively, inductively or abductively valid argument, in which the reason features in the premises and the content of the judgment for which she has a reason constitutes the conclusion.[21]

Reasons, on this conception, are propositions; they transmit warrant to other propositions by links of valid inference. A subject acquires warrant to judge that *p* when she acquires access to the propositions that constitute inferential support for *p*.

On this view, the reasoning that displays the rationality of a judgment made on a particular basis is something to which the subject has access, not something that the subject merely conforms to. This is in contrast to externalist views of warrant, according to which a judgment is warranted when it meets some truth-conducive condition that may be quite inaccessible to the subject. On the externalist conception, warrant attaches in the first instance to judgments, when those judgments are made in appropriate ways, rather than to propositions.

The internalist view can be more or less demanding in its conditions on grasping an argument. According to Brewer (1999), the premises of the argument must be explicitly entertained by the subject as the contents of mental states. On a less demanding version of internalism, such as espoused by Bonjour (1978) and Lehrer (1974), the argument must be available to the subject, so that she could put it forward in defence of her judgment or belief. The minimal commitment is that the premises of the argument be access*ible* to the subject's reasoning, even if not accessed.

I now want to argue that internalism is incompatible with my claim that enjoying a conscious episode can suffice for having a non-introspective reason to self-ascribe that episode, and in fact would go naturally with an introspectionist account. I take this as a motivation to seek an alternative conception of warrant-by-reasons, which I will do in the next subsection.[22]

---

21 See Brewer (1999, Chapter 5).

22 If internalism is incompatible with my claim, why should we reject internalism rather than rejecting my (as yet undefended) claim? For one thing, I have already given independent reasons for rejecting the introspectionist account that goes with internalism. Secondly, the incompatibility would put pressure on us to reject my claim only if it were plausible that internalism provides a generally correct conception of the warrant provided by reasons. But see Martin (2001) for a powerful argument against a sophisticated internalist view regarding perception, that of Bill Brewer. Thirdly, there is, as I will show, a more plausible alternative conception of warrant by reasons, which is compatible with my claim.

On the internalist conception, having a reason to judge a self-ascriptive content, such as "I M that *p*", requires having available a valid argument whose conclusion is the content "I M that *p*". But that makes it hard to see how ordinary first-order experiences and thoughts could give reasons for self-ascriptions. Such experiences and thoughts do not typically have contents that would inferentially justify self-ascriptive contents. To judge that *p*, say[23], is not to have available an argument to the content "I judge that *p*", or even "I am thinking that *p*" (where 'thinking that' is neutral as to what type of mental episode is occurring), since there is no valid inference of any kind from *p* to either of the latter contents. If having a reason is grasping an argument, then most first-order experiences and thoughts do not give their subjects reasons for self-ascriptions.

One option for the internalist is to hold that there *is* some proposition available to thinkers that, combined with the contents of ordinary, world-directed experiences and thoughts, yields arguments to self-knowledge. But what might this proposition be? It would have to be some general principle such as "If *p* then I am thinking that *p*". The subject could then invoke the content *p* of her occurrent thought, to infer the conclusion "I am thinking that *p*". The problem with this suggestion is that there is no plausible candidate for such a general principle: it is false, and nobody has any reason to believe it. So it could hardly help to justify self-ascriptions, or anything else.

Alternatively, the internalist might claim that self-knowledge is somehow inherent in our ordinary, first-order experiences and thoughts themselves, and does not require an independent principle. The suggestion here is that all our conscious states and episodes already constitutively involve self-consciousness: when you have a thought that *p*, you *ipso facto* have the thought that you are thinking that *p*, or an experience as of your thinking that *p*. It is the content of this higher-order thought or experience that warrants a self-ascriptive judgment, on this view. This suggestion is a version of Cartesian introspectionism: it holds that any conscious episode constitutively involves a representation of the subject as enjoying that episode, which the subject can take at face value in order to judge that she is enjoying that episode. I have already rejected that view (Chapter 3).

Epistemic internalism, applied to self-knowledge, does indeed naturally suggest an introspectionist approach. If internalism is right, then a knowledgeable self-ascription must be based on a proposition, to which you have access, and from which the self-ascription can

23 The point is not restricted to propositional, and therefore conceptual, contents. Nor do I believe that only states and episodes with conceptual contents give reasons. The internalist, however, must hold that only conceptual contents can be reasons, since only conceptual contents can be used in arguments (see Brewer, 1999).

be inferred. As we have seen, this proposition cannot generally be the content of the state or episode you are self-ascribing. It is hard to see what it could be except a self-ascriptive proposition that seems to you to be true. And it is hard to see what this seeming could be, if it is not introspective (in the sense I employed in Chapter 3).

For the case of self-knowledge based on reasons, internalism is too demanding in the way it connects, on the one hand, the argument that articulates why a reason is a reason, and, on the other, what is accessible from the point of view of the subject who makes the judgment for that reason. (That is not to suggest that an internalist conception doesn't capture other types of warrant by reasons.) But we should not deny that there is some such a connection. When a subject has a reason for a self-ascriptive judgment there will exist such an argument, and having a reason is essentially connected to what is accessible from the subject's point of view. To deny that there is such a connection would be, effectively, to adopt a purely externalist conception of warrant, that gives no role at all to reasons or rationality. I have argued (section 1.4) that such a conception does not capture the warrant for self-knowledge.

What is required, then, is an account of warrant-by-reasons that respects the connection between reasons and the subject's point of view, without endorsing the internalist conception of what that connection is. This account will serve as the context in which I can put forward my explanation of the warrant for self-ascriptions.

4.3.3 A hybrid view

In this subsection I want to present a hybrid view of warrant, and demonstrate its advantages over the internalist view I discussed in the last subsection. It is an account of a type of epistemic warrant that can be enjoyed by subjects in virtue of having reasons for judgment. In giving a central role to reasons, it is non-externalist and aims to take on board what is correct about internalism; on the other hand, it rejects internalism's demanding conception of what it is to have a reason.

According to this hybrid view, warrant attaches to judgments, and the warrant for a judgment is determined by the way in which the subject comes to make the judgment (as suggested in section 2.2). Certain ways of making judgments are rational. A judgment made in one of those ways will be warranted. Support for propositions is nevertheless crucial for warrant, on this view. Not all truth-conducive ways of making judgments are rational; mere reliability is not sufficient for warrant. One rational, truth-conducive way of coming to make a judgment is by basing it on a good reason; and a good reason for a judgment with a particular content

is something that supports the truth of that content.

The hybrid view is governed by the constraint that, when you do something for a reason, what you do (or your doing it) makes sense, or is appropriate, from your point of view. When you come to make a judgment by basing it on a reason, you will be taking as true a content that you appreciate (or 'misappreciate') is true, or likely to be true, in virtue of whatever consideration constitutes your reason. You will be making some sort of connection between the reason and what is involved in the content of the judgment being true; you will appreciate that the truth-conditions of the content are likely to be met, and make the judgment because of that appreciation. Your judgment will not be a 'leap in the dark'. None of this need involve reasoning or reflection, on your part, that employs the concepts of truth, of a content, of truth-conditions, or of reasons. What I have described is simply the procedure of accepting what strikes you, in virtue of some consideration, as being the case, or what you ascertain to be the case. A simple example would be judging that *p* when it perceptually seems to you as though *p*; in this case, you accept a content because its truth-conditions seem to you to be met.

How does a view with the general features set out above meet this constraint, without reverting to internalism? I will now set out the hybrid view in more detail, to show how it does so.

I start from the following claim: a consideration *is* a good reason for you to judge a particular content if, in basing a judgment with that content on that consideration, you are likely to judge truly. This follows from the point that a reason for Φing is a consideration in virtue of which Φing is likely to serve some good or goal, and the point that the goal of judging is to judge truly (see above, section 4.2).

Thus, when a judgment is made for a good reason, there is a truth-conducive relation between the reason and the judgment-content. If the you base the judgment on the reason appropriately, you rationally exploit that truth-conducive connection, and your judgment is warranted. I claim, however, that you need not understand *why* that connection is truth-conducive. That is, you may lack epistemic access to the argument that would articulate the truth-conducive connection. Here, then, is the distinctive claim of the hybrid view: a judgment can be rational, and thus warranted, in virtue of being based on a good reason, even though the subject does not have access to the *reasoning* that would articulate why that reason is a good reason.

This raises two crucial questions. First, how can you exploit a truth-connection if you do not

understand the connection? Secondly, why should judgments made in the way described count as *rational*?

The answer to the first question begins with the point that you can be sensitive to the fact *that* there is a truth-connection between considerations of a certain type and contents of a certain related type, without having knowledge of *why* there is such a connection. Such sensitivity will be manifested in, for example, your willingness to judge contents of the second type in the presence of considerations of the first type. If your dispositions to judge follow closely the pattern of the truth-connection—if, for example, you are disposed to withhold judgment when that connection is undermined—then the judgments that manifest those dispositions must be attributed to sensitivity to that truth-connection, rather than to a simple tendency to respond blindly to considerations of the first type with the relevant type of judgment. For example, our colour judgments are not mere reactions to colour appearances. We take colour appearances to indicate colour properties, but we are also sensitive to defeaters for that connection. We refuse to make colour judgments when we know the lighting conditions to be abnormal. Our dispositions to make and withhold colour judgments constitute sensitivity to the connections between colour appearances, lighting conditions and the identity conditions for colours. Many subjects have these dispositions despite lacking any epistemic access to the theory that articulates those connections. When such subjects make colour judgments, they are exploiting a truth-connection between colour appearances and colour properties, without understanding that connection.

What of the second question—why do such judgments count as rational? Sensitivity to the truth-connection between a certain type of consideration and a certain type of content can be grounded in the conceptual and cognitive capacities that are constitutive of understanding contents of the type in question. Those capacities ground your willingness to judge a content of that type based on a consideration of the former type—they ground the sensitivity you manifest in so judging. I claim that when your coming to make a judgment is a manifestation of a sensitivity that is grounded in conceptual and cognitive capacities in this way, the judgment will be rational from your point of view.

Consider again our dispositions to make and withhold colour judgments. These are not brute dispositions, but are explained by what is involved in possessing the various colour concepts and having the perceptual capacities that underlie them. To possess a concept is to know its satisfaction-conditions; thus, possession of colour concepts involves, *inter alia*, having some sort of grasp of the distinction between the appearance of an object on a particular occasion and its stable colour properties. The grasp of this distinction will typically consist not in

theoretical knowledge, but in various abilities, such as the abilities to sort by colour in various conditions. These abilities, taken together, constitute a capacity. Willingness to make a simple colour judgment is grounded in that capacity. But that capacity just is knowledge of the condition for an object to fall under a particular colour-concept. In exercising the capacity, you will be reacting to the apparent fulfilment of that condition. You will not be making a leap in the dark, but doing what makes sense from your point of view, given your grasp of what it is for the colour judgment to be true.

That the content of the judgment, in such a case, seems true to you, is explained by your having the capacity in question. But having the capacity is not a matter of having full knowledge of the argument that articulates the truth-connection between the reason and the content.

An assumption I am making here is that what is rational for a subject can depend on, or be explained by, the capacities the subject has. For any given capacity, there are certain ways of exercising the capacity such that a disposition to exercise it in those ways is entailed by possessing the capacity. To exercise a capacity in such a way is not, typically, to do something blindly. When the capacity is a conceptual capacity, and the relevant ways of exercising it involves making certain judgments, those judgments will not be made blindly because judgments aim at truth and the capacities in question constitute grasp of truth-conditions. Suppose, for example, two subjects react to the same colour appearance by saying “That is red”. Suppose that one of the subjects is sensitive to defeaters for colour judgments, while the other simply reacts to colour appearances with affirmative judgments. My claim is that the utterance of “That is red” expresses a rational, warranted judgment on the part of the first subject, but not on the part of the second, precisely because the first grasps the truth-conditions of that content, while the second doesn't.[24]

Let me summarise the hybrid view I have presented. A judgment can be warranted by a reason when the subject's coming to make the judgment by basing it on the reason involves a sensitivity to a truth-connection between the reason and the judgment-content. The sensitivity need not involve an understanding *of* the truth-connection, in a sense that requires access to the argument that would articulate it. It can consist in dispositions to make and withhold judgments, as well as practical abilities. If these dispositions and abilities are grounded in possession of the concepts involved in the judgment-content, then, , when the

24 This also shows that the second subject doesn't genuinely possess the colour concept *red*, hence couldn't genuinely *judge* “This is red”. That is why possession of the concept is tied to a disposition to get applications of it right in a certain range of cases. See Peacocke (1992).

subject judges in that way, she does what is appropriate given her grasp of what it is for the content to be true.

Now we can formulate a sufficient condition for warrant as follows. A judgment that *p*, made by a subject S, is warranted if S comes to make the judgment based on some consideration R, and S's willingness to judge that *p* on the basis of R is a manifestation of sensitivity to the fact that the truth-conditions of the content '*p*' are likely to be met in the presence of R, a sensitivity grounded in the subject's conceptual and cognitive capacities.

That completes my presentation of the hybrid view. I want to note its affinities with two other non-internalist views of warrant, before saying briefly why I think we should prefer it over its internalist and externalist rivals.

In its emphasis on the subject's cognitive and conceptual capacities, the hybrid view presented here makes some gestures towards a virtue-epistemological conception of warrant. In particular, it has similarities with Greco's (2002) 'agent reliabilism'. On Greco's view, the warrant for a belief or judgment depends on its being produced by "one or more of [the subject's] cognitive abilities or powers" (*ibid.*, p. 21). Greco is also keen to accommodate internalist intuitions: he holds that a belief or judgment produced by certain cognitive dispositions will be "subjectively justified" for the subject, because her patterns of judgment will manifest "an awareness of sorts that some relevant evidence is a reliable indication of some relevant truth." (*ibid.*, p. 22.) In the foregoing few pages, I have endorsed and expanded on versions of both of these points.

However, I don't think the hybrid view presented here falls within virtue epistemology. Virtue epistemology attempts to understand the warrant for particular judgments in terms of normative epistemic properties of persons—namely, their intellectual virtues. On the view I have presented, the warrant for particular judgments is explained in part by properties of persons—their conceptual and cognitive capacities. But these capacities need not themselves be characterised in normative epistemic terms. Thus, the hybrid view is not committed to the 'direction of analysis' thesis distinctive of virtue epistemology (c.f. Greco, 2004).

The conception of warrant endorsed by the hybrid view shares a number of features with the notion of entitlement which has been developed in slightly different ways in the work of Burge and Peacocke. A judgment made in a particular way can be knowledge because you have an entitlement to judgments made in that way, even though you are not capable of understanding your entitlement:

> "entitlements are epistemic rights or warrants that need not be understood by or even accessible to the subject. We are entitled to rely, other things equal, on perception, memory, deductive and inductive reasoning, and on … the word of others. […] Philosophers may articulate these entitlements. But being entitled does not require being able to justify reliance on these resources, or even to conceive such a justification." (Burge 1993, pp. 458-9.)

Not just any reliable way of coming to judge will be one to which you are entitled. The entitlement to a judgment made in a particular way is grounded in constitutive features of the states and contents involved in that way of judging—including, perhaps, what is involved in understanding the judged content.

On Burge's view, there is an incompatibility between entitlement and reasons, because basing a judgment on a reason involves understanding why the reason warrants the judgment in a way that is precisely not required by entitlement (Burge, 2003). On the hybrid view, however, there is a way of basing a judgment on a reason that does not involve having a full grasp of why the reason warrants the judgment, or employing the concept of a reason. That there is such a way is provided for by the distinction between being sensitive to the fact *that* R makes *p* likely to be true, and grasping the argument that articulates *why* it does so. The former, but not the latter, is required for a judgment that *p*, made in a certain way, to be warranted by the reason R.[25]

Why should we prefer this hybrid conception of warrant over internalism and externalism, for the case of self-knowledge? Simply because it preserves what's plausible about each of those views, while avoiding their problems.

Externalism correctly holds that the way in which a subject comes to make a judgment is crucial to its warrant. The problem with externalism about self-knowledge is that purely external properties of a way of coming to judge, independent of its rationality, are not sufficient for the sort of warrant we ordinarily enjoy when we make self-ascriptive judgments. This is not a problem for the hybrid view, since it gives a central epistemic role to reasons, which it conceives of as considerations that, for the subject, point to the truth of the content judged.

This central role for reasons and rationality is what is attractive about an internalist conception of warrant. As we saw, however (4.3.2), the internalist conception cannot capture the warrant we enjoy for self-ascriptions, without a commitment to the discredited introspectionist account of self-knowledge. The source of this difficulty is its requirement

25 Susan Hurley (2001) also defends the view that doing things for reasons doesn't require having the concept of a reason.

that the warranted subject grasp a valid argument with the content of the warranted judgment as its conclusion. The hybrid view avoids this difficulty, since it rejects that requirement. I will show in the next two chapters that there is a non-introspectionist epistemic account of self-knowledge that is compatible with, and indeed relies on, the hybrid view of warrant.[26]

I conclude that the hybrid view is more promising than either externalism or internalism in accounting for the warrant ordinarily enjoyed by self-ascriptions. Thus, there is a way of coming to know, based on reasons, that does not involve having access to the argument that would articulate the truth-conducive connection between your reason and the content known.

4.4 Judging for a reason

I have offered an account of what it is for something to be a reason to judge a given content, and an account of what it is for a subject to have that reason. But a judgment is warranted by a reason only if it is made *for* that reason. So a full account of self-knowledge, based on reasons, must involve an account of judging for a reason. That is what I offer in this final section of the chapter. I will argue that, when you judge for a reason, the consideration that constitutes your reason must be consciously entertained in your coming to judge.

To say that you act, or judge, for a reason, is to say that your action or judgment is caused or brought about in a certain way: it must be caused or brought about in the right way by your recognising, in some sense, the force of that reason in recommending the action or judgment you are performing (or trying to perform). This recognition, I have argued, need not be a judgment or belief. It is just whatever explains the fact that, in acting for that reason, you are not acting blindly, but doing what is appropriate from your point of view.

Recognising the force of a reason in recommending a judgment with a given content is recognising the force of that reason in making that content likely to be true. You recognise the force of a reason in this way when, in virtue of that reason, the truth-condition of the

26 The hybrid view also delivers intuitive results for the case of colour judgments, which I have been using as an example. Our colour judgments manifest a sensitivity to the way in which particular colour appearances, in the right conditions, indicate that objects are coloured in certain ways. This sensitivity is fundamental to the full understanding of the contents of such judgments. A subject who lacked such sensitivity would be a subject who, for example, took colour appearances at face value even in manifestly abnormal lighting conditions. Intuitively, such a subject would lack knowledge even when he made a correct colour judgment (if he were capable of genuine colour judgments: see note 24, above), because the judgment would not be a manifestation of a sensitivity to the truth-conducive connection between colour appearances and the colour properties of objects. On the other hand, those of us who *do* manifest the appropriate sensitivity do so without any knowledge of theoretical principles that explain why the truth-conducive connection holds.

content strikes you as being (probably) met.

When you come to perform an action, there will be many considerations that contribute to the action's being appropriate from your point of view, and whose doing so plays *some* causal role in your coming to act. But by no means all of these considerations will count as reasons for which you act. A reason-attributing explanation of your act of Φing is, in part, a story of your *coming to* Φ. The appropriately explanatory considerations will include only those that make a relatively direct contribution to your Φing, and so are salient in such a story.

This is true of judgment, as much as of any other action. When you judge that your train will arrive late at its destination, various physical principles—for example, that time is proportional to distance and inversely proportional to velocity, and that space is approximately Euclidian—will contribute to the appropriateness, from your point of view, of that judgment, and will be such that, were they not features of how the world strikes you, you would not have made the judgment. But they will not count as reasons for which you make the judgment. Your reason will be, say, that the train departed its origin late. This consideration, through featuring from your point of view, is directly causally involved, and plays an appropriate explanatory role, in your coming to judge that the train will arrive late.

So, there are at least two necessary conditions for something to count as the reason for which you act: you must recognise it as favouring the type of action you try to perform, and your so recognising it must play the right sort of role in your coming to act.

For any act of Φing, there will typically be many descriptions of your coming to Φ at various subpersonal and personal levels. The descriptions that mention the reasons for which you Φ will be the one that displays how, *in* coming to Φ, you were doing what made sense, what was appropriate, to you. This is the point of reason-attributing explanation. It involves not just your doing something that in fact you could make sense of, but your coming to do it *in the light of* its making sense. The reason for which you Φ is not *merely* a feature of how things are for you (a potential reason), but is *brought to bear* in a salient way in your coming to Φ (an actual reason for which you Φ).

Judging is by its nature a conscious, reason-guided mental act.[27] The description of your coming to judge that mentions the reasons for which you judge will be a description at the level of conscious thought and experience. For this is the level at which it is determined that

27 It is not always guided by good reasons. I will say more about the nature of judgment in Chapter 6.

in coming to perform a conscious, reason-guided act, you are acting appropriately from your own point of view. It is only by entering into the contents of the thoughts and experiences by means of which you come to judge, that a consideration can play *both* the rational role *and* the salient causal or bringing-about role of the reason for which you judge. Nothing else could ensure that it is in virtue of that content that, *in* judging, you are doing what makes sense to you. If we supposed that the reason for which you judge need not be involved in the thought and experience by which you come to judge, we would be severing the connection between reason-attributing explanation and your consciously doing what is appropriate from your point of view, in the light of its appropriateness.

My claim applies to acts of judgment, not to all actions performed for reasons. Perhaps it is possible to act for a reason without that reason showing up in consciousness at all. Perhaps habitual actions are of this sort; habitual actions are arguably performed for reasons, and are often performed thoughtlessly, so to speak. But judgments are not at all like that. Judgments are episodes in conscious thought, directed towards ascertaining the truth, and guided by reasons pertinent to that goal.

It might be said that at least one species of judgment resembles habitual action—namely, perceptual judgment. After all, perceptual judgments are not usually preceded by deliberation; they can seem almost automatic. But in fact, the reasons for which perceptual judgments are made *do* show up in consciousness—they show up in the conscious perceptual experiences on which such judgments are based. It's not that you must have a thought *about* your perceptual experience, in order to make a perceptual judgment; the experience itself is a conscious episode that gives you a reason (or, more precisely, gives you *access* to a reason). Compare an ordinary perceptual judgment to the sort of judgments made by patients with blindsight. Intuitively, blindsight patients lack reasons for their forced-choice judgments, at least until they learn that their guesses are reliable. They lack such reasons, it seems, because for them no reason-giving consideration comes consciously to mind. The perceptual judgments of ordinary subjects, by contrast, are based on reason-giving considerations consciously presented to those subjects.

In claiming that reasons for judgment must show up in thought or experience, I am not implying that judgments are always preceded by deliberation, in any natural sense of that term. You often judge that *p* just because it seems, perceptually or otherwise, that *p*. In such cases there is no deliberation, but *p*, or something related, must show up in the occurrent thought or experience by which you come to judge. Otherwise it would be a mere leap in the dark, like the judgment of a blindsight patient. None of this requires you to reflect on your

reasons, conceived as such, for and against *p*.

Let me finish by making some clarificatory remarks about the argument I have given in this section.

I am appealing to the distinctive explanatory role of motivating reasons. Earlier (4.1), I distinguished motivating *reasons* from *motivations* in a broader sense. Thus, I am not claiming that our motivations, for judging or anything else, are always explicitly entertained in thought. I am claiming that when our motivations are not explicitly entertained, they are not our *reasons* for acting. There is empirical evidence that subjects can be influenced in expressed preferences by factors that do not enter their conscious thought at all (Nisbett and Wilson, 1977). These factors can be said to motivate the subjects. They might, in some cases, count in favour of the preference they motivate: they might be reasons for the subjects to choose as they do. But they are not the subject's *own* reasons *for* choosing as they do, because they do not cause the subject's choice *by* being factors *in the light of which* the subject chooses.[28]

The argument of this section does not involve the claim, either as premise or as conclusion, that we always know our own reasons for judging. To say that your reason for judging must show up in thought or experience is not to say that you must think of it *as* a reason, nor is it to say that you must have higher-order awareness of your experience or thought. (I have claimed that you must recognise your reason's normative force, but I emphasised that this recognition can be understood quite minimally.) For some fact (say) to show up in thought is simply for you to have an experience or thought that represents that fact, not for you to think "This fact is a reason to judge that *p*", or to think "I am thinking of this fact". And the argument for motivating reasons showing up in consciousness depends on the explanatory role that motivating reasons for judgment must have, not on any self-knowledge that subjects must have. There is no suggestion that you must be in a position to grasp the motivating-reason explanation of your judgment.

To make clear what account of judging for a reason is being defended here, we can consider the following example. Suppose you are presented with a blue shirt and a yellow shirt, and you judge that the blue one is nicer. Suppose your judgment can be explained in terms of a reason (or putative reason) that you have in virtue of believing that blue is nicer than yellow.

28 Another pertinent line of empirical evidence comes from commissurotomy patients, who can be induced to act appropriately in response to a stimulus of which they apparently lack conscious awareness, and will 'confabulate' a mistaken explanation of their action (e.g. Sperry, 1985; Gazzaniga, 1995).

My claim is that this consideration, that blue is nicer than yellow, must have showed up in the thoughts or experiences that led to your choosing the blue shirt. You need not be able to report that it is your reason. It could simply be that the blue shirt appeared to you to be the nicer in respect of its colour, and this appearance was involved in your coming to judge, such that your judgment is explained by the blue shirt's seeming nicer to you in that respect. You might mistakenly report that the excellent stitching was your reason. But if the superiority of blue did not show up at all for you, then it would not be true that blue being nicer than yellow was *your* reason *for* judging—even if, by some mechanism, you always pick blue things over yellow ones.

4.5 Conclusion

A reason to judge a given content is a consideration that makes that content likely to be true. You *have* a reason to judge a given content when such a consideration is accessible from your point of view, in such a way that you could come to see that content as likely (in that respect) to be true. You can be rationally sensitive to a reason-giving, truth-conducive connection without having access to the inference that would explain that connection. When you judge *for* a reason, you recognise the consideration that constitutes your reason as favouring the truth of the content judged, and that consideration shows up in the conscious thoughts or experiences by means of which you come to judge.

The next chapter will address the question of how you know the *contents* of your conscious episodes, and how you know that *you* are the subject of those episodes. I will use the foregoing account of reasons for judgment in order to defend the claim that the occurrence of a conscious episode with a given content can give you a reason to self-ascribe that content, a reason that you can exploit in coming to self-ascribe it.

CHAPTER 5

SELF-KNOWLEDGE OF CONTENT

This chapter begins to set out and defend an epistemic, non-introspective account of self-knowledge that fits into the epistemological framework outlined in Chapter 4. I call it "the moderate epistemic account", because it claims that self-knowledge is based on reasons, but that we have these reasons in virtue of having ordinary, first-order experiences and thoughts directed on the world, rather than introspective experiences *of* those experiences and thoughts.

There are, recall, three components of a self-ascription: the first-person component, the type component, and the content component. This chapter deals only with the first and third of those components. It offers an account of how, when you self-ascribe a conscious state or episode with the content C, you know that *you* (rather than someone else, or nobody) are enjoying a certain conscious state or episode, and how you know that its content is *C* (rather than C*). It thus answers two questions: how enjoying a conscious state or episode, even when it is directed on the external world, can give you knowledge about *yourself*, and how merely entertaining some content can give you knowledge *about* that content. The account of the second component of self-ascriptions, the type component, will be presented in Chapter 6.

Thus, what I offer in this chapter is an account of the warrant for judgments of the form, "I am entertaining the content C", where 'entertaining' means enjoying a conscious state or episode with that content. The way in which I formulate the judgment is not supposed to reflect the way in which subjects actually express their judgments. (A more natural locution, at least for cases in which the content is a proposition, would be "I am thinking that *p*", where "thinking" does not indicate any particular *type* of episode.) The formulation is merely supposed to pick out which aspects of the phenomenon of self-knowledge I am explaining.

The chapter will proceed as follows. In section 5.1 I will set out the claims to be defended. Section 5.2 will argue, in three subsections, that a self-ascription of a content, based on the reason you have by virtue of entertaining that content, can amount to knowledge. 5.2.1 will describe the method for issuing such self-ascriptions and show that it is reliable. 5.2.2 will show that the availability of this method does not presuppose self-knowledge, given the epistemological framework of Chapter 4. 5.2.3 will argue that self-ascribing in the way described is rational, because it is an exercise of a capacity that is fundamental to our

thought, namely a practical grasp of the first-/third-person distinction. In 5.3 I will show that self-knowledge arrived at in this way will have the distinctive features of self-knowledge. Finally, in 5.4 I will tie up a loose end by showing that the moderate epistemic account applies to contents not represented as true (as when you merely suppose that it is raining), as well as to those that are represented as true (as when you see or judge that it is raining).

5.1 The moderate epistemic account for self-knowledge of content

The account I wish to defend, regarding the first-person and content components of self-knowledge, comprises the following two theses:

(**R**) A subject who enjoys a conscious state or episode with some content thereby has a reason to make the self-ascriptive judgment that she is entertaining that content, and a self-ascription of content made for that reason amounts to knowledge.[1]

(**N**) When a subject has a good reason for a self-ascription of a content in virtue of enjoying a conscious state or episode with that content, her having the reason does not depend on experiential awareness of herself, or of her state or episode, or of its content, whether this is construed as constitutive of the state or episode or as distinct from it. The modification of consciousness on which her having the reason depends is the self-ascribed state or episode itself.

This account is epistemic because it holds that self-knowledge is genuine knowledge, and that its status as such is explained by its being based on reasons. It is like introspectionism in the further respect that it claims that the very state or episode you self-ascribe plays a central role in explaining your warrant to self-ascribe it. But it is non-introspectionist (hence moderate) because it does not attribute any role to introspective experience of experiences or thoughts in the ordinary warrant for self-knowledge.[2]

It is a challenge to explain how the moderate epistemic account could possibly be true.[3] Most

1 See Chapter 4 for my view of what it is to have a reason.
2 The account does not entail that there is no such thing as introspective experience. Perhaps such experiences can be an source of additional warrant for self-ascriptive judgments.
3 A challenge of this sort is articulated by Martin (1998), directed at the account offered by Peacocke (1998), which is similar to mine.

experiences and thoughts concern the world outside the subject. That is, their contents are first-order; and in having such an experience or thought, you direct your awareness to some aspect of the external world. How could such an experience or thought, or its content, give you a reason to make a judgment about *yourself*?4 And how could it give you a reason to make a judgment about a *content* of experience or thought (as opposed to its object)? The challenge is pressing precisely because the claim is that having an experience or thought gives a *reason* for a self-ascription. It may be that there is a *reliable* connection between experiences and thoughts on the one hand, and self-ascriptions of content on the other.[5] But judging for a reason, as we saw, involves appreciating the connection between the reason you have and the truth of the judgment that is rationalised by it. It is not clear what connection there is between the consideration you are aware of in entertaining an ordinary first-order content C, and the truth of the self-ascriptive content "I am entertaining the content C".

It is important, in appreciating this challenge, to distinguish between *entertaining* the content C and being *aware of* the content C. When you entertain a content, what you are aware of is typically some putative fact or state of affairs in the world. Contents are senses; what you are aware of is referents. In entertaining the content that it is raining, you are thereby aware, in a particular way, of a putative fact or state of affairs—its raining. It is another matter to be aware of the content 'it is raining', conceived as a content (a sense). To be aware of a content is to be aware of one of the ways of thinking about some putative fact or state of affairs. To be aware of a first-order content is to entertain a second-order content. So the transition from *entertaining* a content to *ascribing* that content is far from trivial: it is a transition from awareness of a putative worldly fact or state of affairs to awareness of a particular way of thinking about that putative fact or state of affairs.

---

4 The challenge also arises for experiences and thoughts that concern oneself. Suppose you think, "I am cold", and move from that thought to the self-ascription "I am thinking that I am cold". Although the reason-giving first-order thought does involve an occurrence of the first-person, it's very unclear that this occurrence does anything to rationalise the first occurrence of the first-person in the self-ascription. So the challenge does not turn on whether our thoughts concern ourselves or the outside world.
On some views, there is an explicit first-personal element in the content of some experiences. A visual experience, for example, can represent the content "I am in front of a table." Even if that is correct, it is a challenge to explain how such an experience could rationalise the first occurrence of the first-person pronoun in the judgment "I see that I am in front of a table."

5 Davidson (1987) and Burge (1996) have argued that, in some cases, there is a constitutive connection between the contents of self-ascriptions and the contents they self-ascribe (or between what is involved in thinking the former contents and what is involved in thinking the latter contents). Such self-ascriptions will be infallible about the content self-ascribed. But that still doesn't explain why contents give (access to) *reasons*.

I will claim that entertaining a content can amount to having a reason, and the reason you have in virtue of entertaining a content is determined by that content. When I talk about a content giving a reason, I am referring to the reason a subject has in virtue of entertaining that content. To give a reason, in this sense, is to provide access to a reason, not to *be* one. The content (sense) determines a referent—a fact or state of affairs. What the subject is aware of, and what constitutes her reason, when she entertains the content, is that fact or state of affairs.[6] When you judge that *p* for the reason given by a content C, you are making a connection between the fact or state of affairs you are aware of in entertaining C, and the truth of *p*.

There is another complexity to be addressed. Martin (2002) distinguishes between two conceptions of intentional content, the stative conception and the semantic conception. On the stative conception, something has intentional content when it puts something forward as being the case—as true. That is, representation in the stative sense is non-neutral with respect to the truth of what is represented. On the semantic conception, something has intentional content when it is *about* states of affairs or objects or properties. Something can represent in the semantic sense, while being neutral about whether the state of affairs represented obtains. Thus, experiences and judgments are representational in the stative sense, whereas many of our other conscious thoughts—for example, when you suppose or imagine that something is the case—are representational only in the semantic sense.

The moderate epistemic account is intended to apply to both sorts of mental episode—those that are representational in the stative sense, and those that are representational in the merely semantic sense. However, it might seem as though the account applies more straightforwardly to stative than to semantic representation. It is relatively clear, given the aim of judgment, how a thought or experience that puts something forward as true can give a subject a reason for judgment. It is less clear, however, how a thought or experience that does not put forward its content as true could give a reason for judgment. How could (say) merely supposing that it is raining give you a reason to judge anything—to think anything in particular is *true*? For most of the rest of this chapter I will leave this distinction aside; in effect, I will talk as though all conscious thoughts and experiences are representational in the stative sense. In section 5.5 I will return to the distinction, and argue that the account I have

6 Here I leave aside various philosophical perplexities about the ontology of facts and their relation to the sense-reference distinction. I will use the natural-sounding locution 'aware of a fact', in order to bring out the difference between entertaining a content (being aware of a putative fact) and being aware of a content (which is different from being aware of a fact). I intend to remain substantially neutral on any controversial disagreements about facts.

offered applies to semantic as well as to stative representational episodes.

5.2 Why self-ascriptions of content are knowledge

The task of defending the moderate epistemic account falls into two broad subtasks. First, I must show that there is a way of coming to make self-ascriptive judgments, based on reasons, such that judgments made in that way are true, or likely to be true. Meeting that challenge will amount to showing why the reasons for which those self-ascriptions are made really are good (normative) reasons for self-ascriptions. A reason for a judgment, as we saw in Chapter 4, is a consideration that supports the truth of the content of the judgment—support that can be articulated in an argument. If there is a truth-conducive way of basing judgments on a consideration, then that consideration supports the truth of the content judged. And the support is articulated by the argument that that way of making judgments is indeed truth-conducive.

However, not just any truth-conducive way of coming to make self-ascriptive judgments will do. The account must involve plausible claims about what ways of coming to judge are available to thinkers. I cannot simply assume from the start that you will be willing to self-ascribe conscious episodes when, and for the reason that, you are enjoying them. Given the accessibility requirement on reasons for which you judge, such an assumption would smuggle in what I am supposed to be explaining—your epistemic access to your own conscious episodes.

The second subtask in defending the moderate epistemic account is to show that in making a self-ascriptive judgment in the way to be described, you can be judging rationally. This is a distinct task because, as we saw in the last chapter, your judgment is rationalised by a reason only when you recognise the reason's force. In the context of the hybrid view of warrant (section 4.3.3), the challenge will be to show that making a self-ascription in the way to be described can be a manifestation of sensitivity to the truth-connection between reasons for self-ascriptions and self-ascriptive contents, a sensitivity grounded in the capacities constitutive of the understanding of such contents.

I will begin with the first of these subtasks.

5.2.1 The reliability of self-ascriptions of content

I claim that the way in which we come to make self-ascriptions of content is this:

(**W**) For the reason given by the content C, the subject judges "I am entertaining the content C".

I say "the reason *given* by the content C" to indicate that the reason is given to the subject in virtue of her *entertaining* that content, and thus being aware of the world, not in virtue of her being aware *of* that content as such. The reason itself is the fact or state of affairs she is aware of in entertaining C.

(**W**) is deliberately formulated in order to be neutral on whether the content C is conceptual or nonconceptual. But the claim that is being made can be made somewhat clearer by considering the case where the content is conceptual. When a subject has a thought or experience with a conceptual, propositional content, her reason is given by the proposition she thereby entertains. Thus:

($\mathbf{W}_{\mathbf{conceptual}}$) For the reason that *p*, the subject judges "I am entertaining the content '*p*'".

While not all self-ascriptions will be made according to ($\mathbf{W}_{\mathbf{conceptual}}$), this formulation brings out the point that, in coming to self-ascribe a content (conceptual or nonconceptual), you make a transition from a thought or experience *with* a content, to a judgment *about* that content (and about yourself) (see 5.1). An example of this would be judging "I am entertaining the content 'It is raining'" for the reason that it is raining.

Why will self-ascriptions made according to (**W**) be true? In many cases, the truth of a judgment based on a reason is explained by the reason-giving content's *truth*, and the truth-conducive connection between that content and the content of the judgment.[7] However, the explanation of why a self-ascription of C based on the reason given by C will be true appeals not to C's truth, but to the conditions for its being a reason for which a subject makes a judgment. According to the account of judging for a reason that I offered in Chapter 4 (section 4.4), the reason for which you judge is a consideration that shows up in your consciousness, in your coming to judge. Therefore, when you have that reason in virtue of a

7 Or, if the reason-giving content is nonconceptual, we appeal to the truth-preserving character of the non-inferential transition from that content to the judgment-content.

particular content, that content must be the content of a conscious state or episode that you enjoy in coming to judge. If you make a judgment for the reason given by C, C is the content of a conscious state or episode that you enjoy in coming to make that judgment. So it is true of you, in that case, that you are entertaining the content C. Thus, you could not make a self-ascription according to (**W**) if the self-ascription were not true.[8]

When I claim that (**W**) specifies a method of arriving at unique judgments, and a method that is available to us, I am making an important assumption, which I should make explicit here. In making the self-ascriptive judgment "I am entertaining the content C", you employ a higher-order concept, expressed by "the content C", to pick out the content C of your first-order thought or experience. Call this higher-order concept Con(C). I have been assuming that Con(C) picks out the very content C that gives you the reason for the self-ascriptive judgment whose content contains Con(C). Thus, I have been assuming that for any content, there is a unique 'canonical' concept of that content, and you can react to the reason given by the content by employing the canonical concept of that content.[9] This assumption is essential for my account to work for the content component of self-ascriptions.

How can you react in this way to the reason given by a content, without first having higher-order experiential awareness of your entertaining that content? The answer is that it is constitutive of possession of the canonical concept of some content that you be willing to employ it when you entertain the content it picks out. More generally, it is constitutive of being able to think *about* contents—of taking a perspective on contents—that you be willing to make this transition from entertaining a content to entertaining in judgment the canonical concept of that content. The possibility of being sensitive in judgment to what contents you are entertaining ought not be controversial. Thought and reasoning of almost any sort requires it: we could not make inferences if we were not sensitive to what contents we were entertaining, since these contents constitute our premises.[10] The puzzle is rather how it can

---

8 This does not mean that our self-ascriptions of content are infallible. It shows only that, when you exercise (**W**) *correctly*, you will not go wrong.

9 This view is defended forcefully by Burge (1979, 2004), and also, more recently, by Peacocke (2007), who has changed his mind from an earlier position denying that there are concepts of contents distinct from those contents themselves (Peacocke, 1996).
The claim I am making is closely related to a claim that originally came up in the debate over whether content externalism is consistent with authoritative self-knowledge—namely, that a self-ascription of an externally individuated content will not go wrong because the self-ascriptive content will inherit the externalist character of the content it ascribes (see Davidson, 1987). Sometimes this is expressed by saying that the content of the ascribed thought is "carried over" to the self-ascription. But, as Burge has emphasised, the self-ascription need not involve a content *identical* to that of the ascribed thought. A weaker relation will do, as long as the ascribed content determines the second-order content.

10 If perceptual experiences have nonconceptual representational contents that make a difference to

be *rational* to make a transition from entertaining a first-order content, and thus being aware in some way of some fact or state of affairs in the world, to entertaining the canonical concept of that content, and thus making a judgment about that content. I will deal with that puzzle in section 5.2.3.

5.2.2 The availability of (**W**) and the hybrid view

I noted above that an account of how we come to make self-ascriptions must describe a procedure that we can actually follow, and that does not presuppose self-knowledge. In this brief section I want to show that (**W**) meets those criteria, in the context of the hybrid view of warrant presented in Chapter 4.

When you judge for a reason, you have a certain sort of epistemic access to your reason. (**W**) attributes to the subject only the reason given by the first-order content of some thought or experience of hers. It is uncontroversial that the contents of thoughts and experiences you have can give reasons available to you: entertaining such a content just is being aware of some potentially reason-constituting fact or state-of-affairs. (**W**) does not presuppose anything it is intended to explain.

The argument I offered for the reliability of (**W**) mentions in its premises not the truth of the reason-giving content, but its giving a reason for which a subject judges, a reason which the subject therefore *has*. In the case of self-knowledge, something's *being* a reason for a self-ascriptive judgment cannot be explained independently of a subject's *having* that reason. A fact or state of affairs is a reason for you to self-ascribe a certain content only when, and because, you have that reason in virtue of entertaining that content.[11]

If we conceive of knowledge along internalist lines, then this way of explaining the reliability of (**W**) is incompatible with (**W**) being a way of acquiring knowledge. According to the internalist view, your having the reason given by C to judge that *q* requires not only that C features in your point of view, but that you have access to the premises of the argument that articulate the truth-connection between C and *q*. But the argument I have

---

what judgments we are willing to make on the basis of those experiences, then we must equally be sensitive to what nonconceptual contents we are entertaining. Is there a canonical concept of any given nonconceptual content? Plausibly, the canonical way of picking out a nonconceptual content is demonstrative. For example, when presented in experience with an object of a certain unconceptualised shade, you can think of the object's looking *that* shade. See Peacocke (2001).

[11] In section 4.1 I claimed that statements of the form (b), 'You have reason R to Φ', do not generally follow from statements of the form (a), 'R is a reason for you to Φ'. My claim now is that, in the special case where Φing is self-ascribing some content, (b) does follow from (a).

offered involves, as a premise, that you are entertaining the content C. To have access to that premise, you would have to already have a higher-order awareness of your entertaining the content C. Thus, if following (**W**) could give you self-knowledge, conceived in this internalist way, you would have to already have self-knowledge. This is another version of the circularity problem faced by internalism in explaining self-knowledge (section 4.3.2).

The hybrid view has no such difficulty. It is consistent with the hybrid view that C's giving a reason for you to issue a self-ascription is explained and constituted in part by your having that reason. This is in part because having a reason does not require awareness of the premises of the argument for the truth-conducive connection between the reason and the content for which it is a reason. On the hybrid view, that a consideration's being a reason depends on your being aware of it by entertaining some content does not entail that, when you *are* aware of it, it is not consideration itself (rather than your being aware of it) that constitutes the reason; i.e. it does not entail that simply entertaining the content alone is not sufficient for warrant. The warrant for a judgment can depend on being aware of reasons for the judgment, without depending on awareness of the full conditions that make them good reasons for that judgment.

It might seem puzzling that something's *being* a reason for judgment can depend on your *having* that reason for judging. But why suppose that there can't be cases where something's being a reason for something is explained by someone's having that reason? For a consideration to *be* a reason, recall, is just for it to make likely the truth of some content. For you to *have* a reason, is for it to feature from your point of view. There is nothing circular about the claim that a consideration supports the truth of some content only when and because it features from someone's point of view. What it is for a consideration to feature from someone's point of view is not explained in terms of what it is for that consideration to be a reason for something.

I have shown that there is a way of coming to make self-ascriptions of contents based on reasons such that the self-ascriptive judgments will be true. I have shown that if the internalist conception of warrant were correct, this could not be a way of acquiring self-knowledge, but that this difficulty is avoided if we adopt the hybrid view. I have yet to show, however, that the full conditions for knowledge, as conceived by the hybrid view, will be met by a subject who makes self-ascriptive judgments according to (**W**). That is my next task.

5.2.3 The rationality of self-ascriptions of content

To defend the moderate epistemic account of self-knowledge of content, I must show that self-ascribing content according to (**W**) will be rational from the subject's point of view.

Since (**W**) is reliable, the reasons for which self-ascriptions of content are thereby made are in fact good reasons for such self-ascriptions. To show that self-ascriptions are rational, however, I must show that subjects can make them because they recognise the force of those reasons. I must show that subjects, in judging according to (**W**), are manifesting sensitivity to the connection between the reason for which they self-ascribe, and the truth of the self-ascription.

My claim in this section will be that (**W**) is available only to subjects for whom making judgments in that way will indeed be rational. The argument is that certain capacities involved in being able to make judgments in that way are sufficient for a subject to have a rational sensitivity to the connection between the reason given by C and the truth of the content "I am entertaining the content C".

One way to make such an argument is simply to state the necessary conditions for possessing certain concepts. We could hold that willingness to make self-ascriptions of contents, when entertaining those contents in conscious episodes, is a condition on possessing the canonical concepts of those contents and the first-person concept—the concepts employed in a self-ascriptive judgment. Such self-ascriptive judgments, we could add, are rational simply because willingness to make them in the appropriate circumstances is constitutive of understanding what it is for their contents to be true. The strategy of trying to explain the rationality of certain judgments in this way is familiar from the work of Chris Peacocke (1992; he uses this strategy to explain the rationality of certain self-ascriptive judgments in his 1998, 1999). The shortcoming of this 'concept-possession' account, as it stands, is that it simply stipulates that making self-ascriptive judgments in the specified way is constitutive of understanding them, rather than explaining why it is rational in a way that respects the connection between rationality and the subject's point of view. Peacocke states that it is constitutive of possession of a concept that you find "primitively compelling" certain ways of judging contents involving that concept: but this is not yet to connect *what* a subject finds compelling with the knowledge that the subject has of what it is for something to answer to the concept she possesses. Peacocke himself acknowledges this point, in relation to the concept of conjunction:

> "The possession-condition offered for the logical concept of conjunction mentions what the thinker finds primitively compelling. But rational judgment is not simply yielding to what one finds primitively compelling. […] [T]he need for a better form of account cannot be met simply by restricting the requirement that something be found primitively compelling to judgments which are properly written into the possession-conditions for the concepts in question. What we are missing is a requirement that the judgment or transition be rational from the thinker's own point of view." (Peacocke, 2004.)

In pursuing a more satisfying explanation, we might ask what is involved in possessing the various concepts involved in a self-ascriptive judgment "I am entertaining the content C". It is platitudinous that possessing a concept consists in knowing what it is for something to answer to the concept. A subject who possesses concepts of contents knows what it is in general for someone to entertain a content. Let us use the term "awareness" to denote the property a subject has when she entertains some content in a conscious state or episode, regardless of what *type* of state or episode she is enjoying. Then, a subject who can *self*-ascribe a content—that is, any subject capable of judging according to (**W**)—has some grasp of the conditions under which it is true that: someone is in a state of awareness (is entertaining a content), and that someone is herself.

But it is not obvious how this helps. When a subject comes to make a judgment by valid inference, we can explain how judging in that way is rational from the subject's point of view by saying that the subject knows what it is for the premises of the inference to be true, and knows what it is for the conclusion to be true, and thus knows that when the truth-conditions for the premises are met, the truth-conditions for the conclusion will (probably) also be met. But of course there is no valid inference from '*p*' (say) to "I am entertaining the content '*p*'", and a subject who understands both contents knows that the truth-conditions for the former can be met without those for the latter being met.

Indeed, the problem is not just that there is no parallel with valid inference, but that, on the face of it, the subject's knowing what it is for a self-ascription to be true does *nothing* to explain the rationality of (**W**). To see why, consider a subject who possesses canonical concepts of contents, and the first-person concept. Such a subject knows what it is for herself to be in a state of awareness with some content. Why should such a subject, when aware, in entertaining the content C, of some state of affairs in the external world, suppose that that state of affairs has anything to do with her own state of awareness? After all, the contents of our thoughts and experiences are not typically given to us *as* contents of thoughts and experiences; rather, with them we are given (in particular ways) states of affairs in the world. We can imagine, without any surface incoherence, such a subject being prepared to make

ascriptions of content to others and to herself, on the basis of behavioural evidence, and yet seeing no connection between (say) the state of affairs that it is raining and the state of affairs of her entertaining the content "It is raining"—an apparently independent psychological phenomenon. She sees no systematic connection between states of affairs in the world, and her own state of awareness. Such a subject will be unwilling to self-ascribe according to (**W**). The concept-possession account of the rationality of (**W**) claims that such a subject is not genuinely possible. But then it is a challenge to explain why.

It seems to me that we can make progress towards a satisfying explanation by considering what would be *lacking* in a subject who (*per impossibile*, I will claim) was competent with the various concepts involved in a self-ascription, but refused to make self-ascriptive judgments on the basis of contents of present thought and experience. The subject we considered in the last paragraph has a detached understanding of what it is for herself to be in a state of awareness. She understands it as a state of affairs of the very same type as occurs when some other person is in a state of awareness, the only difference being that the someone, in this case, is herself. She understands it as something she could come to know about in just the same ways as she could come to know about the awareness of others. Now, a full understanding of awareness and the first person does indeed involve a grasp that for oneself to be aware is for one to have the same property as any arbitrary person does when they are aware. But there is a fundamental distinction between two ways of conceiving of awareness: conceiving of it third-personally, and conceiving of it from the first-personal perspective. Our subject fails to conceive of awareness from the first-personal perspective. To conceive of awareness from the first-personal perspective is to conceive of it from the inside. Corresponding to this way of conceiving of awareness is a distinctive way of knowing about awareness, which is available to you only with respect to that awareness whose first-personal perspective you occupy. In failing to pursue this way of knowing, our subject is failing to understand awareness in its first-personal dimension. That is the idea I wish to develop in the rest of this section.

Making self-ascriptions of content (of awareness, in the above sense) on the basis of reasons given by the contents of present thought and experience is an exercise in competent negotiation of the first-/third-person distinction. Refusing to make such self-ascriptive judgments is a failure to grasp that distinction. I want to argue that grasp of the first-/third-person distinction is a general capacity that grounds, and thus helps explain the rationality of, particular self-ascriptive judgments.

We can get a grip on what the first-/third-person distinction involves by considering how it is

negotiated in (**W**). At one level, there is a transition from a third-person content to an explicitly first-personal content: from the content C to the content “I am entertaining the content C”. At another level, there is a move *away* from the first-person perspective. From a state of awareness that constitutes the subject's immersed, first-person perspective on the world (awareness that it is raining, say), the subject moves to a higher-level, reflective perspective in which her immersed first-person perspective is represented (the judgment “I am entertaining the content 'it is raining'”). Roughly, there is a transition from a third-personal content represented from a first-personal perspective, to a first-personal content represented from a third-personal perspective.

The fact that the transition the subject makes is a transition between two perspectives explains how the *transition* can be a valid one, even though there is no relation of valid inference between the *contents*. It is more akin to a warranted translation procedure than to a warranted inferential step.

An informative analogy can be drawn from David Velleman's paper “Self to Self” (1996). Velleman describes a subject who makes transitions between, on the one hand, perception of the layout and features of his immediate environment, and, on the other hand, representations that locate himself and various environmental features in an objective spatial frame of reference, given by a map of the region. The subject is able to make the transitions because of a legend on the map, which says “This map is here”, and indicates a particular location on the map. In making such a transition, the subject does not *infer* that a state of affairs obtains, from the premise that some other state of affairs obtains. Rather, he grasps how a state of affairs, as presented to one spatial perspective, can be represented from a 'higher-order' spatial perspective that includes that first perspective in its field of vision (so to speak). Similarly, a subject who follows (**W**) grasps how the state of affairs presented to her immersed, first-order, first-personal *cognitive* perspective, can be represented from a second-order cognitive perspective that is a perspective on, *inter alia*, that first-order perspective.

In Velleman's case, the map's legend plays the role of providing an algorithm to translate the coordinates of space as given in perception to the coordinates of space as represented by the map. In the case of (**W**), the translation between cognitive perspectives relies on the fact that the awareness from which one begins is (as it must be) *one's own*. This fact gives a cognitive location to the third-personal content from which the translation begins, just as the map's legend gives a spatial location to the perceptually presented features from which one begins in locating oneself. Here, it is useful to invoke John Perry's (1986) distinction between a thought or experience carrying information *about* something, and its carrying information

concerning that thing. A thought (say) can carry information *concerning* something without *representing* that thing—without carrying information *about* it. The thought that it is raining carries information concerning a certain place (the thinker's locality, usually), but does not represent that place (and the thought could be had by a subject who did not grasp that weather is localised). I claim that our thoughts and experiences carry information *concerning* ourselves, even when they are not about ourselves. A thought or experience carries information concerning you in virtue of being yours. Subjects with the requisite cognitive and conceptual capacities can exploit this feature of thoughts and experiences to follow the 'translation' procedure (**W**) and make self-ascriptive judgments.

A subject who successfully exploits this feature of thoughts and experiences manifests an appreciation that what she is aware *of*, though given in some way simply as a bit of the world, reflects two different things: how the world is, and her own perspective on the world. That is, she grasps that it is not only a part of the world, but a part of the world as it is for her, given in a certain way. A subject can be attributed with this grasp when she manifests an understanding that the world extends beyond the world-as-it-is-for-her.[12] She grasps that the world-as-it-is-for-her is a portion of the world *simpliciter*, given in a certain way, and that how it is given is a perspective on it.[13]

I am suggesting that we should think of the transition from a state of awareness directed on the world to a self-ascription of content on the analogy of a translation between spatial coordinates. It is a translation between the first- and third-person perspectives, governed by a grasp of the first-/third-person distinction. I now want to defend the claim that this grasp, which is involved in the understanding of self-ascriptive contents, can explain the rationality of self-ascriptions. I will argue that we do indeed have that grasp, and that it is a general capacity, fundamental to our thought, and prior to particular self-ascriptive judgments. It can thus ground those judgments. According to the epistemological framework of Chapter 4, a judgment that is grounded in a cognitive capacity in this way will be rational.

What evidence is there that we do have such a capacity? Attributing a grasp of the first-/third-person distinction to subjects is necessary to explain various competences that they manifest. A relatively primitive marking of the first-/third-person distinction can be found in

12 Here there is a crucial role for objective spatial and temporal thought: thought about other times and places, and of one's present time and place as fitting into an objective order. This has received much attention in the neo-Kantian tradition (see, for example, Campbell, 1994; Cassam, 1997).

13 It is important that what the self-knowing subject self-ascribes is not just the putative fact or state of affairs of which she is aware, but a way of thinking about that fact or state of affairs. This is part of why it is tricky to capture the rationality of the transition from entertaining a content to thinking about a content.

the navigation behaviour of even some non-human animals. Animals that display *perspectival sensitivity* mark the distinction. Perspectival sensitivity (the notion is due to Peacocke, 1983, and is developed by Bermudez, 1998) is, roughly, sensitivity to how the spatial relations of objects or locations, including those that are currently unperceived, to oneself, vary as a function of one's changing position in space. An animal that displays perspectival sensitivity can navigate directly towards an unperceived target object, whose location it has learned, not by following a learned route, but by calculating the target object's relative position anew, on the basis of the animal's own present location. Such an animal must be able to work out its own location in objective space, based on the content of its present experience, and then calculate the positions of objects relative to that location. It treats the content of experience as yielding information not only about the environment, but about itself. The animal is manifesting a grasp of the relations between the objective spatial arrangement of the objects of its experiences, its own movements through space, and what it perceives when at a particular location. Behaviour that requires attribution of perspectival sensitivity has been elicited from dogs (Chapuis and Varlet, 1987) and chimpanzees (Menzel, 1973).

This sort of spatial reasoning involves something more like a translation procedure, governed by the first-/third-person distinction, than a valid inference. The animal must make a transition from a perceptual representation of a familiar landmark as being a certain egocentrically specified direction and distance away, to a representation of itself as being at a certain objective location. Here we have a translation between two frames of reference in which spatial coordinates are specified (somewhat like Velleman's case, described above). In the perceptual experience with spatial content, it simply seems to the creature as if the environment is a certain way. But because the creature has an understanding of the relation between itself and its environment, it can recognise the double aspect of the environment's seeming that way: its simultaneous reflection of the environment's *being* that way, and *its* being in a certain location in that environment. This understanding enables it to translate the environment's seeming that way into a representation of its own location in that environment.

When perspectival sensitivity is combined with certain other skills, the resulting abilities involve a more sophisticated grasp of the first-/third-person distinction. Grasp of the distinction can be manifested in various kinds of judgments that are not self-ascriptive judgments. A subject can manifest it in certain kinds of temporal thinking, and in generating explanations and predictions of the course of her experience and the consequences of her

actions (see Campbell 1993, especially pp. 207ff.). For example, a subject might notice some marks on the ground, and reason that another person or animal passed by while she was engaged in some activity; the subject thus manifests a grasp that such goings-on are independent of, but can be taken up in, her point of view.

Although these examples are relatively primitive, it is plausible that self-ascriptive judgments made by following (**W**) count as manifesting, in a more sophisticated way, the very same capacity. Such judgments involve precisely the grasp of oneself as the occupant of a perspective (a cognitive, not merely spatiotemporal, perspective), and of the double aspect of the world's presenting itself to that perspective—its reflecting both the world and the perspective on it.

These examples illustrate that grasp of the first-/third-person distinction is not theoretical. It does not consist in knowledge of propositions. It is a capacity that is manifested in various skills. The exercise of these skills consists in particular behaviours and judgments. A subject's behavioural and cognitive repertoire can require attribution of the *marking* of a distinction that the subject would not be able to think reflectively *about* in its generality. A subject can be aware *of* a fact without being capable of conceptual articulation of that fact. Likewise, a subject's awareness of a property can consist in sensitivity, in thought and behaviour, to its instantiations, rather than in bringing that property under a concept, à la Kant. Such a subject can be said to be aware of the distinction, fact or property in a practical or implicit way.

Thus, grasp of the first-/third-person distinction is not a *source* of reasons or of warrant; it is not a set of warranted beliefs that give reasons, or from which warrant is transmitted. Rather, it is a context in which certain sorts of transitions are rationally warranted. What is grasped in this practical or implicit way will not itself be a reason available to the subject, but can help explain why other considerations are reasons for the subject.

This type of explanation—of the rationality of particular judgments by their being grounded in a practical capacity—is powerful and well established. To return to an earlier example (section 4.4), for many (or maybe all) subjects there are various principles of physics that help to ground the rationality of certain judgments they make, but that those subjects could not think about as such. Such principles help explain why your train's departing late is a reason you have to judge that it will arrive late. But they are not themselves your reasons for the judgment (if you make it), nor is their role one of transmitting warrant from their contents. Your implicit grasp of those principles is a context in which the judgment that the train will arrive late is warranted by the fact that it departed late. There is much more to say

about this sort of practical awareness and its rationalising role, but for my purposes it is sufficient to note that it exists and has explanatory power.

Finally, grasp of the first-/third-person distinction does not consist in, or otherwise depend on, but is explanatorily prior to, self-ascriptive judgments. The examples I have given show that grasp of the distinction is a basic cognitive capacity that can be manifested by subjects who lack the cognitive or conceptual capacity to self-ascribe their conscious episodes. Our possession of the capacity can help explain the rationality of particular self-ascriptive judgments, because it grounds the willingness to make those judgments, and does not presuppose it.

Thus, there is a practical capacity that grounds self-ascriptive judgments, whose possession is prior to those judgments, and that any subject capable of self-ascribing will possess. This explains, within the framework of the hybrid view (section 4.3.3) the rationality of judgments made according to (**W**). Recall that according to the hybrid view, a judgment made on the basis of some reason is warranted if the subject's willingness to make that judgment is a manifestation of sensitivity to the fact that the content of the judgment is likely to be true, given the reason, and if that sensitivity is grounded in the subject's cognitive or conceptual capacities. What I have shown in this section is that grasp of the first-/third-person distinction is a basic cognitive capacity, and that it can ground a subject's willingness to self-ascribe some content for the reason given by that content. Given that capacity, the subject will be sensitive to the truth-conducive connection between what she is aware of in entertaining a content, and the self-ascription of that content. Grasp of the distinction is not a source of reasons or warrant, in addition to the reason for which the subject makes the self-ascriptive judgment; rather, it is a context within which states of affairs a subject is aware of are reasons for self-ascriptions.

That completes my account of the rationality of (**W**). It is worth noting that this account is not a competitor to the 'concept-possession' account I considered at the start of the section (according to which self-ascriptions are rational because willingness to make them is constitutive of possession of the concepts involved). Rather, it can help to deepen that account. For, plausibly, grasp of the first-/third-person distinction is involved in possession of the first-person concept and in possession of canonical concepts of contents. To employ the canonical concept of a content, in response to entertaining that content, *is* to think of the part of the world given to you as a part of the world-as-it-is-for-you, given in some way. To possess such concepts is to be able to take a third-person perspective on your own

perspective on the world.[14]

I have now answered what I called "the knowledge question" (Chapter 1) about self-ascriptions of content. Self-ascriptions are knowledge because they are made by a procedure, (**W**), that is reliable and that makes the resulting self-ascriptions rational. In answering the question I have relied on the epistemological framework set out in Chapter 4.

In the next section I will answer what I called "the specialness question".

5.3 The specialness of self-knowledge of content

We saw in Chapter 1 that self-knowledge appears to have a number of distinctive features, and that an account of self-knowledge ought to explain these apparent features. I added a few more distinctive features in subsequent chapters. What I have offered so far in this chapter is an account of how you know *you* are in some state of awareness, and how you know that the content of that awareness is C (say), rather than C*. I will now show that self-knowledge of content, in both of these respects, has the various distinctive features.

I omit the feature of *commitment* discussed in Chapter 3, since it is associated with self-knowledge of certain *types* of state and episode, rather than with mere self-knowledge of content.

(a) *Security*

Security is the feature that, ordinarily, when you make a self-ascriptive judgment, it is exceptionally difficult for that judgment to be wrong; only in modally distant circumstances could it be wrong, and there is no analogue to perceptual illusion or a sceptical scenario in which you would be mistaken. Self-ascriptions of content made according to (**W**) will be secure, with respect to both their first-person component and their content component.

When you self-ascribe according to (**W**), you could not be wrong about *who* is enjoying the specified conscious episode, nor could you be wrong because *nobody* is enjoying it. The

14 It is part of this explanation that, as Burge (2004) and Peacocke (2007) emphasise, the canonical concept of a content is determined by that content itself. When a subject entertains a content, there is a unique, privileged way in which she can come to think *about* that content—by employing its canonical concept. This way of thinking about a content is made available to the subject, and made rational for her, simply by her entertaining the content, and by her having the capacity to think about contents—to think about what is given as representing a perspective—rather as one who understands a sentence, and understands the use of quotation marks, can talk about that sentence by putting quotation marks around it (this analogy is put forward by Burge, *ibid.*).

availability of the reason for which you judge depends on its being represented in a conscious episode of yours. That is just the nature of judging for a reason.

In fact it seems that self-ascriptions made according to (**W**) will be immune to errors of misidentification relative to the first person—you couldn't come to know about someone else's conscious episode in this way.[15]

Nor could you easily be wrong about which content you are entertaining. A content makes available a canonical way of thinking of *that* content and no other. Slips will perhaps be possible—you may, due to inattention or confusion, call on the wrong canonical concept of a content—but they will be limited by the nature of the procedure and the constitutive requirements of possessing the concept of the content.

These explanations of security do not appeal to brute causal mechanisms that could easily break down, but to the nature of reasons and of concepts. Thus they respect the modal character of security. They also help to explain why widespread error among a subject's self-ascriptions is impossible (see section 3.4).

(b) *Salience*

Salience is the feature that, for any conscious episode, it is exceptionally difficult for its subject (if adequately conceptually equipped) not to be in a position to know about it. The explanation of this is simply that the procedure (**W**) is made available by any conscious episode with a content, since any such episode will provide access to a potential reason for judgment.[16] The procedure enables the subject to know the content, since any content makes available the canonical concept of that content, provided the subject grasps the first-/third-person distinction. And it enables such a subject to know who is entertaining that content, since the first-person component of the self-ascription is made available by the state of awareness being the subject's own.[17] So conscious episodes will be salient with respect to the

15 See Shoemaker (1968) and Evans (1982) for immunity to errors of misidentification. Such immunity comes in stronger and weaker forms. The sort of immunity enjoyed by (**W**) is perhaps only *de facto*: error might be possible in grotesquely abnormal circumstances.

16 I here ignore the distinction between stative and semantic content—it might be thought that this claim is more compelling when applied to episodes that have content in the stative sense than when applied to episodes that have content in the merely semantic sense. I will deal with that issue in the next section.

17 An exception to the salience of (**W**) for the first-person component can apparently be found in the phenomenon of 'thought-insertion'—a symptom suffered by some schizophrenic patients. These patients are unwilling to self-ascribe certain thoughts that occur to them, and that they claim are inserted into their minds by other agents. One way to treat these cases is to hold that the inserted thoughts genuinely aren't the subjects' own thoughts—there is a fracturing of the unity of

first-person and content components of (**W**).

(c) *First-person privilege*

This is the feature that the ordinary way of self-ascribing is available to the subject of the ascribed state or episode, and to nobody else. This will indeed be true of (**W**). You cannot, except in the most outlandish science-fiction circumstances, come to know *what* content another subject is entertaining, or *which* other subject is entertaining a particular content, simply by ascribing the contents you are entertaining. Thus, judgments made according to (**W**) will be first-person privileged with respect to their first-person and content components.

(d) *Authority*

The authority of self-knowledge consists of features of our 'language-game' practices. First, it is inappropriate for an interlocutor to gainsay or raise doubts over a sincere self-ascriptive utterance. Second, it is inappropriate to demand justification for such an utterance—to ask "How do you know?", or "How can you tell?".

The first aspect of authority can be explained by appeal to two other features, namely security and first-person privilege. Since a subject's self-ascriptions made according to (**W**) cannot easily be wrong, and since interlocutors have no comparable procedure for coming to make judgments about the subject's conscious states and episodes, an interlocutor will almost never have sufficient grounds to gainsay such a self-ascription. This aspect of authority is thus a matter of our language-games respecting the actual features of our judgments. It will hold for both the first-person and content components.

The second aspect of authority is not so straightforwardly explained. It might be thought that the moderate epistemic account is actually incompatible with it. The account attributes to the self-ascribing subject a *reason.* This seems to suggest that the subject does have a justification that she could offer, if it were demanded. Why, then, should such a demand be inappropriate?

Further reflection on the practice of demanding and offering justifications, and on the moderate epistemic account, brings out why self-ascriptions must be immune to demands for

---

consciousness in such cases. What *is* the subject's own is the passive experience of apparently having the thought inserted. This is precisely what the subjects *are* able to know about and are willing to *self*-ascribe.

justification. Demanding and offering justifications is a social practice; it is a way of bringing reasons into a shared domain, where they can be evaluated by any subject. The reasons brought into this domain are evaluated, so to speak, from a shared perspective. That is how rational persuasion can occur. Reasons for self-ascriptions, however, are essentially first-personal; their being reasons for self-ascriptions is essentially tied to their featuring from particular first-person perspectives. That it is raining is a reason for me to self-ascribe the content 'it is raining', but it is not a reason for you to make any ascription to me. If you demand a justification for my self-ascription of that content, and I present my reason by saying "It is raining", you will rightly reply, "That has nothing to do with your state of awareness or its content". I could retort that, by giving that reason, I was indicating that I was entertaining that content (not just indicating that it is raining). But then I have re-asserted the judgment, not justified it. So I cannot offer anything that, from any perspective except my own, will count as a reason for the judgment—for either its first-person or its content component. So for you to demand such a justification would be inappropriate.

(e) *Immediacy*

Immediacy is the feature that we do not come to self-ascribe by acquiring or attending to grounds, and that our self-ascriptive judgments do not express things we have found out about ourselves. Although the moderate epistemic account attributes to the self-ascribing subject a *reason*, the subject does not have to do anything, beyond enjoying a conscious state or episode, to acquire or attend to that reason. Just entertaining a content gives you a reason to self-ascribe it; it allows you to appreciate that you are entertaining that content. This is not *finding out* who is in a state of awareness, or what the content of your state of awareness is. It is knowing about the subject and content of your state of awareness just by being in it (and having the right capacities). Immediacy is thus explained for the first-person and content components.

(f) *Transparency*

By 'transparency' I mean that the procedure for coming to self-ascribe does not involve any introspective experience *of* the properties of your conscious episode; it involves only enjoying that conscious episode (see sections 3.2.2, 3.4). This requirement is a basic commitment, which I earlier labelled (**N**), of the moderate epistemic view. The present chapter has been an attempt to show how the requirement can be met.

5.4 Non-stative contents

My presentation and defence of the moderate epistemic account for self-knowledge of content is now largely complete. There remains, however, the complication that I raised in 5.1 above, when I made a distinction between two conceptions of intentional content: the stative and semantic conceptions. In this section I will show that the moderate epistemic account can explain self-knowledge of conscious states and episodes that are representational (have intentional content) under either of these conceptions.

An experience or thought has content in the stative sense when it puts forward its content as true—it puts forward the state of affairs on which it is directed as actually obtaining. An experience or thought has content in the merely semantic sense when it represents some state of affairs (or object), but does not put forward that state of affairs as actually obtaining. Merely semantic representation can be neutral with respect to the truth of the content represented.

In the last few sections, I have talked as though experiences and thoughts always put forward their contents as true. Thus, the account I have offered applies to states and episodes that are representational in the stative sense. The question is whether (**W**) could also be a way of coming to know about contents represented in the merely semantic sense. We can break that question down into three further questions, each of which represents a challenge to the account I have offered. Firstly, will (**W**) be reliable in the merely semantic case? Secondly, will (**W**) be rational in the merely semantic case? Thirdly, will (**W**) be available at all in the merely semantic case? I will take the three questions in turn.

The answer to the first question is, clearly, yes. The account I offered in section 5.2.1 of the reliability of (**W**) made no appeal to the truth or apparent truth of the reason-providing self-ascribed content. It appealed only to what it is for a content to give the reason for which a subject judges. The claim was that the content must be entertained by the subject in coming to judge. It is immaterial whether it is true or represented as true.

One might object that a content *cannot* give the reason *for which* a subject judges unless it is taken as true. But this is to raise a question about the availability of (**W**), not its reliability. I will deal with that question below. The present point is that a content can give a reason for a self-ascriptive judgment, in the sense of supporting its truth, without itself being true or apparently true.

I turn to the second question: will (**W**) be rational in the merely semantic case? On the

account I offered, (**W**) is rational in the stative case in virtue of the subject's grasp, concerning the state of affairs of which she is aware, that it reflects not only how the world is, but her own cognitive perspective on the world. Can that same grasp, or an analogue of it, play the same role in the merely semantic case?

There is a good, if derivative, sense in which the merely semantic case involves occupying a cognitive perspective on a state of affairs of which you are aware. When you entertain a content, you consider the state of affairs in which the truth-conditions of that content are met. That is what it is to entertain a content. When you suppose that the earth will explode tomorrow, or when that proposition simply pops into your head, you are aware of the putative state of affairs of the earth's exploding tomorrow. To merely consider the content without considering the state of affairs in which its truth-conditions are met would be to think *about* the content. It would be to entertain some other, higher-order content, and thus to consider the state of affairs in which the truth-conditions of *that* content are met.

Thus, any conscious state or episode with content involves consideration or awareness of a state of affairs; the difference between stative and merely semantic is that between non-neutrality and neutrality with respect to whether that state of affairs actually obtains. In this sense, any conscious state or episode constitutes a cognitive perspective on a state of affairs, namely the state of affairs in which the truth-conditions for the content entertained are met. When you suppose that the earth will explode tomorrow, you occupy a cognitive perspective on the state of affairs in which the earth explodes tomorrow, or on the world in which that state of affairs obtains. To entertain a content is to occupy a cognitive perspective—even if it is not to assent to anything, or to find something seeming to be the case. In occupying such a perspective, you can do many of the things you could do if you judged that the earth will explode tomorrow—you can think through the consequences of its being the case that the earth will explode tomorrow, for example.[18] Admittedly, such non-stative representations do not form part of your overall outlook on the world—how you take the world to be. In that sense, they are not part of your own perspective on the world. But each such episode in thought constitutes *a* perspective on its subject-matter. It is a perspective that you occupy, even if it is, for you, only a suppositional, or imaginary, perspective.

18 This idea, that non-stative representation involves a sort of cognitive perspective derivative from that of stative representation, draws support from Spinoza's view (in *Ethics* Part 2 P49 CN), endorsed by Geach (1957), that thought is fundamentally assertoric, and judgment is thus the basic type of mental act. We could add to Geach's claim on the fundamental nature of judgment a corresponding claim about the realm of passive mental occurrences: the fundamental type of passive mental occurrence is a seeming, perceptual or otherwise, that something is the case. If this is right, then we can see why non-stative thoughts and experiences inherit a sort of derivative perspectival status from judgments and seemings.

If (**W**) is to be rational in such cases, the subject must grasp, concerning the state of affairs of which she is aware in this sense, that it reflects a perspective she is occupying. This grasp is arguably more sophisticated than the grasp that how things appear reflects your perspective on the world, since it involves a conception of someone's entertaining a content without there being any commitment to its truth. But it seems to me that any subject who is capable of non-stative representation, and who has mastered the first-/third-person distinction, will have such a grasp. To consider a proposition without commitment to its truth is to embrace the idea of a *putative* state of affairs independent of how the world *is* or presents itself as being. Together with the bare idea of a subject entertaining a content, and thus being aware of some state of affairs, this yields the idea of a subject entertaining a content without commitment to its truth—being aware of some putative state of affairs without being aware of its obtaining. Thus, a subject who can consider contents in a non-stative sense, and who has the capacities discussed in 5.2.3, will grasp, concerning the putative states of affairs of which she is aware, that they reflect a cognitive perspective she is occupying, where this perspective need not be one of commitment. This grasp is grounded in those capacities. It can thus, in the context of the hybrid view (Chapter 4), make rational the transition from a content not taken as true to a self-ascription of that content. (**W**) will be rational in the merely semantic case.

The third question is whether a content not taken to be true could give a reason *for which* a subject judges. A *prima facie* reason to think that it could is that we can and do make rational transitions from contents not endorsed as true. Consider hypothetical reasoning. A conditional proof requires a series of transitions from a content not taken as true, and one that may well strike the reasoner as clearly false. This sort of reasoning would not be possible if we were incapable of making rational transitions from contents, entertained but not taken as true, to judgments.

To answer this third question in the negative would be to claim that we can only make judgments on the basis of what we take to be, or what appear to be, facts. But what would be the basis for such a claim?

It might be defended on the grounds that the constitutive aim of judging is truth. But we can accept that, in judging, we aim at truth (the truth of the content judged), while denying that the reasons for which we judge are always taken to be truths. On the account I have offered, a subject grasps, of the object of her awareness, that it is something on which she has a cognitive perspective. This grasp doesn't depend on that awareness having a stative content. So doing something for the reason given by a content doesn't seem to depend on taking it as

true.

Another way to defend the claim that we can only make judgments on the basis of what we take to be, or what appear to be, facts, would be to appeal to the notion of a point of view. In Chapter 4 I said that a reason a subject has must enter into the subject's point of view. I glossed this by saying it must form part of the world as it is for the subject. Doesn't this suggest that the content that gives the reason for which a subject judges must present itself as being, or be taken to be, true of the world? What I suggested four paragraphs ago is that when a subject entertains a content she considers the state of affairs in which that content is true of the world, even if she doesn't take that state of affairs to obtain in fact. The content thus enters a point of view—not the subject's own committed point of view on the world, but a point of view the subject can occupy in a suppositional, imaginary, or other derivative way. There is, then, a sense in which any content a subject entertains is part of a point of view the subject occupies. This sense is derivative from the stronger notion of a point of view; it is an 'as if' sense. It is not an equivocation on that notion.

Is this weaker, derivative notion enough to allow for reasons for which a subject judges? The original notion of a point of view was introduced to capture *what makes sense to a subject*—what is intelligible to a subject, given the way the world strikes her. I have argued that a thought not endorsed as true can contribute to what makes sense to a subject who grasps an analogue of the first-/third-person distinction for states of affairs not taken as obtaining. So contents can give reasons for judging in virtue of entering a point of view in this derivative sense.

I conclude, then, that the moderate epistemic account applies to merely semantic as well as stative representations. This discussion has brought out, however, that its application to merely semantic representation is derivative, in a certain way, from its application to stative representation. The rationality of (**W**) in the merely semantic case depends on the sense in which entertaining a content without any commitment to its truth amounts to occupying a perspective on a putative state of affairs—a sense that is derivative from the sense in which stative representation involves occupying a perspective on an apparent state of affairs. In this respect, the fundamental case of self-knowledge is knowing how things seem to you to be, or how you take them to be.

5.5 Conclusion

In this chapter I have carried out a major part of the constructive task of this thesis. I have

shown how it is that, in enjoying a conscious state or episode with a particular content, you are warranted in self-ascribing that content. The claim of my moderate epistemic account has been that the content itself, and not the content of some higher-order or introspective awareness, is what gives you (provides you with access to) a reason to self-ascribe that content. I presented a way of coming to make self-ascriptions of content, (**W**), and argued that (**W**) is a reliable, available and rational way of coming to self-ascribe content. In 5.2.1 I argued that what is involved in judging *for* the reason given by a content guarantees that a self-ascription of that content, made by (**W**), is true. In 5.2.2 I argued that (**W**) is available, given the hybrid view of having a reason. In 5.2.3 I argued that any subject so much as capable of judging according to (**W**) will have a practical or implicit grasp that what she is aware of reflects not only how the world is or might be, but also her own cognitive perspective on that world. I argued that, in virtue of this practical or implicit grasp, judging according to (**W**) will be rational for such a subject.

That section relied on a broadly Kantian view of the subject's perspective on the world as essentially involving the potential for a conception of the subjective self. Although self-knowledge involves conceptual capacities that are relatively sophisticated, its roots lie deep in the nature of our engagement with the world. Self-knowledge should not be thought of as a turning-away-from the world, but as grounded in our engagement, as subjects, with the world.

All of this has been in the service of giving an epistemic account of the first and third components of self-ascriptions—the first-person component, and the content component. In Chapter 6 I will offer an epistemic account of the second component—the type component. I will then have given a full epistemic account of the warrant for the self-ascriptive judgments that are the topic of this thesis.

CHAPTER 6

KNOWLEDGE OF TYPE

Self-ascriptions have three components: the first-person component (ascribing first-personally to *yourself*), the type component (ascribing a certain type of conscious state or episode), and the content component (ascribing some particular content). In Chapter 5 we saw how, by entertaining some content C, you can have a reason to self-ascribe that content. That explains the first-person and content components of self-knowledge. The present chapter focuses on the remaining component—knowledge of the type of conscious mental state or episode you are enjoying.

We can knowledgeably self-ascribe a range of different types of conscious state and episode, including judgments, other occurrent thoughts, perceptual experiences, memory impressions and imaginings. Given that some of these types are mental actions and some are passive mental occurrences, there is no guarantee that we know about them all in exactly the same way. This chapter examines in detail two types of conscious episode: perceptual experience and judgment.

In 6.1 I will introduce the general claims of the moderate epistemic account for the type component of self-knowledge.

6.2 will deal with perceptual experience. The claim will be that there is a distinctive way in which states of affairs are given to a subject in perception, in virtue of which those states of affairs, so given, can rationalise not only perceptual judgments concerning those states of affairs themselves, but also self-ascriptions of perceiving. The warrant for such self-ascriptions, I will argue, is in important respects parasitic on the warrant for perceptual judgments. 6.2.1 will discuss this feature of perceptual experience, which I call "directness". 6.2.2 will explain how we come to self-ascribe perceptual experiences, and will show, by appealing to directness, why self-ascriptions made in that way will be knowledge. 6.2.3 will show why such self-ascriptions will have the distinctive features of self-knowledge.

6.3 will deal with judgment. Self-knowledge of judging, I will claim, is a species of self-knowledge of acting (6.3.1). In 6.3.2 and 6.3.3 I will argue that self-knowledge of judging, like self-knowledge of other types of actions, is grounded in *control* of action. Control of judging is a matter of aiming at truth in a certain way; you know that you are *judging* that *p*, when you do so, because you are aiming at truth and the truth-pertinent reasons governing

your inquiry yield the verdict that *p*. I will then deal with some objections to the account (6.3.4), before showing that it respects the specialness of self-knowledge of judging (6.3.5).

6.1 The moderate epistemic account for knowledge of type

The moderate epistemic account for self-knowledge of content, given in Chapter 5, was characterised by theses (**R**) and (**N**). There are two corresponding theses that characterise the approach of the moderate epistemic account for knowledge of type:

($\mathbf{R_T}$) A subject who enjoys a conscious state or episode of a particular type, with a particular content, thereby has a reason to make the self-ascriptive judgment that she is enjoying a state or episode of that type, and a self-ascription made for that reason can be knowledge.

($\mathbf{N_T}$) When a subject has a good reason for a self-ascription of type in virtue of enjoying a conscious state or episode of that type, her having that reason does not depend on experiential awareness *of* that conscious state or episode. It depends on what is involved in enjoying the conscious state or episode itself.

Again, the challenge is to explain how ($\mathbf{R_T}$) and ($\mathbf{N_T}$) could be true. When you enjoy a conscious state or episode with a particular content, you are aware of, or thinking about, some putative state of affairs, determined by that content. That state of affairs typically has nothing to do with your psychological state. How, then, could you thereby have a reason to self-ascribe any particular type of state or episode?

I will claim that subjects enjoying conscious states and episodes with the very same content, but of different types, thereby have reasons *for* different judgments. There are features of different types of state and episode that contribute to what their subjects have reasons for. These features, however, are not reasons that those subjects have, or reasons for which they judge. They make a difference to what it is rational for subjects to do, even though they are not introspectively experienced by subjects. As I pointed out in Chapter 4 (section 4.3.1), factors other than the reasons you have can contribute to what it is rational for you to do.[1]

1 In Lucy O'Brien's terms, the view I advocate is a 'non-content-based view' (O'Brien, 2007). I hold that the epistemic warrant for self-ascriptions is not explained *solely* by reference to the contents

Since the relevant features will differ from one type of state or episode to another, the explanation of how they rationalise self-ascriptions will differ accordingly. I will set out the specific accounts for perceptual experience and judgment.

6.2 Perceptual experience

How do you know you are perceiving (rather than imagining, remembering, judging, or whatever) something to be the case, when you are? According to the moderate epistemic account, in perceiving that *p* (say), you have a reason to judge "I *perceive* that *p*", and this reason is not furnished by any sort of experiential awareness *of* your perceptual experience, but simply by the experience itself.[2]

The defence of these claims will appeal to a feature of perceptual experience which I call "directness". I will discuss directness in the first subsection, 6.2.1. In 6.2.2 I will describe a procedure for self-ascribing perceptual experience, and argue, in the light of the preceding discussion, that such self-ascriptions will be knowledge. In 6.2.3 I will show that they will have the distinctive features of self-knowledge.

---

on which self-ascriptions are based. I agree with O'Brien that we must also appeal to the *conscious* occurrence of the episode in which the relevant content is entertained. I claim, further, that, to explain knowledge of type, we must appeal to distinctive features of different episode- and state-types, features that are not explained solely in terms of consciousness.

2 Similar claims can be found in Peacocke's paper "Another I" (Peacocke, 2005). Peacocke deals with the determinate modality of vision (with self-ascriptions of seeing-that) rather than with the determinable, perceptual experience generally (self-ascriptions of perceiving-that). It may be that we learn to self-ascribe determinates before we come to learn the concept for the determinable—i.e. it may be that we know propositions of the form "I see that *p*" and "I hear that *p*" before we know propositions of the form "I perceive that *p*". What's more, even when we possess the concept of perception, we will tend to make the more determinate judgment or assertion. From the more determinate judgment, the less determinate one can be inferred. I assume that there is nevertheless something general to be said about you know you are perceiving, when you are—even though you usually know more than that. It seems that there must be something to be said about it that does not presuppose an account of how you know you are seeing that *p*, hearing that *p*, or whatever. A full account of how you know that you are perceiving in a particular modality will presumably draw on the account of how you know you are perceiving (and not, for example, engaging in sensory imagination) at all, and presumably that element will be common across modalities. Let me emphasise again that my approach is to try to bring out the nature of the epistemic warrants that self-knowing subjects have, rather than to capture the judgments that they in fact tend to make.
A further difference between Peacocke's paper and the present chapter is that Peacocke does not address the question which drives my inquiry: how, given the nature of reasons, could the claim that experiences give access to reasons for self-ascriptions of perceiving be true?

6.2.1 The directness of perceptual experience

There is a feature of perceptual experience generally, that contributes to what perceptual experiences give their subjects reasons for, but that is not part of what is experienc*ed* in such experiences, and is not a reason that such subjects have. So I will be arguing in this subsection. I will call the feature "directness". The evidence for the existence of directness is phenomenological; the evidence that directness contributes to what subjects have reasons for without itself being a reason derives from what is required for a plausible epistemology of perceptual judgment.

Let me first put forward the phenomenological evidence for the existence of directness. This evidence also helps us get a grip on what directness *is*. Phenomenological reflection suggests that perceptual experiences, in all modalities, stand apart from all other conscious states and episodes, in constituting a uniquely immediate, rich mode of engagement with the environment. John Searle, considering the case of vision, famously remarks:

> "If, for example, I see a yellow station wagon in front of me, the experience I have is directly of the object. It doesn't just "represent" the object, it provides direct access to it. The experience has a kind of directness, immediacy and involuntariness which is not shared by a belief which I might have about the object in its absence." (Searle, 1983, pp. 45-46.)

Similarly, here is Scott Sturgeon:

> "your visual experience will place a moving rock before the mind in a uniquely vivid way. Its phenomenology will be as if a scene is made manifest to you. … Visual phenomenology makes it for a subject as if a scene is simply presented." (Sturgeon, 2000, p. 9.)

In these passages Searle and Sturgeon consider visual experience in particular, and Searle contrasts it with belief. But the point is not limited to vision, nor to belief. Perceptual experiences[3] in all modalities have that combination of what Searle calls "directness, immediacy and involuntariness"; when you touch a surface, or taste a flavour, you are engaging with the object of your experience in a way that having a belief about it does not involve—and nor does thinking about that object, imagining it, or even remembering experiencing it. Only actually experiencing it does.

3 Here I count illusory and hallucinatory experiences as perceptual. If one objected to this, one could say instead that only perceptual experiences and apparent perceptual experiences have the combination of features Searle mentions.

As the quoted passages indicate, what I am calling directness is a multifaceted feature of experience. One aspect of it is the involuntariness of experience. A second is that, in experience, it seems to the subject that things are as the experience represents them to be—the state of affairs on which the experience is directed seems to the subject to obtain. Thirdly, in perceptual experience the constituents of the state of affairs that strike the subject as obtaining are *manifest* to the subject in a unique way. When you perceive that the foliage is yellow, the leaves and their colour are present, are manifest, to you. And fourthly, the state of affairs is present to you in all its detail, in a certain sense. Although a single or momentary experience will represent only some of the detail in a scene or state of affairs, it will also typically present the scene or state of affairs as involving further (perhaps unlimited) detail, accessible by further potential experiences. When you look at the foliage, you may not be able to see the exact shading of a particular leaf, but that property is, in an important sense, present to you nevertheless. Compare imagining or remembering a sound, taste or visual scene. These do not provide the same sense of there being further detail to access, the sense of potential for further discovery.

These aspects of directness are not just more things that seem to you to be the case when you have a perceptual experience. The situation is not that: it seems to you that the foliage is yellow *and* that that state of affairs is manifest to you in all its detail, etc.. Directness is not captured by adding something to the representational *content* of the experience. Any such additional content could feature equally in an apparent memory, a judgment, or a propositional seeming. You could have a thought with the content that the foliage is yellow and that that state of affairs is manifest to you in all its detail. This would not be the same as an experience as of the foliage being yellow.

Directness is a *way in which states of affairs are given* in (and only in) perceptual experience. When you enjoy a perceptual experience with a certain representational content, there is a way that the world seems to be to you—there is a state of affairs that seems to you to obtain, in virtue of that content. For example, it seems to you that certain foliage is yellow. But there is also a way in which it seems to you that the world is that way. This latter way is what distinguishes a perceptual experience as of the foliage's being yellow from any other sort of conscious state or episode that represents the foliage as yellow—that has the same representational content and thus represents the same state of affairs as obtaining. Thus, directness is not part of the representational *content* of experience, in the sense of being part

of what strikes you as true; it is the *way* in which what strikes you as true is given to you.[4]

Let me make one clarification. It would be a mistake to say that the directness of experience is a way in which experience is given to its subject. Rather, directness is a way in which a state of affairs is given to a subject, through experience. Experiences themselves are not given to their subjects in the same sense that objects and properties, or states of affairs, are. To say that experiences are given in that sense would be to say that a subject who enjoys an experience is also thereby conscious *of* the experience, in the sense that I have objected to (see section 3.3).

That is all I will say to characterise and establish the existence of directness. I will now argue that the directness of an experience makes a difference to what it is rational for the subject to do, but does not do so by being a reason that the subject has.

Perceptual experiences give reasons for judgments. A perceptual experience as of the foliage being yellow can give you a reason for the judgment that the foliage is yellow. That the experience gives you a reason for that particular judgment is partly determined by its representational content—that the foliage is a certain shade. But your having a reason for that judgment also depends on that content's being the content of a perceptual experience, and not of some other type of state or episode. You would have no reason for the judgment if you *imagined* the foliage being that shade, or if the thought simply occurred to you that the foliage is that shade. Reasons for perceptual judgments depend, in part, on the nature of perceptual experiences themselves, and not merely on their representational contents.[5]

Thus, if you are rationally to exploit the reasons provided by perceptual experiences, you must be rationally sensitive to when contents feature in perceptual experiences, and when they don't. You must be prepared to base judgments on perceptual experiences, but not on imaginings. You must therefore be sensitive, in your perceptual judgments, to features that

---

4 There may be contents that are made available only by perceptual experiences—demonstrative contents, for example—and that therefore cannot even be entertained in the absence of the appropriate experience. Might directness just consist in the entertaining of such contents? This is the view of Brewer (1999). However, even if a certain content is available to you only *when* you are entertaining a perceptual experience, it does not follow that that content can't be entertained in a distinct conscious episode or state. You might, while enjoying a perceptual experience, simultaneously *imagine* of an object you are experiencing that *that* object—thought about demonstratively, exploiting your experience—is *thus*—picking out demonstratively a property that some *other* object you are experiencing has (see Martin, 2001). This episode of imagining will involve contents that are made available only by a perceptual experience, but it will not have the directness of perceptual experience. Thus, directness does not consist in entertaining such contents.

5 Here I take it that perceptual experience, as a type, isn't individuated *merely* by the types of contents it has. See above, note 4.

distinguish perceptual experiences from other types of conscious state and episode. You must be sensitive to the directness of experience.

Since you (presumably) *do* rationally base judgments on perceptual experiences, you *are* sensitive to the directness of experience. Directness makes a difference, for you, to what it is rational for you to do. But when you make a perceptual judgment, the reason for your judgment is not *that* you are having a certain experience, or *that* you are being given some state of affairs directly. The reason for your judgment is the state of affairs disclosed by your experience—that the foliage is a certain shade, say. Perceptual judgments are based on what appears to be the case in experience, not on self-ascriptions of experience or on introspective experiences *of* experiences. A self-ascription or introspective experience of an experience would warrant a perceptual judgment only in combination with some further belief that the experience is veridical. But you do not typically have any basis for such a belief that is independent of the experience itself. So self-ascriptions or introspection of experiences, and their directness, do not provide a route to warranted perceptual judgments. It is experiences themselves that warrant perceptual judgments, by providing access, through their *contents*, to reasons. Your having a reason for a perceptual judgment depends on your experience in fact being direct, but not on your experiencing that directness, or on directness being your reason.

Thus, the directness of an experience makes a difference to what the content of that experience gives a reason to do, without doing so by itself being a reason you have.

That completes my discussion of the directness of experience. I will now use that notion to address self-knowledge of perceptual experience.

6.2.2 Self-ascriptions of perceptual experience

As ever, I take it that an explanation of why judgments in some range constitute knowledge consists in a description of a way of coming to judge, and an explanation of why coming to judge in that way will be reliable and rational.[6] That is what I will offer in this subsection.

I do not wish to take sides on the question of the nature of perceptual content, so I formulate the way of coming to self-ascribe perceptual experience neutrally between conceptual and nonconceptual content:

6 That way of coming to judge must also be available to subjects (see 5.1). I take it that there is no difficulty with that requirement in this case.

($\mathbf{W_E}$) For the reason given by the content C, where C stands in the appropriate rational relation to *p*, and that reason is given in the appropriate way, the subject judges "I perceive that *p*".

The 'appropriate rational relation' here is the relation that C and *p* stand in when the correctness of C guarantees the truth of *p*, and grasp of the content *p* requires willingness normally to judge that *p* in response to experiences with the content C.[7] The simplest case is that in which C and *p* are identical: having an experience as of *p*'s being true, you judge "I perceive that *p*". A less simple case is that in which you experience the foliage as being a certain nonconceptually given shade, which you know is a shade of yellow, and thus judge "I perceive that the foliage is yellow". The 'appropriate way' in which the subject's reason is given is of course the way associated with the directness of perceptual experience; the reason will not be given in the appropriate way if C features in an apparent memory or an episode of visual imagining.

($\mathbf{W_E}$) captures the moderate epistemic account's claim that the identity of the reason a subject has, in virtue of enjoying a perceptual experience, is fixed by the content of that perceptual experience. It also captures my claim, defended in the last section, that what a reason, given by some content, rationalises for a subject depends in part on the way the state of affairs that content puts her in touch with is given to the subject.

Why will self-ascriptions made according to ($\mathbf{W_E}$) be reliable? A subject can judge for the reason given by a particular content only if she consciously entertains that content—so I argued in Chapter 4 (section 4.4). That the subject who follows ($\mathbf{W_E}$) is entertaining the reason-giving content C in a perceptual experience, or an apparent perceptual experience, is guaranteed by the way in which that reason is given, for its being given in the appropriate way is a matter of the directness of perceptual experience. So if a subject follows the procedure described, her self-ascription will be true, as long as she is enjoying a genuine, veridical perceptual experience, and not merely an apparent or illusory one.

That is, ($\mathbf{W_E}$) will be reliable, in so far as the subject's perceptual experience is reliably veridical. Is this sufficiently reliable for judgments based on veridical experiences to, potentially, be knowledge? Yes. For consider: perceptual judgments made by taking contents of experiences at face value are typically, when true and undefeated, knowledge. Their being so depends on the reliable veridicality of perceptual experience. If perceptual experience is

7 Here I draw on Peacocke's (1992) account of perceptual content.

reliable enough for true, undefeated perceptual judgments to count as knowledge, it is reliable enough for true self-ascriptions of perceiving-that to count as knowledge (provided the other conditions on knowledge are met).

What about the rationality of ($\mathbf{W_E}$)? The question here is why the directness of experience should contribute to making rational a judgment about perceptual experience. What explains your willingness to make a judgment about your own perception on the basis of an experience of some independent state of affairs?

I argued in Chapter 5 that self-ascriptions of content involve a recognition, concerning what you are aware of, that it reflects both how the world is, and your own perspective on the world—I called this "grasp of the first-/third-person distinction". Any subject capable of following ($\mathbf{W_E}$) will grasp the first-/third-person distinction. Any subject capable of following ($\mathbf{W_E}$) must also possess the concept of perceptual experience.[8] If you are such a subject, you will thus have a conception of an objective world, in which states of affairs obtain, and of yourself as a subject who can latch onto such states of affairs by means of perception. You will appreciate that the obtaining of a state of affairs is independent of the perception of it. You will appreciate that a perceptual experience is an episode in which some state of affairs is presented, is made manifest, to a subject.

When you enjoy a perceptual experience of the foliage being yellow, you will know that you are enjoying a conscious episode with the content that the foliage is yellow (as I argued in Chapter 5). What's more, you will know that that state of affairs, the foliage being yellow, actually obtains—for it is apparent to you, in having the experience, that that state of affairs obtains. Thus, you will know that you are presently aware of the obtaining of the independent state of affairs of the foliage being yellow. The leaves and their colour will also be *present*, be *manifest*, to you, in the uniquely immediate way that I have called "directness". Given all of this, and given that you know the truth-conditions of the content "I perceive that the foliage is yellow", it will seem to you that those conditions obtain. Thus, you will be willing to judge that content, and your doing so will be rational.

Once again, the role of directness in this explanation makes no appeal to introspection. It is the way that states of affairs are given in experience, not the way in which experience is

8 There may be subjects who can self-ascribe seeing that *p*, for example, before they possess the general concept of perceptual experience. Of course, such subjects will possess the concept of seeing-that, so will have some grasp, perhaps tacit, of the general notion. I take it that self-ascriptions of seeing-that nevertheless rely, in part, on the capacity to tell when an episode is perceptual, rather than, say, imaginative (see note 2). This capacity, I would suggest, depends on directness.

given, that explains the rationality of ($\mathbf{W_E}$). And your willingness to self-ascribe in that way, when you are enjoying a perceptual experience, depends on certain fundamental cognitive and conceptual capacities, rather than on any sort of introspective experience (see Chapter 5).

It might be objected to all of this that a perceptual experience could at most give you a reason to judge "It perceptually appears to me that *p*", rather than "I perceive that *p*". After all, you have no way of telling from the inside whether your apparent perceptual experience is veridical.

But this objection is wrongheaded. Consider again the perceptual judgment that *p*. This judgment is made rational simply by your enjoying a perceptual experience as of *p*'s being the case, in the absence of defeaters. Having a reason, given by perceptual experience, to judge that *p* does not require first establishing that your experience is veridical. Rather, you have a default entitlement to presuppose that you are perceiving properly. Since you have this entitlement, a perceptual experience as of *p*'s being the case gives you a defeasible reason to judge that *p*, and also a defeasible reason to make the factive self-ascriptive judgment "I perceive that *p*". You can be rational in making the factive judgment, without having to make independent checks that you are perceiving properly.

Note also that, in making the factive judgment, you will be doing what is rational from your own point of view. The factive implication of the judgment—the implication that *p*—is something that strikes you as true when you perceive that *p*. You will not be going wrong by your own lights in making the factive self-ascription "I perceive that *p*".

Of course, you *would* be going wrong by your own lights if you judged "I perceive that *p*" but you were *not* prepared to endorse the content of your perceptual experience, either because you doubted that you were perceiving properly or because you suspected that ~*p*. But if you grasp the self-ascriptive content "I perceive that *p*", you appreciate its factivity, and so you will not endorse it if you are not prepared to take it that *p*. (You will, in such cases, be prepared to make the non-factive judgment "It perceptually appears to me that *p*".) Your grasp of the content ensures that you will be willing to make the self-ascriptive judgment precisely when you are willing to make the perceptual judgment that *p*.

Correspondingly, the warrant for the self-ascriptive judgment ("I perceive that *p*") will track precisely the warrant for the perceptual judgment ("*p*"). Each warrant has the same source: the perceptual experience as of *p*'s being the case. Each warrant rests on the default entitlement to presuppose that you are perceiving properly. Each one can be defeated by a

reason to doubt that you are perceiving properly, or a reason to doubt that *p*.[9]

The upshot of this account is that you will be warranted in judging, and, insofar as you are rational, willing to judge, "I perceive that *p*", precisely when perception warrants you in judging, and makes you willing to judge, that *p*. The account made no appeal to an internalist awareness of your warrant to judge that *p* or an introspective awareness of your willingness to judge that *p*. The warrant for the factive self-ascription depends on *having* a warrant for the perceptual judgment, but not on having an independent warrant to judge that you have that warrant. The willingness to make the self-ascriptive judgment depends on the same conditions as the willingness to make the perceptual judgment, but not on awareness of that latter willingness.

That completes my account of how you come to know that you are perceiving something to be the case, when you are. It remains to show that knowledge acquired in this way will have the distinctive features of self-knowledge.

6.2.3 The specialness of self-knowledge of perceptual experience

(a) *Security*

Security is the feature that self-ascriptions are extraordinarily safe—it is modally difficult for them to go wrong.

Self-ascriptions made according to ($\mathbf{W_E}$) will not easily go wrong in respect of the type of conscious episode you are enjoying. The procedure for coming to self-ascribe depends crucially on the directness of experience—a feature that is, I argued, unique to perception (and apparent perception). What's more, the explanation of the reliability of the procedure did not appeal to any contingent, causal connections, but only on what is involved in the procedure's being available to you at all.

Given its factivity, a self-ascription made according to ($\mathbf{W_E}$) will be secure only when your perception is veridical. Thus, the security of this procedure is a necessity that is contingent on the veridicality of experience. Self-ascriptions of the weaker, non-factive sort—"It

9 The view of perceptual warrant I am assuming is something like Pryor's 'dogmatism' (Pryor, 2000). It is the view that perceptual experiences themselves provide defeasible warrants for perceptual judgments, without those warrants relying on subjects having independent grounds to suppose that their experiences are veridical or reliable. I am suggesting that this account can also explain the warrants for self-ascriptions of perceiving.

perceptually appears to me that *p*"—when made by the same procedure, will be secure *simpliciter*.

(b) *Salience*

Salience is the feature that a subject will not easily be 'self-blind' with respect to any particular conscious state or episode.

Any perceptual experience with a particular content will make available ($\mathbf{W_E}$), for perceptual experiences by their nature give (defeasible) reasons for judgment. Thus, a subject who is conceptually equipped will be able to know, whenever she enjoys a perceptual experience and there are no defeaters, that she is perceiving something to be the case.

(c) *First-person privilege*

This is the feature that each person can come to know about his or her own conscious states and episodes in a way available only to him or her.

A perceptual experience directly offers reasons for judgment only to its subject. ($\mathbf{W_E}$) is available only to the subject of the reason-giving experience. You could not come to know, in that way, the type of any conscious episode of which you were not the subject.

(d) *Authority*

Authority has two aspects. It is typically inappropriate to challenge or gainsay a self-ascriptive judgment. And it is typically inappropriate to demand justification for such a judgment.

I argued in section 5.3 that self-knowledge of content is authoritative *because* it is secure and first-person privileged. Challenging a self-ascription is typically inappropriate because self-ascriptions are secure and an interlocutor has no comparable way of coming to make a judgment about the matter. This point applies as much to self-knowledge of perceptual experience as to self-knowledge of content. Demanding a justification for a self-ascription is inappropriate because the reason for which you make the self-ascriptive judgment is a reason only from your own point of view, and so cannot play a role in the social practice of *justifying*. If asked how you know you are *perceiving* that the foliage is yellow, you could say, "Because it *is* yellow, and it really appears that way to me", or, if pressed, "Because I *am*

perceiving it". But these justifications would not satisfy a third party.

(e) *Immediacy*

Self-ascriptions do not involve acquiring or attending to grounds, or any other way of finding out about your own conscious states and episodes.

($\mathbf{W_E}$) does not involve acquiring or attending to grounds that warrant the type-component of the self-ascription. Just having a perceptual experience puts you in a position to self-ascribe it—no further effort or thought is required for you to be able to tell that it is a perceptual experience. The resulting judgment cannot be characterised as something you have found out.

(f) *Transparency*

This is the feature that many self-ascriptions are arrived at by attending outwardly to the objects of experience (or thought), not by reliance on a representation of your own mind.

There is no role in the account I offered for introspective experience. You come to know that you are perceiving that *p* by perceiving how things are in the world around you.

(g) *Commitment*

Certain self-ascriptions commit their subjects to the contents they self-ascribe (section 3.2.3).

This is true of self-ascriptions of perception: when you judge "I perceive that *p*" you thereby commit yourself to the truth of *p*. However, the explanation is relatively simple in this case. Such a self-ascription (unlike the ones discussed in Chapter 3) is factive, and the subject who is conceptually capable of issuing it appreciates its factivity, so will not issue it unless prepared to commit to *p*.

6.3 Judging

I want now to offer my account of self-knowledge of judging. My question for the rest of the chapter is: when you judge that *p*, how do you know that you are *judging*, not, guessing, supposing or doubting, that *p*? I will argue that judgments are mental actions performed for reasons, and that it is in virtue of their nature as actions that a judging subject has a reason to

judge that she is judging. I will start, in subsection 6.3.1, with the argument that judging is an action, and that self-knowledge of judging is a species of self-knowledge of acting. In 6.3.2 I will outline "the control view" of self-knowledge of acting, according to which, when you engage in an activity, it is constitutive of your controlling that activity that you have practical awareness of the goal of the activity and of the particular acts you are performing in pursuit of that goal. In 6.3.3 I will apply the control view to self-knowledge of judging. When judging, your goal is truth (or the truth with respect to some matter), and you delegate control of particular acts of judgment to the outcome of your reason-guided inquiry into what is the case; so you know that you are judging, when you are, because you have a practical awareness of your goal and you know the outcome of your inquiry. In 6.3.4 I will deal with some objections, before, in 6.3.5, showing that this account captures the specialness of self-knowledge of judging.

6.3.1 Judgment, action and self-knowledge

This section will make two claims. First, judging is a mental action,[10] like visualising the Taj Mahal, calculating the square root of 216, directing your attention to some pressing problem, or making a decision to take a day off. Second, self-knowledge of judging is a species of self-knowledge of acting. That is, self-knowledge that you are *judging*, rather than supposing, doubting, or merely having it appear to you, that *p*, is a species of the same epistemic genus that includes self-knowledge that you are raising your arm when you reach up to grab something, and self-knowledge that you are walking home when you are doing so.

By 'judgment' I mean a type of occurrent event in consciousness, a conscious episode, in which a thinker takes it that something is the case. Judgments have propositional contents that they represent as true. It is in the nature of judgment that an episode of judging that *p* tends to constitute the acquisition or retention of the belief that *p* (but beliefs can also be acquired with no conscious episode of judging).

There are certain features of judgment that strongly suggest it is an action.

First, judging has a goal. Roughly, the goal of judging is to judge truly and avoid judging falsely. Judgments are also constituents of activities governed by more specific goals, such as the goal of coming to a correct verdict on some matter. In such cases the subject's thinking is guided by that more specific goal.

10 This view is defended by Geach (1957). It can also be found in Soteriou (2005) and Peacocke (1999).

Second, you are responsible for your judgments. There can be reasons for and against making them, and when you judge you are responsible to those reasons. It is up to you to keep your judgment in line with what reasons recommend.

Third, judgments are paradigmatically made *for* reasons, good or otherwise. The attribution of judgment (and belief and other propositional attitudes) to a subject is normatively constrained: it makes sense only when the subject's thought and behaviour can be seen as responsive to reasons. A capacity for responsiveness to reasons is not merely a capacity to do what reasons recommend; it is a capacity to do things for those reasons. A subject who judges, judges for reasons.

Fourth, judgments are committing. They have specific normative and causal implications for the subject's attitudes and future actions. If you judge that the leaves are yellow, you are committed to their being yellow in a way that merely experiencing them as yellow does not entail. You are rationally bound to act and judge in certain ways when the colour of the leaves is relevant.

Each of these features provides evidence that judging is an action; together they constitute a very strong cumulative case. First, goal-directedness is characteristic of action, as opposed to mere behaviour. Second, responsibility can be attributed only to agents, and it can be attributed to them primarily with respect to their actions (including acts of omission).[11] Third, doing something for a reason is always an exercise of agency. It does not make sense, except perhaps metaphorically, to offer a reason-giving explanation of something that is not an exercise of agency: there is a deep connection between these two notions. Fourth, it is unclear how a judgment could constitute the undertaking, by the subject, of a commitment, if it were not an act performed by the subject. A perceptual experience or other seeming cannot by itself commit the subject to its content, even though it is non-neutral with respect to the truth of its content. The subject herself becomes committed to that content when she undertakes a commitment by performing an act of endorsing the content.[12] No amount of passively undergoing experiences could be sufficient for the undertaking of a commitment.

---

11 We are responsible for certain of our attitudes, such as beliefs, which are not themselves actions. But we are responsible for them in the sense that we are responsible for acquiring, maintaining and rejecting them, all of which (I would argue) can be actions.

12 Again, beliefs are committing, and you can acquire a belief without performing a mental act of judgment. When you observe a scene, you form many beliefs 'automatically', by accepting what you see without any act of judgment. But if the acquisition of the belief *is* a conscious episode, that episode must be an act—for how else could it be the conscious entering of a committing state? Beliefs that are not manifested in consciousness in this way nevertheless owe their nature, in part, to the conscious acts that *would* manifest them.

In sum, it is hard to see how we could make sense of these features of judgment without assuming that judgments are actions.

It might be objected that judgments cannot be actions in any strict sense, because they are not voluntary. You do not and cannot judge that *p* simply by deciding to do so. And, if you are considering whether *p*, whether you judge that *p* is not up to you, but is determined by the strength of the reasons you perceive for thinking that *p*; once you have a determinate sense (say) that the reasons decisively establish that *p*, the judgment is effectively made for you. If you see the yellow leaves in plain view, with no apparent reasons for doubt, you cannot refrain from judging that they are yellow, if you consider it.

There are two points in response to this. The first is that an act of judging that *p* often is voluntary, under some description other than 'judging that *p*'. Although it is not up to you which particular contents you judge true, the matters on which you come to make judgments are largely under your control. You can direct your thought: you can attend to particular subject-matters. If you see apparently yellow foliage, it will often be of no interest to you to ascertain whether it is really yellow, and you will not try to make a judgment on the matter. However, if you are surveying the colours of the local foliage, you will voluntarily try to determine whether the foliage really is yellow. Doing so is an activity, governed by a goal; the judgment you come to is a constituent act of that activity The act of judging *is* your coming to a verdict on whether the foliage is yellow—here we have one and the same act under two different descriptions. The act is voluntary under the description 'coming to a verdict on whether the foliage is yellow'.

Peacocke describes well the nature of control in thought:

> "When a thinker is engaged in directed thinking, he is in effect selecting a certain kind of path through the space of possible thoughts—thought contents—available to him. There is not selection for particular thoughts, of course: that would involve the rejected view that there are intentions to think certain particular thoughts. But there is selection of a certain kind of thought, given by the content of the thinker's aim in thought. Without such selection, human thought would be chaotic, at the mercy of associational connections not necessarily at all pertinent to the thinker's current goals." (Peacocke, 1998, p. 70.)

The second point in response to the objection is to do with the sense in which judgments are non-voluntary, and brings out some deeper points about judgment and agency. Judgments are subject to the authority of reasons: the properly functioning judger will judge in the way recommended by the reasons as she sees them (of course, she may misperceive the relevant

reasons). When her inquiry into whether *p* yields the verdict that *p*, she will judge that *p*. But the authority of reasons is not a matter of *compulsion*, or of *reflex*. Reasons do not overpower your will. It is a matter of *reason*. The non-voluntariness of judgment still leaves room for the subject's careful reflection on what is the case, and for the subject to be responsible for the outcome of that reflection. It is not at all like the non-voluntariness of falling over when pushed.

Importantly, non-voluntariness also leaves room for the subject to withhold judgment in the content *p*, even when her inquiry has yielded the verdict that *p*—for it is always open to a subject to reconsider her verdict, to ask whether her inquiry was competently carried out, and thus effectively to reopen the question whether *p*. When she does so, she makes her sense of the pertinent reasons indeterminate again: she doubts whether she was seeing the reasons properly when inquiring into whether *p*, and it is an open question for her whether the pertinent reasons really do determine the verdict that *p*. Reopening this question and withholding judgment is something that the subject can do voluntarily. Such a withholding of judgment may be motivated by a genuine doubt over whether the inquiry into whether *p* yielded the correct outcome. But it also may be non-epistemically motivated: the subject may have some other reason not to want to accept that *p*. Thus, the non-voluntariness of judgment allows for a significant degree of autonomy.

These considerations shed light on the connection between the goal of judgment, the role of reasons in judging, and the non-voluntariness of judging. The goal of judgment is, as a constitutive matter, truth. It is thus constitutive of judging that you are guided by those considerations that strike you as pertinent to the truth-values of contents. Those considerations are your reasons (not all your reasons, in this sense, will be genuine reasons; see section 4.1). To judge that *p* when the truth-pertinent reasons do not yield the verdict that *p* for you, would be to fail to aim at the truth; but it is constitutive of judging that you aim at the truth; that is why it is impossible simply to judge that *p* without regard for the reasons pertinent to whether *p*. Thus, the non-voluntariness of judgment is part of what it *is* for judging constitutively to aim at truth. The non-voluntariness of judgment, far from arguing against its being an action, is a central aspect of its nature as an action.

But now it might seem that I have made too many concessions to non-voluntarism. For it might be asked: isn't there a phenomenon of akratic judgment? That is, don't we sometimes judge to be true contents that we are aware there is not sufficient reason to think true? Thus, haven't I overstated the extent of non-voluntarism?

I reply that the phenomenon of akratic judgment, when correctly described, is not

incompatible with my claims. It is true that people make rash judgments, and sometimes do so by ignoring reasons against judging, which they are in some sense aware of. But even in these cases, I claim, the subject must get herself to occupy a perspective from which the reasons pertinent to whether *p* (those she is taking into account) really do yield the verdict that *p*. That is, in order to judge that *p*, she must first manipulate her sense of the reasons pertinent to whether *p*. In such a case, the subject's process of reaching a verdict may be in tension with other things the subject is aware of; she may be irrational in reaching it. It will usually involve a certain sort of wilful blindness and self-deception. The subject will perhaps tell herself that certain reasons are more powerful than they are, and avoid taking others into account. Goals other than truth will be involved in this process of reaching a verdict, although the subject will have to deny that she is pursuing anything other than truth. It is a vexed question how such self-deception and irrationality is possible. But in any case, so-called akratic judgment does not show that the act of judging itself can be voluntary in any sense that is contrary to my discussion. The akrasia lies in the setting up of the inquiry into whether *p*, not in the act of judging itself.

I have argued that judging is a type of action, and made some claims about its character as an action. What is the relevance of all of this to self-knowledge of judging?

If what I have said is correct, knowing that you are judging is knowing that you are performing a certain type of action. But I want to make a stronger claim: that self-knowledge of judging is a species of self-knowledge of action. This is a claim about the sort of knowledge you have when you know you are judging, not merely about the content of that knowledge. It is the claim that the knowledge you have is of the same *epistemic* genus as self-knowledge of physical actions. By this I mean that the way in which you come to know you are judging falls under the same type as the ways in which you come to know you are performing physical actions; there is significant kinship between the epistemic accounts of why self-ascriptions in each case constitute knowledge; and accordingly the two species of self-knowledge share a number of distinctive epistemic features.[13]

There are *prima facie* reasons to believe that self-knowledge of judging is not only knowledge of something that is in fact action, but also of the same epistemic genus as self-knowledge of physical action.

13 To say that self-knowledge of judging is a species of self-knowledge of action is to say that it is a species of the same genus under which self-knowledge of physical actions falls, not that it is just the same as self-knowledge of physical actions. It may be an irreducible species in itself, requiring its own treatment.

One reason is that it is knowledge of what you are doing, that you get just by doing it. When you make a judgment, nothing further is required, it seems, in order for you to know you have judged. You do not, as I have argued (Chapter 3), have to introspect your mental events or states. The same seems to be true of certain physical actions. Simply raising your arm seems to put you in a position to know you are doing so. You do not have to make some independent check that that is what you have done. The exercise of agency seems to be sufficient, *ceteris paribus*, for putting you in a position to know you are judging, and to know you are raising your arm.

Secondly, the exercise of agency plays a necessary role in your coming to self-ascribe judgment or physical action. In each case the ordinary first-person way of coming to make the self-ascription depends crucially on agency. Merely having it appear to you that *p*, no matter how forcefully, will not be sufficient for willingness to self-ascribe judging. In the case of raising your arm, you will not be prepared to self-ascribe the act of raising your arm if your arm is propelled upwards, outside your control; nor will you be prepared to self-ascribe the act if you merely *intend* to raise your arm, but make no effort to do so. In each case, the exercise of agency is a crucial element in the way of coming to self-ascribe.

Thus it is *prima facie* very plausible that self-knowledge of judging is a species of self-knowledge of action. I will not attempt directly to defend that hypothesis any further. Rather, I will offer an account of self-knowledge of judging in which the hypothesis is a central pillar. If an attractive account of this sort can be developed, that will constitute further reason to believe the hypothesis.

6.3.2 Self-knowledge of acting: the control view

In this section I will describe two traditional types of view of self-knowledge of acting, and show that neither view can account for self-knowledge of judging. I will then outline an alternative view—the control view—which I can go on to apply to judging. First, I must say a few things to clarify the phenomenon of self-knowledge of acting.

When you act, you typically know what you are doing, under certain descriptions. You know, as you walk home, that that is what you are doing. If you raise your arm in order to grab something, you know that you are raising your arm. You know these things 'from the inside', in a way that is unavailable to other people, and that does not depend on the evidence of what you do. You do not need to wait and see which route you take in order to know you are walking home; and your knowledge that you are raising your arm does not depend on

observing your arm rise.[14]

The knowledge I am concerned with is knowledge of the *content* of your action—knowledge of *what you are doing*. This is to be distinguished, firstly, from knowledge of ownership—knowledge that *you* are the agent of the action. I am concerned with how you know that it is Φing, rather than Ψing, that you are doing, rather than the question of how you know it is you, rather than someone else, that is Φing.[15] It is to be distinguished, secondly, from knowledge that you have succeeded in executing the action you are performing. That is, the question of how you know that it is Φing, rather than Ψing, that you are doing, is distinct from the question of how you know that you are in fact Φing and not merely trying to Φ. So, when I say that you know you are Φing, I mean that you know that Φing, rather than anything else, is what you are doing, or at least attempting to do.[16]

First-personal self-knowledge of acting does not extend to all descriptions of actions. When you intentionally walk home, you know first-personally that you are walking home and that you are moving your legs. But you may not know that you are taking 5,000 paces, or that you are walking due east. An account of self-knowledge of acting ought to give intuitively correct results about which descriptions are those under which you know first-personally what you are doing, and which are not. I leave aside the question of how to draw a principled line between the two categories.

Two quite different sorts of accounts of self-knowledge of acting have dominated the philosophical literature. I will call them, respectively, 'practical reasoning views' and 'experience-based views'.

According to practical reasoning views, the agent's knowledge of what she is doing is epistemically grounded in the practical reasoning, reasons or intentions that led her to do it.[17] It is by reasoning about what to do, or by forming an intention to do something, that you

14 Although it is not essential for my argument, and I won't defend it, it is notable that self-knowledge of action within a certain range seems to share with self-knowledge of conscious states and episodes the distinctive features of security, salience, first-person privilege, authority and immediacy.

15 Plausibly, the ordinary first-personal way of coming to know what you are doing is constitutively tied to the sense, or the fact, of ownership; so you could never know what you are doing in this way without knowing that it is you that is doing it. Nevertheless, the questions are distinct.

16 This is not to suggest that knowledge is contrastive: that you can know that *p* relative to the contrast-proposition *q*, but fail to know it relative to the contrast proposition *r* (for which view see Schaffer, 2005). Regardless of whether knowledge is contrastive, the question of *how* you know something *can* have a contrastive dimension, depending on the interest of the person asking it. We can, as philosophers, be interested in one aspect of your knowledge rather than another.

17 This sort of view was pioneered by Anscombe (1957). It has been carried on by, *inter alia*, Velleman (1989), Dunn (1998) and Moran (2001).

come to know what you will do—and that, when the time comes for you to do it, you know what you are doing. You know what you are doing because you know what you intend to do, or what you take yourself to have most reason to do, and because that is what you do.

Experience-based views hold that the agent's knowledge of what she is doing is epistemically grounded in an experience of acting in that way.[18] That is, when you intentionally Φ, you simultaneously enjoy an experience in virtue of which it seems to you that you are Φing, and you can come to know that you are Φing by taking the content of that experience at face value. The experience need not be characterised as a sort of perception, but there are epistemic parallels with perception. In each case, you come to judge by enjoying an experience in which something strikes you as being the case, and endorsing the content of that experience; and the warrant for the judgment consists, in part, in a warrant to rely on such experiences.

An experience-based view of self-knowledge of action should not be identified with the claim that the agent is typically aware of what she is doing when she acts. The experience-based view goes far beyond this claim: it makes a specific epistemic claim about how the agent knows what she is doing. It is important here to distinguish between being aware of something's being the case, and enjoying an experience as of its being the case. You are typically aware of what you are doing when you do it, in the sense that you know (perhaps without making any conscious judgment) what you are doing. The experienced-based view makes the substantive claim that this knowledge is based on a distinctive experience. To offer a rival account is not to deny that the agent is aware of what she is doing—it is to offer an alternative explanation of that fact.

Equally, a non-experience-based view can accept that there is a distinctive phenomenology that accompanies agency. The proponent of such a view will hold either that that phenomenology does not play a role in how the agent comes to know what she is doing (though its absence may be a defeater for a self-ascription of action), or that the phenomenology should be partly explained *by* the agent's knowledge of what she is doing.

Can either of these traditional views explain self-knowledge of judging? It seems not.

The practical reasoning view cannot explain how self-ascriptions of judging will track particular acts of judgment. Judgments are not typically intended, so a practical reasoning

18 See Peacocke (2003, 2007, forthcoming) for an experience-based view of self-knowledge of physical and mental actions. This sort of view is also prevalent in the psychological literature. See, for example, Frith (1992), and the more empirically-orientated contributions to Roessler and Eilan (2003).

view would have to appeal to *reasons* for judging, rather than intentions to judge. It would have to claim that you know you are judging that *p* because you are aware of reasons for judging (rather than, say, supposing) that *p*, rather than claiming that you know you are judging that *p* because you intend to judge (rather than to suppose) that *p*. At any given moment, you will typically have, and be aware of, reasons for many different judgments—judgments about conditions in your environment, about what you are doing, about the fact that you exist, and so on. But you do not in fact make all of these judgments. You may not make any of them. Whether you make a judgment, and which judgment you make, depends on your current goal or direction in thought, not just on what reasons for judging you are aware of. So, if self-ascriptions of judging were based on reasons for judging, they would not reliably track actual acts of judging. So this cannot be the correct account of self-knowledge of judging.

Experience-based views, on the other hand, are ruled out by the arguments of Chapter 3 (section 3.2). Experience-based views are introspectionist views. They hold that self-ascriptions of judging are based on experiences that present acts of judging to their subjects—introspective experiences. Introspectionist views, I argued, fail to capture what is distinctive about self-knowledge and about the subject's perspective on her own conscious judgments.

The correct account of self-knowledge of judging must capture what is right about each of these views, while avoiding what is problematic. Like experience-based views, it must offer an explanation of how self-ascriptions of judging track particular episodes of judging. Like practical reasoning views, it must respect the self-knowing subject's perspective on her own judgments.

I think that self-knowledge of at least certain actions is best understood by looking not at practical reasoning or at experience, but at *control*. And I think that a view based on control can extend to self-knowledge of judging.

By 'control' I have in mind a personal-level notion: what is sometimes called 'intentional control'. There is a thinner sense of ‘control’ in which any piece of behaviour guided by a representation of a goal and by monitoring of performance is controlled. But not every such piece of behaviour counts as something *you* control in the personal-level sense. For example, very fine-grained movements that are part of things you do intentionally (like minor adjustments of finger position when reaching to grasp something) are often initiated and guided by wholly subpersonal processes. These movements are controlled—that is why they are successful—but they are not under your personal-level, intentional control. On the other

hand, the larger action of which those movements are constituents (e.g. reaching to grasp that box) is something you do control. At this level, the action is something you initiate and guide.

One reason that control is an attractive resource to appeal to is that control is a feature not just of individual acts, but of *activities*. To be engaged in an activity is typically to exercise control. You exercise control over the activity; and part of doing that is controlling (selecting, initiating, guiding, intervening, ceasing) individual acts in light of the goal of the activity of which they are part. Epistemologists of action have tended to focus on self-knowledge of discrete individual acts, in isolation from any context, but in fact many of our acts are carried out as part of broader activities, and our perspective on them as agents takes in not just their intrinsic features as acts, but their roles in activities.

When you judge that *p*, that particular judgment is a constituent act of an activity in thought. That activity might be simply coming to a verdict on whether *p*. It is guided by the goal of ascertaining the *truth* with respect to *p*. Your judgment is a response to considerations pertinent to that goal. For example, you see apparently yellow foliage, but instead of automatically accepting what you see, you wonder whether it is really yellow. So you engage in the activity of coming to a verdict on whether it is yellow, perhaps by considering whether the appearance in this case can be trusted. Your activity has a goal—to come to a correct verdict on whether the foliage is that colour. Eventually you reach the verdict that it is indeed yellow, and thus your activity culminates in an act of judgment, the judgment that the foliage is yellow.

There are similar cases in the realm of physical action. Sometimes you are engaged in a goal-directed activity that involves individual acts performed in response to considerations that are pertinent to your goal. For example, suppose you are driving from Edinburgh to Glasgow. This is an activity guided by a goal—to arrive in Glasgow in a safe and timely manner, let's say. In performing this activity you perform many small acts in response to cues that are relevant, given the goal of your activity. You see a red light and you brake. Braking is itself something you control: *you* initiate the act of braking, guide the force with which you do it, and stop braking when appropriate. Controlling the individual act is part of controlling the activity, and you control the individual act in the light of the goal of your activity (you would use the brake less often and less forcefully if safety were not part of your goal).

Note that you will not run through any reasoning before braking. Nor need you form a prior intention to brake. You simply brake because, from your point of view, that's what the red

light tells you to do; you know what to do when there is a red light. Nevertheless, you have an agent's distinctive knowledge, from the inside, of what you are doing when you brake. You do not simply find yourself braking.

I want to suggest the outline of an account of self-knowledge of physical acts of this sort. The account will not carry over precisely to judgment, which I deal with below. But it suggests a way in which control can be central to self-knowledge, and points the way towards an account for judgment.

Take a controlled activity A, guided by a goal G, and an act of Φing that is a constituent of A, such that the agent's control of A involves selection and control of the act of Φing. I claim that:

(**1**) It is constitutive of G's guiding A that the agent have practical awareness of G.

And:

(**2**) It is constitutive of the agent's control that, when the agent Φs, she takes it that she is Φing.

Let me defend these claims, before saying how they can be recruited to explain self-knowledge of acting.

Why must the agent have practical awareness of her goal? Because if she did not, the goal would not be guiding her activity; it would not *be* the goal of the activity. A goal guides an activity via the agent's control, which in turn involves the agent's knowing what to do in the course of the activity—knowing which actions to select, for example. Thus, if your goal is a safe and punctual arrival in Glasgow, this involves your responding to cues, such as traffic lights, by selecting certain actions, such as braking, that will further your goal. But cues tell you what to do only *in the light of* the goal you are aiming at. A red light alone, in abstraction from a goal, doesn't tell you to do anything. You know what to do in response to a red light only in the light of your goal.[19] To respond to cues in the light of your goal requires having

19 Of course you need other information, about the rules of the road, etc., too. But even if we fill in all the factual information, we will not get a verdict as to what you should do without specifying the goal guiding your activity.

some practical awareness of your goal.[20] Thus, it is constitutive of an activity's being goal-directed and controlled that the agent have practical awareness of the goal.

What of (**2**)? Why must the agent take it that she is Φing, when she Φs as part of a controlled activity? Because failure to take it that she is Φing would undermine her control in two ways.

Firstly, it would undermine her control (guidance) over that particular act of Φing. When you Φ as part of an activity, you often need to exercise control over the execution of the action—over *how* you Φ—in order to fit it into the activity. Thus, you need to brake with the correct firmness, and stop braking at the right moment, if the act is to serve your goal. Control of a particular act, during its execution, presupposes that you are performing that act. You could not know to stop braking, or to brake harder or softer, if you did not take it that you were braking in the first place. And you could not control your braking if you had to wait and see what you did before you knew what you were doing: control guides actions; it doesn't follow them. So, in order to control a particular act, during its execution, you must take it that you are performing that act, and you must do so from the agent's perspective, not by awaiting evidence of what you do.[21]

Note that it would not be sufficient that it *seemed* to you that you were braking, for that still leaves it open for you whether you *are* braking, and to control your braking it must not be an open question for you whether that is what you are doing.[22] You must be *committed* to that being what you are doing.

Secondly, the agent's failure to take it that she is Φing, when she is, would undermine her control of the activity of which the act of Φing is a constituent. Control of an activity involves a certain amount of coordination. An uncoordinated activity is an uncontrolled activity. What the agent must do next often depends on what she is doing now, as well as on

---

20 The term 'practical awareness' is supposed to indicate a type of knowledge that need not involve propositional attitudes. See section 5.3.

21 A similar argument can be found in the work of Lucy O'Brien:

> "If what I am doing can be said to be controlled by me, I must at least have the power to initiate it or will it to cease when I have reason to do so. The control and regulation of my actions as the actions of a unified agent seem to require this. And surely if *I* have the power to initiate or to stop what I am doing, then what I am doing must normally be in some way accessible to me. Thus for an action to be within a subject's control and responsibility the subject must be capable of knowing what she is doing." (O'Brien, 2007, p. 170.)

O'Brien holds that control is a crucial resource in explaining self-knowledge of action, and she also offers an account of self-knowledge of judging that draws on that conception. However, her way of developing that claim is somewhat different from mine.

22 Of course you can control your braking even though it is an open question for you whether you are *successfully* braking. But it mustn't be an open question for you whether braking is what you are at least attempting to do.

the goal of her activity. So to know what to do next, she must take it that she is doing now whatever she is in fact doing. When you start to brake in response to a red light, you'd better also be preparing to change down the gears. And when you do change gears, it is because you take it that you are indeed braking.[23]

So, when control your activity A, guided by goal G, you have a practical awareness of G. In virtue of that practical awareness you can select acts, in response to cues, performing which is constitutive of engaging in A. When a cue tells you, in the light of G, to Φ, you not only Φ, but also, in coming to do so, thereby take it that you are Φing (or at least attempting to). These points follow from what is involved, constitutively, in an activity being controlled in a goal-guided way.

This is not yet to explain self-knowledge of acting. To explain that, we need to explain how you can come to judge that you are Φing, when you are, in a way that is reliable and rational.

The commitment you have (the awareness that you are Φing) when you Φ is not itself a judgment and does not depend on your making any judgment. You need not make any judgment about what you are doing: you need not consciously articulate the commitment. What's more, a subject may Φ without even having the conceptual resources to self-ascribe her action in judgment; in such a case her commitment will be a sort of practical awareness of a fact, rather than an attitude to a proposition she can entertain (see section 5.3 on practical awareness of facts). However, a subject who *does* have the conceptual resources to self-ascribe her action can come to judge that she is Φing just by articulating the commitment that she has, as part of coming to Φ.

The commitment that such a judgment articulates is, plausibly, an irreducible, though not inexplicable, feature of agency. That is, it is simply part of the nature of an agent that she takes it that she is Φing when she intentionally does so. An explanation of her having that commitment would consist in an explanation of her having the capacities of an agent, and particularly the capacity to control her activities and their constituent acts. If this is right, then there is no explanation, in any particular case, of *why* an agent takes it that she is Φing, beyond what is involved in the fact that she *is* intentionally Φing.

There *is*, however, an explanation of why an agent in this position *judges* that she is Φing, if the question of what she is doing comes up: she comes to judge that she is Φing because she answers the question by articulating the commitment that she already has, as part of coming

23 This argument is indebted to Velleman's "What good is a will?" (Velleman, forthcoming), although Velleman's argument is focused on intentions.

to Φ. I suggest that coming to make a judgment by articulating a commitment of this sort will be rational, for a subject: a judgment so made will be a rational judgment. If a subject is committed to the proposition that she is Φing, then, if the question arises of what she is doing, the correct answer, as far as she is concerned, is that she is Φing. She will not be going wrong by her own lights in making that judgment. The commitment that the subject has in coming to Φ, as well as being a *commitment*, constitutes a sort of non-experiential *access* to the fact that she is Φing. When she is so committed, it is a feature of how things are for her that Φing is what she is doing. A subject who comes to judge that she is Φing by articulating this commitment can be said to be making that judgment *for the reason* that she is Φing.

In such a case, neither the fact that the subject is Φing, nor her access to that fact, are constitutively independent of her commitment to that fact. But that her access to the fact is not independent of her commitment to it is precisely why an articulation of that commitment can be a judgment made for the reason that she is Φing. She doesn't need any other, independent access to the fact, in order for it to be a reason for her.

As well as being rational, a judgment made in this way will be reliable. Coming to judge that you are Φing in this distinctive way is possible only when you really are at least attempting to Φ. If your control of your activity did not involve such an act, you would not be committed in that way to the proposition that that is what you are doing.[24]

That, I claim, is why self-ascriptions of at least certain types of action are knowledge. Needless to say, there would be a good deal more work involved in setting out and defending the control view fully. In particular, we would need a worked-out account of the nature of agency in order to achieve a deeper understanding of the connection between intentional action and the sorts of commitments I have mentioned. The sketch I have offered will suffice for present purposes.

I mentioned earlier that you typically know what you are doing under some descriptions and not others, and that an account of self-knowledge of acting ought to classify such descriptions in an intuitively correct way. The control view does so. Consider again the example of braking at a red light. Suppose you have a prior intention to stop, which you do by braking. You control your stopping the car. You control your braking. You also control

24 This claim would need more defence in a fully worked-out account of self-knowledge of action. Such a defence would perhaps appeal to the claim that it is a feature of agency that you won't take it that you are Φing if that is not what you are at least attempting to do. Such mistaken commitments would radically undermine the coordination of action.

your leg and foot movements at a certain grossly specified level. You do not control the very fine-grained adjustments of position of your knee and ankle that are required to keep your foot steady on the pedal (although they are *controlled* in the thin sense). Nor do you control what effect you are having on the anxiety of your passenger. Each of the things you control is a thing you have an agent's distinctive knowledge of; each of the things you don't control is a thing you lack such knowledge of.

Note too that the notion of control has the potential to meet the criteria I specified at the end of the last subsection. Like experience, it is tied to particular acts: at any moment you are controlling precisely those acts and activities that you are presently performing. Thus, a control-based account can explain how self-ascriptions track the acts that make them true. Like practical reasoning, the account involves the unique perspective of the agent-subject. It is not vulnerable to the objections against an introspectionist account.

I have tried to argue that the notion of control has an important explanatory role to play in self-knowledge of certain kinds of actions. I will now turn to the question: how precisely can this point be applied to self-knowledge of judging?

6.3.3 The control view and self-knowledge of judging

Judging that *p* is an act that takes place as part of a controlled activity—the activity of coming to a verdict on whether *p*, say. But the account just given won't carry over to self-knowledge of judging, because the control you have over particular judgments is not the same as the control you have over acts such as braking at a red light. Judging is in an important sense non-voluntary (section 6.3.1), and not something that you select and execute in the same way that you select and execute a driving manoeuvre. Nevertheless, I will argue, the control you have over judgments is a form of control that allows you to know that you are judging that *p*, when you are.

My claim (**1**) above was that you must have practical awareness of the goal guiding a controlled activity of yours. The argument for this claim carries over to the case of judging. The activity of coming to a verdict on whether *p* requires control, just as much as the activity of driving does. This is control of your direction in thought, guided by your goal of getting to the truth with respect to *p*, as discussed in section 6.3.1. You must be able to confine your attention to matters pertinent to whether *p*. And you must be able to move on to other matters when you conclude by judging that *p*. That is, you must know what to do in order to successfully carry out the activity of coming to a determinate verdict on whether *p*. But you

could not know what to do unless you had some sort of awareness of your goal of coming to a (true) verdict on whether *p*. Thus, when you judge that *p*, you have a practical awareness of the goal guiding your activity.

Note also that any subject who genuinely *judges* must have a certain understanding (not necessarily theoretical) of this goal, because she must have some conception of appearance-independent truth. For a subject's acceptances of contents to count as judgments, there must be a difference between something's appearing to the subject to be true, and the subject's taking it to be true; otherwise she does not judge, but merely reacts to how things appear. If there is such a difference, the subject must be disposed, in certain circumstances, to withhold endorsement of appearances—not to take things to be how they appear to be. And doing this involves an appreciation that how things are need not be how they appear to be. It involves an appreciation of the appearance-reality distinction,[25] and thus of truth. A judging subject will thus have a certain understanding of, as well as an awareness of, the goal she is pursuing in judging.

My claim (**2**) above was that you must take it that you are Φing when you Φ as part of a controlled activity in light of the goal of that activity. However, neither of the arguments I offered for (**2**) carries over to the case of judging. The first argument appealed to the need to control the execution and cessation of an action; there is no such need in the case of judging. The second argument appealed to coordination of the activity; your judgments and attitudes must indeed be coordinated, but this, plausibly, can be effected by storing the *contents* of those judgments and attitudes, rather than by having higher-order attitudes directed on those judgments and attitudes themselves.

But you do control your judging in a different way: via reason. That is, in coming to judge you cede or delegate control to perceived reasons—to considerations pertinent to your goal of ascertaining the truth with respect to *p* (say). As I argued in section 6.3.1, ceding or delegating control in this way just is aiming at the goal of truth in the way constitutive of judgment. Judgments are guided by perceived reasons, similarly to how certain physical acts (e.g. braking) are guided by perceptible cues (e.g. red lights), except that perceived reasons for judgment constrain judgment in a stronger way than perceptible cues constrain physical acts. Of course you can always step in and reopen the question whether *p*, thus undoing your sense of what verdict the pertinent reasons point to.

This type of control can ground self-knowledge of judging. In the case of braking at a red

25 Allen (1997) argues in some detail that judgment requires not just the capacity to represent, but a marking of the appearance-reality distinction.

light, recall, the cue tells you to brake, in the light of your goal; you respond by doing so, and as part of doing so you take it that you are braking. In the case of judging, the perceived reasons do not *tell* you to judge. Rather, the reasons stacking up in a certain way *suffice* for you to judge, provided you don't step in and reopen the question of how the reasons stack up. The issuing, by your reason-guided inquiry, of a verdict, *constitutes* your judging. You know the verdict of your reason-guided inquiry—you know how the reasons stack up for you, just as you know what the red light tells you to do. In knowing the verdict of your reason-guided inquiry, you can, in the light of the goal guiding that inquiry, know that you are *judging* that *p*. This is how control of judging grounds self-knowledge of judging.

To fill out this account of self-knowledge of judging more explicitly, I must say how we come to self-ascribe judgments, and why self-ascriptions made in that way will be reliable and rational.

Here, I suggest, is how we come to self-ascribe judgments:

> (**W$_J$**) When the subject is aiming (in the right way) at the truth with respect to whether *p*, and her inquiry yields the verdict that *p*, for the reason that *p* the subject judges "I judge that *p*".

The reliability of (**W$_J$**) is a straightforward consequence of the procedure. If the two conditions, that the subject is aiming in the right way at the truth with respect to *p*, and that her inquiry yields the verdict that *p*, are met, it typically will be true that she judges that *p*. The subject *may* withhold judgment, but to do so is to reopen the inquiry and make its verdict indeterminate for the time being. As soon as judgment is withheld (so to speak), the second of the two conditions is no longer met.

What about the rationality of (**W$_J$**)? This is explained in a number of steps.

Firstly, I argued above that a subject who judges will be aware of the goal governing her present activity, namely truth, or the truth with respect to whether *p*.

Secondly, when a subject's inquiry yields the determinate verdict that *p*, the subject will be aware that that is the verdict. It is *her* considered sense of where the reasons point that determines the verdict of inquiry. And her considered sense of where the reasons point is just where they point, on consideration, from her point of view. But she knows where the reasons point, on consideration, from her point of view, simply by considering where they point; it is

her answer to this latter question that fixes the answer to the former. Thus, when consideration of where the pertinent reasons point leads her to the verdict that *p*, she will be aware of *p* as where the reasons pertinent to her goal point. She will be aware of it as the verdict of the inquiry. If she is not aware that *p* is her verdict, then it is still open for her where the reasons point with respect to whether *p*.[26]

So the judging subject is aware that she is aiming at the truth with respect to *p*, and that her verdict (*the* verdict, as far as she is concerned) is that *p*. Therefore (thirdly), she will be aware that her verdict is that *p* is *true*.

Fourthly, the subject will know about the particular conscious episode in which that verdict is represented—in which *p* is represented as true. That is an instance of what I explained in Chapter 5.

Fifthly, the subject who can follow ($\mathbf{W_J}$) will appreciate the internal connection between judgment and truth. That is because she must, in order to be capable of self-ascribing judgment, possess the concept of judgment, and the connection between judgment and truth is internal to the concept.

Since the subject knows that she is having an episode, representing the verdict that *p* is true in the context of the goal of ascertaining the truth with respect to *p*, it will be rational for her to take it that the conscious episode is a *judgment*. Thus, when a subject judges that *p*, she thereby has a reason, in the context of her practical awareness of her goal in judging and her appreciation that *p* represents the verdict of an inquiry guided by that goal, to self-ascribe that act of judgment.

To make clear how this explanation works, we can consider how the subject knows the episode is *not* of certain types. How does the subject know she is not merely *supposing* that *p*? Because she is aware of the goal governing her activity, a goal distinctive of judgment. How does she know that she is not merely having it *seem* to her that *p*, without performing the act of judgment? Because she is aware of *p* as a verdict, as the outcome of an inquiry governed by the goal of reaching a verdict on the truth with respect to whether *p*. To withhold judgment, once the verdict is in, would be to nullify its status as a verdict; to reopen the question of what verdict the reasons (*her* reasons) determine. Finally, how does the subject know that she is not merely *guessing* that *p*? Because the procedure I described involves not only the goal of truth (which also guides guessing), but aiming at that goal in

26 The argument of this paragraph clearly owes a lot to Richard Moran's insights (Moran, 2001). But I wish to make those insights work in a slightly different way. I will compare my view with Moran's in Chapter 7.

the right way. It involves the subject's sense of the reasons as yielding a decisive verdict—as establishing that *p*. (Truth guides guessing in a less constraining way than it guides judging.) So the subject will be judging, not merely guessing.

6.3.4 Objections and replies

This account has drawn on many claims about action and about judgment in particular. I want to finish this section by answering a couple of important objections.[27]

One way to attack the account would be to object to my reliance on the notion of control. This attack could take either of two forms. [28]

Objection 1

It might be claimed that control is not a rich enough notion to capture the epistemic phenomenon of self-knowledge of action. According to this objection, the intuitive notion of a controlled action merely has to do with the action's entering into the right sorts of causal relations. But those causal relations won't suffice for the agent to have any self-knowledge. Unless self-knowledge is just built into the account of control, something's being controlled will leave it open whether the agent knows about it.

My reply is that I did not rely on a generic notion of control, but on a specific, personal-level notion of control. This notion is not merely causal-explanatory, but intentional. Under this notion, pieces of behaviour that are controlled in the thin sense will not count as being acts that the agent controls. Admittedly, I have not offered a systematic analysis of this notion of control that I have relied on. And my characterisation of it has arguably presupposed the notion of intentional action. But this does not strike me as problematic for the account. The notion of intentional control has a firm enough intuitive basis to offer the sort of explanatory

27 The most obvious line of attack against the view I have presented would be to object to the two conditions mentioned in ($\mathbf{W_J}$) for self-ascribing judgment. It might be argued that these conditions are not sufficient for judging, or that they are not necessary. If either of these claims is correct, self-ascriptions made according to ($\mathbf{W_J}$) will not track acts of judgment. I do not treat either of those objections here, because I answered both of these objections in the course of presenting the account. To say that the two conditions are not sufficient is to say that the subject could refrain from judging even when they are met; I have argued that to refrain from judging would be to undo the verdict of the inquiry into whether *p*. To say that they are not necessary is to appeal to the possibility of akratic judgment, which I dealt with in section 6.3.1.

28 Thanks to Chris Peacocke, who pointed out these potential objections to any account based on control. A version of the first one can be found in his "Mental Action and Self-Awareness II" (forthcoming).

support I demand of it. And it is not a flaw that an account of self-knowledge of action would presuppose a notion of intentional action: it is very plausible that full-fledged self-knowledge is explained *by*, and not explanatory prior to, rational agency.

Objection 2

It might be claimed, alternatively, that control is too demanding a notion to capture self-knowledge of action. A person can be out of control and yet know what she is doing: someone in a fit of rage may lack control of her actions, but she will not be blind to them.

But the actions of someone in a fit of rage will count as being controlled by the agent, in my sense, as long as they are not externally compelled, but are selected, initiated, guided and stopped by the agent. The sense in which such a person is out of control is just that she is behaving contrary to certain constraints, and perhaps contrary to her own best judgment, not that her actions are driven by some compulsion or external force.

Objection 3

A different line of objection appeals to pathological cases to attack my account of the connections between control, action and self-knowledge. Some schizophrenic patients suffer from delusions of control: they believe that certain bodily acts and thoughts of theirs are initiated and controlled by some external agent (see Frith, 1992). One way to understand this symptom is this: the patient intentionally acts, or thinks a thought, but is not willing to self-ascribe that action or thought. This suggests that intentional action, and control, are *not* sufficient for the 'commitment' I described above: an agent can intentionally Φ and yet not take it that she is Φing. An experience-based view of self-knowledge of acting offers a natural explanation of this phenomenon: these schizophrenic patients simply lack the experience of acting that normally accompanies the exercise of agency (see Peacocke, 2003). But the phenomenon seems inexplicable on a control-based view.

This objection assumes that normal intentional control of action can be intact, even when self-knowledge of acting is absent or impaired. In fact, however, delusions of control seem to be associated with deficits in control (Frith and Done, 1989). So the symptoms described above may be explained by a disorder of agency, rather than a disorder of experience: it may simply be wrong to describe them as involving normal agency with abnormalities of self-knowledge. Indeed, abnormalities and deficits in willed agency are common symptoms of

schizophrenia (Frith, 1992).

In any case, even if, as seems likely, delusions of control involve abnormalities in conscious experience, and not just abnormalities in agency, it doesn't follow that the abnormality consists in the absence of an experience on which self-ascriptions of acting are normally based. Perhaps the patient undergoes an experience as of a thought being inserted in her mind, but she does not perform any *act* of thinking that thought. Or perhaps the patient thinks a thought, but lacks a certain phenomenology normally associated with thinking thoughts, and the absence of this phenomenology is a defeater for a self-ascription of acting; even then, it wouldn't follow that, in non-pathological cases, the *presence* of this phenomenology is ordinarily the epistemic basis for self-ascriptions of acting.

6.3.5 The specialness of self-knowledge of judging

Self-ascriptions made according to ($\mathbf{W_J}$) will have the now-familiar distinctive features of self-knowledge.

(a) *Security*

If you are aiming in the right way at the truth with respect to whether *p*, and your inquiry yields the verdict that *p*, then, as long as you don't reopen the inquiry, it will be true of you that you *judge* that *p*. It will not be a supposition, a seeming, a guess, or whatever. So self-ascriptions competently made according to ($\mathbf{W_J}$) will be true.

The account also respects the modal character of security (not only are self-ascriptions invariably right, but it is modally difficult for them to go wrong). ($\mathbf{W_J}$) does not involve any contingent mechanism for hooking up to facts about your judgment, but only on the nature of judging. Its availability guarantees the obtaining of the relevant facts.

(b) *Salience*

It is, I have argued, *constitutive* of judgment that an act of judging is coming to the verdict of an inquiry governed by the goal of truth (or the truth with respect to a particular proposition). So whenever you perform an act of judgment, and provided you possess the requisite concepts, ($\mathbf{W_J}$) will be available, and you will be in a position to know about the judgment.

(c) *First-person privilege*

(**W$_J$**) is available only to the agent-subject who judges. It depends crucially on the subject's own inquiry into whether *p*.

(d) *Authority*

Once again, authority can be explained by security and first-person privilege. Self-ascriptions of judging are not to be gainsaid because they are exceptionally secure and reached in a privileged way. The demand for justification is inappropriate because no *third-personal* justification can be offered: at best the subject could reiterate that she really has come to the verdict that *p*, which, from the perspective of an interlocutor, is no justification at all.

(e) *Immediacy*

(**W$_J$**) does not involve acquiring or attending to grounds for the self-ascriptive judgment, nor does it involve finding anything out. The mere fact of coming to the verdict that *p* puts you in a position to self-ascribe the judgment that *p*. No further checking is required.

(f) *Transparency*

The procedure I described is not introspective. You come to know that you judge that *p* by considering whether *p*: you aim to ascertain whether *p* is true, and do so by attending to what you take to be the relevant considerations.

(g) *Commitment*

Commitment, recall, is the feature that a subject who judges "I judge that *p*" thereby commits herself to the truth of *p*. I argued in Chapter 3 (section 3.2.3) that this feature must be explained by the way in which the subject comes to make the self-ascriptive judgment. The procedure that I have described inherently involves the subject's coming to the verdict that *p*; she could not come to self-ascribe by that procedure if it were an open question for her whether *p*. The self-ascription of judgment is effectively parasitic on the judgment it self-ascribes. It will inherit, and not merely be accompanied by, the relevant commitment. Thus, she could not be alienated from her self-ascribed judgments.

6.4 Conclusion

In this chapter I have addressed the question of how you know what *type* of conscious state or episode you are enjoying. I considered in detail two cases: perceptual experience and judgment. Perceptual experience, I argued, is characterised by *directness*—a distinctive way in which states of affairs are given in perception. I argued that directness can rationalise self-ascriptions of perception, even though it is not introspected. Judgment, on the other hand, is a type of mental *action*, and self-knowledge of judging is a species of self-knowledge of action. I argued that self-knowledge of action is grounded in *control*, rather than in practical reasoning or in experience. I claimed that you know you are *judging*, when you are, in virtue of your practical awareness of the goal by which you control your judging, and your awareness of the verdict you reach in pursuing that goal.

That completes my positive account of self-knowledge of conscious states and episodes.

The next and final chapter will consider the role of self-knowledge in our rational and cognitive lives. It will compare my explanation of how self-knowledge plays that role with other accounts that make that role more central to the epistemology of self-knowledge.

CHAPTER 7

THE ROLE OF SELF-KNOWLEDGE: VARIETIES OF WARRANT AND THE DIRECTION OF EXPLANATION

Lucy O'Brien has written:

> "Self-knowledge is not just another epistemic acquisition, like knowledge of trains or stamps. It is knowledge that lies at the core of our understanding of what it is to be, and what is important in being, a person." (O'Brien, 2003, p. 375.)

An epistemology of self-knowledge must respect its central role in our personhood. But we can also ask: how does this role of self-knowledge relate to the *warrant* for self-knowledge? Is the role of self-knowledge in personhood partly constitutive of, or explanatory of, its warrant? Or is the warrant for self-knowledge independent of, and partly explanatory of, its having and fulfilling that role? The account that I have offered over the last two chapters makes no explanatory appeal to the role of self-knowledge and thus fits into the latter of these two strategies. This chapter will show that my account can help explain the role of self-knowledge, and will argue that an account of this sort enjoys advantages over those accounts according to which the role of self-knowledge is constitutive of, or explanatory of, its warrant.

I will start, in 7.1, by briefly characterising the role of self-knowledge and showing that my account can help explain it. In 7.2.1 I will distinguish three varieties of warrant: top-down, mixed and bottom-up. On the account I have offered, the warrant for self-knowledge is bottom-up. A warrant that constitutively or explanatorily depends on the role of self-knowledge would be top-down, or mixed. 7.2.2 will consider Tyler Burge's top-down account and argue that it does not give a satisfying explanation of the warrant for self-knowledge. 7.2.3 will consider Richard Moran's account, which can be read as positing a mixed warrant. I will argue that the top-down element of his account is based on an implausible claim about rational agency, and should be rejected. The problems with both of these accounts suggest that we should adopt the bottom-up strategy of my moderate epistemic account, which can accommodate the important insights of Burge and Moran.

7.1 The role of self-knowledge

Why is self-knowledge especially important? After all, we can imagine subjects who lack the conceptual resources to make self-ascriptive judgments or hold self-ascriptive beliefs, and who therefore lack self-knowledge of this sort, and yet who have conscious representational states, who have goals, and who judge and act intelligently in response to reasons. Such subjects would be unable to reflect on her thoughts, actions and goals; but thinking, acting and having goals do not seem in themselves to require such higher-order reflection—they seem to require only first-order representation. Such subjects would arguably be unable to think *about* their reasons, considered as such; but, again, responding appropriately to reasons does not seem to require thinking about them *as* reasons (see Hurley, 2001). In short, it seems that a subject can think and act rationally, in an important sense, without having self-knowledge of the sort I have been discussing in this thesis. Plausibly, young or very young children are such subjects (as I will argue in section 7.2.3).

Mature human rational agency goes beyond merely judging and acting intelligently in response to reasons. It also involves the capacity to *reflect* on your own judgments, attitudes and actions, and to consider explicitly whether they are in line with your goals, your assessment of reasons, and so on. I will call reasoning that involves such higher-order reflection "reflective reasoning". The assessments you reach in your reflective reasoning (for example, "My reason for judging that *p* was a bad reason") contribute to what you ought, at the first-order level, to do or think (for example, revise your attitude to *p*). But they can do so only because such reflection involves *knowledge* of your thoughts, experiences and attitudes. If you were not knowledgeable about what you think (and want and intend and do), and why you think it, you could not effectively employ higher-order reflection to rationally assess and revise what you think:

> "If reflection provided no *reason-endorsed* judgments about the attitudes, the rational connection between the attitudes reflected upon and the reflection would be broken. So reasons could not apply to how the attitudes should be changed, suspended, or confirmed *on the basis of* reasoning depending on such reflection. …
>
> "[Furthermore, i]f reflective judgments were not normally *true*, reflection could not add to the rational coherence or add a rational component to the reasonability of the whole process." (Burge, 1996, pp. 101-2. Emphasis added.)

Reflective reasoning of this sort, I will now argue, is not just *more* reasoning, to go alongside first-order reasoning. It changes the nature of your rational agency. It adds a dimension of

freedom, and of responsibility, that is beyond the *merely* rational agent who cannot reason reflectively—a dimension of freedom and responsibility that is characteristic of persons. Thus, the importance of self-knowledge is at least partly to do with its contributing to this freedom and responsibility.

In her judgments and actions, the agent who is merely rational, and lacks reflective reasoning, is in a sense at the mercy of how reasons stack up for her, given her goals. When the agent has sufficient reason, given her goals, for a certain act or judgment, she will (all going well) simply be moved to perform that act or make that judgment. She will lack a further conception of what she is doing—a conception of herself as performing such-and-such an act for such-and-such a reason (that would be to have reflective reasoning). By contrast, for the reflective, self-knowing agent, being moved by reasons is not the end of the story. Certain further questions are inescapable for the self-knowing agent. Is my assessment of these reasons correct? Shall I be moved by them? Is the goal that this action serves really of value? Does this action fit in to my broader plans and self-conception? The self-knowing agent thus faces the choice of reflectively endorsing, or refusing to endorse, those acts and judgments towards which she finds herself moved by reasons. She is free to *commit to*, to *identify with*, those judgments, attitudes and actions, or to withhold such commitment and perhaps refrain from judging, from acting, or whatever. The merely rational agent simply judges, believes or acts. Accordingly, the self-knowing agent is responsible for her judgments, attitudes and actions, in a deeper way than the merely rational agent.

Frankfurt (1982) famously makes a point along these lines, about desire. The agent who knows her desires can reflectively endorse, and thus commit to, those of her desires that fit into her conception of what is worth wanting. She can shape her desires accordingly. She may desire things that she thinks are not worth wanting, and be unable to rid herself of such desires. But she thereby becomes alienated from those desires in a certain way: we are tempted to think of them more as addictions or urges than as genuine desires of the agent.[1] A subject who knows her desires is self-determining in a way that the agent who lacks self-knowledge is not. The agent who lacks self-knowledge wants what she wants, and does what her desires motivate her to do. There is no distinction, in the case of such an agent, between reflectively endorsed desires and desires from which she is alienated.

---

1 Frankfurt does not use the terminology of commitment and alienation. He would say that the agent who cannot rid herself of a desire that she refuses to identify with thereby lacks freedom of the will. Only self-knowing agents can have or lack freedom of the will. Agents who lack self-knowledge can *act* freely—they can do what they want to do—but the notion of freedom of the will is not applicable to them.
See Chapter 3, section 3.2.3, for a discussion of alienation and commitment, focused on judgment.

The freedom and responsibility of self-determination that go with reflective reasoning and self-knowledge, are aspects of what it is to be a person (Burge, 1999; Frankfurt, 1982). Thus, self-knowledge is central to the nature and importance of personhood.

This important role of self-knowledge is, as Lucy O'Brien hints, related to the fact that self-knowledge is not a contingent acquisition—not something we could easily lack. It is difficult to make sense of the possibility of a subject who has the conceptual repertoire of a self-knower, but somehow lacks epistemic access to her own conscious states and episodes—a subject who is capable of wondering about her present conscious state, but for whom no answer is forthcoming (c.f. Shoemaker, 1996). Equally, it is difficult to make sense of the possibility of a subject who *systematically* makes mistaken self-ascriptions, or who self-ascribes judgments and beliefs that she refuses to commit to (see section 3.2.3). We are inclined to say, of such cases, either that something has gone wrong in the subject's rationality or understanding, or that there is no single, unified subject of both the first-order states and the self-ascriptions of those states.

The account I have offered in this thesis respects the role of self-knowledge in personhood, and helps to explain how it can fulfil that role. Let me explain how it does so, with reference to the points made in this section.

On my account, self-knowledge is grounded in certain of our fundamental cognitive capacities, in what it is to enjoy a conscious state or episode at all (see Chapter 5), and in the very natures of various types of conscious state and episode (see Chapter 6). Any subject who possesses the relevant cognitive capacities will be capable of knowing her own conscious states and episodes, and will avoid mistaken self-ascriptions, provided she has the appropriate conceptual repertoire. My account made no appeal to any contingently reliable mechanism for hooking up to facts about your own conscious states and episodes. Rather, it rooted self-knowledge deeply in our natures as conscious subjects. No wonder, then, that self-knowledge is not something a subject could easily lack. And no wonder, therefore, that it is apt to play a central role in personhood.

There is no difficulty, on my account, in explaining how the self-ascriptions made in the course of reflective reasoning can make a rational contribution to what a subject ought to do or think at the first-order level. I have assumed from the beginning that the warrant for self-ascriptions is a rational warrant. In light of that assumption, I have offered an account on which self-ascriptions are rationalised by reasons. They are also normally true. Thus, self-ascriptions help to sustain a rational connection between reflective reasoning and the first-order thoughts, attitudes and actions such reasoning is directed on.

In particular, my account promises to explain the inescapable rational *commitment* involved in certain instances of self-knowledge. I argued in Chapter 6 (section 6.3.5) that self-ascriptions of judgment will inherit the commitments of the judgments they self-ascribe. More generally, I have been keen to emphasise that your perspective on your conscious states and episodes is radically unlike your perspective on the world outside you. Your conscious states and episodes *are* your perspective on the world outside you.

At the same time, my account allows that there can be intentional action and rational agency without full-fledged self-knowledge. While self-knowledge is rooted in certain fundamental capacities and features of our subjecthood, it also goes beyond those capacities and features. It involves conceptual sophistication that is not guaranteed by those capacities and features. I have been anxious to emphasise (see especially section 5.2.3) that conceptually articulated knowledge in some domain can be partly grounded in a practical awareness of some fact or distinction—a practical awareness that does not consist in particular judgments and that can exist without conceptually articulated knowledge. On the account I have offered, any rational agent will have *some* of what is required for self-knowledge. Perhaps any rational agent must have some sort of primitive self-awareness. But a rational agent may lack full-fledged self-knowledge. Thus, the account respects the point I made at the beginning of this section.

In the light of all this, it seems to me that my account of self-knowledge can contribute to the explanation of how we come to have the freedom and responsibility of self-determination that is characteristic of personhood. It is partly because of the resources that my account appeals to—certain fundamental cognitive capacities, what is involved in enjoying a conscious state or episode, and the natures of various types of conscious state and episode—that we can meet the conditions for being persons. The epistemic character of self-knowledge explains how it can fulfil the role that it does.

7.2 Varieties of warrant and the direction of explanation

On the form of account I have offered, self-knowledge can play its role partly because it has the epistemic character it does. But there are alternative views of the connection between the warrant for self-knowledge and the role of self-knowledge. In this section I will consider those alternative views.

7.2.1 Top-down, mixed and bottom-up warrants

The taxonomy of top-down, bottom-up and mixed is due to Lucy O'Brien (2005). It is a taxonomy of types of warrant that can attach to a judgment made on some basis, where having that basis involves enjoying some mental state or episode.

A judgment has a top-down warrant when the warrant is explained at least partly by features of the warranted judgment itself rather than the basis on which it is made—metaphysical features or features of some broader role that judgments of that type play in the subject's rationality, cognitive economy, or whatever. The judgment's being made on a particular basis may explain why it is reliable, but the 'top-down' features of the judgment itself will, on this picture, explain why the judgment is not merely reliable, but also epistemically warranted.

A judgment has a bottom-up warrant when the warrant is explained solely by features of the basis on which the judgment is made. If having that basis involves enjoying a mental state or episode, the warrant will be explained by the content of that state or episode, and the type of state or episode it is.

Mixed warrants have both top-down and bottom-up elements. A mixed warrant is one that a judgment has solely in virtue of being made on a basis with a particular nature, but such that the nature of the basis, and thus its providing that warrant, depends on top-down features of the judgment it warrants.2

According to my account, the warrant for self-ascriptions is bottom-up. The bases for self-ascriptions of conscious states and episodes are simply the occurrences of those states and episodes themselves, which provide access to reasons. These states and episodes provide warrants for self-ascriptions in virtue of what it is for a state or episode to occur consciously, and of features of particular types of state and episode, as well as the subject's cognitive capacities. They do not in their nature constitutively involve self-knowledge or self-ascriptions.

But does this account get right the relation between the warrant for self-knowledge, and the role of self-knowledge? In the rest of the chapter I want to consider the plausibility of the alternative suggestion, that the role of self-knowledge somehow explains its warrant, making it a top-down or mixed warrant.

---

2 O'Brien also uses 'mixed' to describe theories according to which judgments in a certain range have two independent warrants, one top-down and the other bottom-up. This sort of theory is less interesting for my purposes. I want to focus on the nature of the warrants themselves; if a judgment has two independent warrants, that is not in itself relevant to the nature of either warrant. While my view is that the warrant for self-ascriptions is bottom-up, that does not entail that there aren't further, top-down sources of warrant.

7.2.2 The top-down account: Burge

The most sophisticated top-down theory in the literature is that of Tyler Burge (1996, 1999). Burge concentrates on self-knowledge of beliefs, rather than of conscious episodes. However, his form of account can also be applied to self-ascriptions of at least some conscious episodes, including judgments. I will call this application of his theory 'the Burgean view'. I do not claim that Burge endorses this application of it.

After outlining the Burgean view, I will state a *prima facie* problem for the view, and show that no adequate response to the problem can be given that remains within the Burgean approach. I will conclude that my own account enjoys an important advantage over the Burgean account.

The warrant for self-ascriptive judgments, on the Burgean view, is an entitlement: a warrant that a subject can have and exploit without having any understanding of the warrant (Burge, 1993). The entitlement to self-ascriptive judgments is explained by the role of those judgments in a reflective, self-critical sort of reasoning which he calls “critical reasoning”. Burge defines critical reasoning as:

> “reasoning that involves an ability to recognise and effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such.” (Burge, 1996, p. 98.)

The key element here is that the critical reasoner's judgment or belief is not just sensitive to reasons, but that she is able to reflect on those reasons, conceived as such, and on their force, to evaluate her own attitudes and reasoning, and to implement those evaluations in her reasoning.

Self-knowledge is essential to critical reasoning. For it is essential to critical reasoning that the reasoner's self-ascriptions of thoughts and attitudes generate reasons, in line with the norms of critical reasoning, for revision or reaffirmation of those thoughts and attitudes. For example, if, when reasoning critically, you realise that an attitude you have self-ascribed is unjustified, then you have a reason to revise that attitude. Self-ascriptions can generate such reasons only if they are knowledge (Burge, 1996, pp. 98-103; and see section 7.1 above).[3]

3 Although Burge's own claim is that such oughts can be generated only if one's self-ascriptions of belief are knowledge, it is very plausible that the activity of critical reasoning also requires self-

According to the Burgean view, it is this role of self-knowledge as a transcendental condition on critical reasoning that explains the entitlement to self-ascriptions:

> "[The entitlement to self-ascriptive judgments] depends on the judgments' being instances of a kind essential to critical reasoning. Critical reasoning presupposes that people are entitled to such judgments. Since we are critical reasoners, we are so entitled.
>
> "Epistemic entitlement derives from jurisdiction—from the place of the judgments in reasoning." (Burge, 1996, p. 116.)

This sort of account, understood as an epistemic account of self-knowledge, faces a problem. *Prima facie*, it does not seem to *explain* why particular self-ascriptive judgments are warranted. The transcendental argument establishes *that* we have self-knowledge, and that self-knowledge has a special status as a condition on critical reasoning. But that is not the same as giving a satisfying explanation of the epistemic status of particular self-ascriptions—of how we come to make self-ascriptive judgments in such a way that they are warranted. The fact that critical reasoning generates reasons, in line with certain norms, shows that the critical reasoner's self-ascriptions are warranted, but it seems plausible that she has those reasons *because* her self-ascriptions are warranted, rather than that they are warranted because they generate those reasons.

One reply to this worry would be to deny that there is any need for a *further* explanation of the warrant for self-knowledge, beyond the explanation offered by the transcendental argument. The transcendental argument demonstrates that true self-ascriptive judgments are licensed by norms of reason. According to this reply, judgments made in ways that are licensed by norms of reason are *ipso facto*, and without need for further explanation, judgments to which their subjects are rationally entitled. Therefore, our self-ascriptive judgments are rationally entitled in virtue of those norms.

This is arguably Burge's own view, when he describes entitlement as "a status of operating in an appropriate way in accord with the norms of reason" (Burge, 1996, p. 93).

This reply thus challenges the conception of rational warrant that underlies my objection. In my view, to explain the rational warrant for a judgment is (*inter alia*) to show why the subject is rational, from her own point of view, in making that judgment. The reply I am

---

ascriptions of many conscious episodes to be knowledge. Consider, for example, a judgment based on perceptual experience. If a critical reasoner comes to recognise that the conditions for perception are poor, she will be under a requirement to revise the judgment. But she can be under such a requirement only if she knows that she has made the judgment, *and* that she made it based on a perceptual experience. This, at any rate, is the *Burgean*'s claim.

considering involves a thinner conception of rational warrant, according to which it derives simply from acting in accord with certain norms.[4]

The thin notion seems to me to be too thin to capture the sort of warrant we have for self-knowledge. We should distinguish between making a judgment that is in accord with a norm, and making a judgment because it is in accord with a norm—just as we distinguish between conforming to a rule and following a rule. The Burgean conception of warrant appeals to the former notion; but only the latter notion can capture our warrant for self-knowledge. A judgment that is merely in accord with some norm, but not made *because* there is such a norm, is made blindly, in a certain sense. Imagine a subject who, without any thought, comes up with the answers to difficult square-root calculations: they simply pop into his head. The subject, in coming to judge by this method, will certainly be operating in accord with a norm, for he is performing the right mathematical operations. But he will not be making those judgments because they are in accord with a norm. If the subject has no idea that his guesses are correct, he will be operating blindly. By contrast, a subject who gives certain answers because they are in accord with a norm will be giving answers that seem appropriate to him. Our self-knowledge is like the latter rather than the former. Self-knowledge is a paradigm of judgments that are rational from the subject's point of view (see section 1.4).

So, an account that merely shows that self-ascriptions are in fact in accord with a norm will not be as satisfying as an account that shows why self-ascriptions are rationally warranted in the thicker sense (i.e. made because they are in accord with a norm). Given that my account does just that, it has a substantial advantage over the Burgean account, unless the Burgean can offer some further explanation of the warrant for self-knowledge.

Can the Burgean provide such a further explanation? The challenge is to explain why particular self-ascriptive judgments are warranted, as opposed to merely establishing that they are warranted. The Burgean must, in doing so, make some explanatory appeal to critical reasoning, on pain of abandoning his top-down approach. I will now show that the Burgean lacks the resources to provide such an explanation.

What further resources could the Burgean appeal to?

---

4 This reply is to be distinguished from a slightly different one. The Burgean could claim that his project is more modest than I am suggesting: he is aiming only to give an anti-sceptical reassurance that we have self-knowledge, not to give an epistemic explanation of that knowledge. So my explanatory demand is illegitimate. If that were the Burgean's project, his view would not be a rival to mine. The reply presently being envisaged is that the Burgean is engaged in the same project as me, but his conception of what is required to carry out that project is more minimal than my own.

One resource is concept-possession. Burge himself does appeal to this resource, saying: "Understanding and making [self-ascriptive] judgments is constitutively associated both with being reasonable and with getting them right." (Burge, *ibid*..) The claim here is that it is constitutive of possessing the various concepts involved in self-ascriptions—the concept of the first person, concepts of the mental states and episodes one self-ascribes, etc.—that your self-ascriptive judgments are reliable and rational (see also Burge, 1999).

As we saw in Chapter 5 (5.3), however, the mere fact that possession of certain concepts necessitates willingness to make certain judgments does not itself give a satisfying explanation of why particular judgments are rationally warranted. There must be some further explanation of why, in certain particular circumstances, a subject who possesses those concepts will be willing to apply them. I offered such an explanation as part of my account. The Burgean must offer an alternative explanation; and his explanation must appeal to critical reasoning, if that part of his account is not to become redundant. Thus, the appeal to concept-possession leaves us back where we started—needing an explanation of why particular self-ascriptions are warranted, that appeals to critical reasoning..

Burge does hold that a full understanding of the first person, and of propositional attitudes, consists, in part, in being a critical reasoner (Burge, 1999). What is really needed, then, is an account of critical reasoning that explains why the self-ascriptive judgments made by a critical reasoner will tend to be reliable and true. According to Burge, what distinguishes critical reasoning from the mere making of higher-order judgments and evaluations is that critical reasoning consists in embracing one's reflective second-order judgments and the first-order attitudes on which they are directed within the same first-person *point of view* (see Burge, 1996, pp. 108ff). Thus, a second resource to which the Burgean might appeal is this notion of a point of view.

But what does this notion amount to? What individuates points of view is, roughly, how reasons transfer across and within them: point of view A and point of view B are identical just in case the following conditional holds: if, from point of view A, there is good *prima facie* reason for the occupant of point of view B to Φ, it *immediately* follows that from point of view B there is *prima facie* reason to Φ (this is my paraphrase of *ibid*., pp. 108-9).[5] Certain reasons apply within points of view but do not transfer across points of view. For example, if from your point of view there is reason for S to take the next train, and you are S

[5] In the context of the view of reasons I set out in Chapter 4, this point may be better formulated with reference to *having* reasons than with reference to there *being* reasons. Here I try to follow Burge's formulation. Nothing turns on it for present purposes.

(and you know it), then it immediately follows that, to that extent, there is reason from S's point of view to take the next train. By contrast, if from your point of view there is reason for *me* to take the next train, it doesn't immediately follow that there is reason, from my point of view, to take the next train. Thus, for critical reasoning to take place from a single point of view is for the reasons available from the point of view of the critical reasoner to apply immediately to the point of view of the attitudes being reflected on—it is for the reasoning immediately to give reasons for revision or endorsement of those attitudes, from the point of view of their subject.

So critical reasoning takes place from a single point of view when it generates reasons, in line with the norms of critical reasoning, for the revision or endorsement of the attitudes being reflected on. But this account just takes us back to the original claim that featured in the unsatisfying transcendental argument—that critical reasoning constitutively involves the generation of certain reasons that require self-ascriptions to be knowledge. In trying to pursue a further explanation, we have come upon that very claim again. The notion of a point of view does not add any more explanatory power to that argument. The Burgean account seems to run out of resources here. The generation of reasons according to the norms of critical reasoning just won't explain the warrant for self-knowledge.

An alternative way to try to rescue the Burgean account would be to deny that it is merely being *subject* to those norms that explains why critical reasoners' self-ascriptions are warranted. It might be said that the explanatory work is done by the critical reasoner's *understanding of* those norms.[6] This version of the view attributes to the self-knowing subject some grasp of the theory of critical reasoning. Given such a grasp, a subject won't self-ascribe a judgment, for example, unless she recognises some reasons as sufficiently favouring that judgment, and is prepared to make the appropriate commitment.

But this is much too demanding. It is not plausible that all subjects capable of knowing their own conscious episodes have such a grasp of the theory of critical reasoning. Children aged 3 to 4 years can report the occurrence of a conscious episode in which they learned something, but tend not to understand the connection between the occurrence of that episode and their possession of knowledge. They can say, "I saw it", but they cannot say "I know because I saw it"; nor, in many cases, do they seem to have implicit knowledge of how they know (Haigh and Robinson, in press). They are also poor at demonstrating a grasp of how (e.g. by what modality) you could come to know something (e.g. the colour of something) (O'Neill and Chong, 2001). Generally, children are poor at reflecting on the credentials of

6 Sometimes this seems to be Burge's position: see Burge (1999, pp. 41-2).

their knowledge until around 7 years (Robinson et al., 2007, p. 3). All of this suggests that grasp of a theory of reasoning and knowledge, particularly a sophisticated theory like Burge's, is developmentally far behind the ability to make knowledgeable psychological self-ascriptions. Thus, the warrant for such self-ascriptions does not depend on grasp of such a theory.

I take the above discussion to have exhausted the most obvious ways of developing a Burgean account of self-knowledge that responds to the explanatory problem I presented. I do not see any other available way of trying to explain self-knowledge in terms of critical reason or a similar notion. If there is an available way, the burden is on the defender of the Burgean view to show that it is viable. In the meantime, we can conclude that the Burgean view has a serious problem: it cannot offer a satisfying epistemic explanation of the warrant for self-ascriptions.

Although I have dealt specifically with the Burgean view, the sorts of considerations I have adduced can be directed against any top-down account of the warrant for self-knowledge. It is hard to see how top-down resources can provide a satisfying explanation of that warrant. No matter what role self-ascriptive judgments play in a subject's cognition, it is difficult to see how this role could explain why, when a subject comes to make a self-ascription, she is operating not merely in accordance with a norm, but in a way that is appropriate from her own point of view. Such an explanation would involve an account of why the subject is willing to make those judgments. But it seems that this willingness would have to be explained, in a given case, by the particular circumstances in which the judgment is made, in combination with the subject's understanding of the content of the judgment. Top-down features, like the role of self-knowledge, are not apt to feature in such an explanation. There may be an argument from the presence of those features to the existence of certain rational norms that self-ascriptive judgments must meet. But it is another matter to give an illuminating account of why, in making self-ascriptive judgments, subjects are not merely operating in accord with norms, but doing what is appropriate because it is appropriate.

I have not disagreed with Burge's claims about the role of self-knowledge; on the contrary, I think he has deepened our understanding of the importance of self-knowledge. We can respect these claims, while eschewing a top-down *epistemic* account of self-knowledge. We can hold that self-ascriptions are warranted on a basis that is independent of their top-down role, and it is partly because they are warranted in this way that a self-knowing subject can be a critical reasoner. We can respect the thought that self-knowledge is deeply rooted in our nature by holding that it is grounded in the occurrence of conscious states and episodes, and

in certain fundamental cognitive capacities, rather than in some introspective faculty. This is the sort of account I have defended. Since this view contributes to the explanation of the features of self-knowledge that Burge emphasises, while also offering a satisfying epistemic explanation of self-knowledge, we should prefer it to the Burgean view.

7.2.3 The mixed account: Moran

A mixed account is one according to which self-ascriptive judgments are warranted wholly in virtue of being made on a certain basis, but the nature of the basis, and therefore its providing a warrant for self-ascriptions, depends on a certain feature of self-ascriptive judgments. Richard Moran can be read as offering such an account (Moran, 2001).[7] Like the top-down theorist considered above, the mixed theorist appeals to the connection between self-knowledge and *reasons*, *reasoning* and *deliberation*. But the mixed theorist holds that the *basis* for self-knowledge *consists in* the reasons, reasoning or deliberation that leads to the state or episode known about.[8] He claims, further, that the nature of reasoning and deliberation itself depends on self-knowledge (this is the top-down element). Thus a self-ascription is warranted wholly by the basis on which it is made, but that basis in turn depends on self-knowledge.

I will argue that such an account misconstrues the relation between self-knowledge and deliberation: the sort of deliberation that might plausibly constitute the basis for self-knowledge does not in turn presuppose self-knowledge. First I must outline Moran's account in more detail.

Moran holds that knowledgeable self-ascriptions are the outcome of rational agency—that the procedure for coming to know your own judgments (say) is parasitic on the procedure for coming to make judgments.[9] His take on this idea is that a knowledgeable self-ascription of the judgment that *p* will be a self-ascription that you arrive at by employing your ordinary method for coming to judge that *p*—for ascertaining whether *p*. A self-ascription of a judgment is answerable to the reasons that rationalise that judgment, rather than to evidence

7 There is more than one way to read Moran's account. In *Authority and Estrangement* (Moran, 2001) which is the main source for this subsection, it seems to be a mixed account. In some later work (e.g. Moran, 2004, pp. 465ff.) he seems to shy away from the top-down element that is present in the book, and endorse a bottom-up account. My concern is with the plausibility of the mixed account, not with whether that is ultimately the account that Moran wants to offer.

8 Obviously, this account applies only to self-knowledge of reason-driven states and episodes, such as belief and judgment.

9 This thought has strongly influenced my own account of self-knowledge of judging, in Chapter 6.

about whether you in fact make that judgment. Your self-ascription of judgment constitutes self-knowledge when you treat the self-ascription, not as the report of a pre-existing thought that you have discovered, but as the formation or re-affirmation of an attitude—as the making of the commitment constitutive of judging.

To self-ascribe in this way is to assume that your judgment (or whatever) is *up to you*—that the facts about your judgment are fixed by your own deliberation as to what is the case. And you are entitled to that assumption, according to Moran, because making it is constitutive of being a rational agent at all:

> "One must see one's deliberation as the *expression and development* of one's belief and will, not as an activity one pursues in the *hope* that it will have some influence on one's eventual belief and will. Were it generally the case ... that the conclusion of deliberation about what to think about something left it open for him what he *does* in fact think about it, it would be quite unclear what he takes himself to be *doing* in deliberating. It would be unclear what reason was left to *call* it deliberation if its conclusion did not count as his making up his mind; or as we sometimes say, if it didn't count as his coming to *know* his mind about the matter." (Moran, 2001, pp. 94-5; emphasis in original.)10

The top-down element of this view (the element in virtue of which it is mixed, rather than bottom-up), is this: assuming that your judgments and attitudes are up to you *consists*, in part, in being willing to self-ascribe those thoughts and attitudes that are the outcomes of deliberation. On Moran's view, knowing your thoughts and attitudes is constitutive of being a rational agent:

> "When there is an attitude of mine that I cannot become aware *of* through reflection on its *object*, it suggests that the attitude is impervious to ordinary rational considerations relevant to the maintaining or revising of the attitude." (*Ibid*., p. 107. First emphasis added.)

For Moran, understood as a mixed theorist, if you are not prepared to self-ascribe the judgment that *p*, then you are not judging that *p* as an exercise of your rational agency.

I think we should reject this top-down element of Moran's account. Propositional knowledge of your own thoughts and attitudes is not constitutive of being a rational agent, in the sense

10 As is made clear by the rest of the book, and elsewhere (e.g. Moran, 2003), Moran's use of the term 'deliberation' is not meant to indicate any second-order, reflective reasoning. For Moran, purely first-order reasoning, that results in formation or revision of attitudes, without being *about* attitudes, counts as deliberation. So, coming to believe that *p* because it perceptually appears to you that *p* is an episode of deliberation.

that plays a role in Moran's argument.

Rational agency, for Moran, means acting and forming attitudes on the basis of reasons. This doesn't seem to require the subject to have any higher-order beliefs. A subject who lacks the concepts required for having such beliefs could nevertheless be a rational agent. It therefore does not require the subject to have any self-knowledge, *pace* Moran (*qua* mixed theorist).

An example of the exercise of rational agency, in Moran's sense, would be: judging that the foliage is yellow for the reason given by a visual experience as of its being yellow. In such a case, things strike the subject as being a certain way, and she accepts that that is how things are. This is something that a subject can simply do straightaway. There is no justificatory or enabling role, in this sort of activity, for a judgment or belief about the visual experience, or about the subject's attitudes. The subject need not have these concepts, in order to accept the content of a visual experience.[11]

Or consider Moran's own example, quoted above, about maintaining and revising an attitude. Maintaining or revising an attitude, in response to rational considerations, does not depend on awareness *of* the attitude. A subject who judges that the foliage is yellow may come to the view that the foliage is in fact green, when the conditions for seeing improve. This does not require any sort of higher-order reasoning. It requires that the content of the first judgment be stored in memory as true, and that it be made available for later reasoning and potential rejection.

Note also that Moran's argument applies to the practical as well as to the theoretical domain. But practical rational agency does not constitutively involve self-knowledge. Many subjects who are capable of such rational agency—of choosing one course of action over others on the basis of considerations that favour it—lack the conceptual ability to have a propositional attitude about their actions and reasoning. This includes very young children, and also, plausibly, some non-human animals (see Hurley, 2001).

Perhaps very young children and some non-human animals have some sort of non-conceptual proto-knowledge, or practical awareness, of their thoughts, attitudes and actions. If so, there is a sort of self-awareness that could perhaps be constitutive of rational agency. But that would not help the mixed theorist. The mixed theorist is a theorist who appeals to some feature of self-ascriptive judgments themselves in explaining the warrant for self-

11 I argued in Chapter 6 that a subject who judges must have some practical awareness of her goal in coming to judge, and a grasp of the appearance-reality distinction. These are capacities that help to ground willingness to make self-ascriptions, but they do not already involve self-knowledge. A subject could possess these capacities without yet having the conceptual repertoire to self-ascribe her judgments.

knowledge.[12]

As I argued in section 7.1, my own account allows for the possibility of rational agency without self-knowledge, while also allowing that rational agency may involve practical self-awareness of some kind.

It might be replied to my argument that Moran is appealing to a notion of reflective rational agency, rather than to mere ground-level rational agency. And reflective rational agency does indeed constitutively involve self-knowledge. But if that were his approach, Moran would face the same problems as Burge: this sort of reflective reasoning seems to presuppose self-knowledge, rather than explaining it.

The problem with Moran's argument is that he slides from the correct point that your reason-driven thoughts, attitudes and actions must in fact be up to you, to the claim that, in order to be a rational agent, you must assume that your thoughts, attitudes and actions are up to you. The plausibility of this latter claim depends on what is meant by "assume". If the claim is that you must have a propositional attitude, conceptually articulated, towards your reasoning and attitudes, then it is not plausible—as we have seen, rational agents need not have such attitudes. On the other hand, if the claim is that, in order to be a rational agent, you must have some sort of practical awareness of the point of reasoning, then it is plausible—after all, to reason properly, you must know how to reason. However, this latter sort of practical awareness does not constitutively involve self-knowledge of *particular* states and episodes. It is much more plausible that it helps to ground such self-knowledge.

These considerations tell against the mixed strategy generally, not just against Moran's view. A mixed strategy holds that self-knowledge derives its warrant from a basis that is itself explained, in part, in terms of self-knowledge. If this basis is *reasoning*, it must be a relatively modest sort of reasoning, and not reasoning of the reflective or critical sort, since, as we saw, the latter does not explain self-knowledge; but reasoning in the modest sense just doesn't constitutively involve self-knowledge.

The argument of this section, taken together with the earlier discussion of the top-down approach, suggests that the correct account of self-knowledge is bottom-up. Rational agency of the modest sort discussed by Moran does not constitutively require self-knowledge; rather, the conditions for such rational agency are among the conditions that ground self-knowledge. On the other hand, those more sophisticated sorts of reasoning that *do* require self-

[12] Again, Moran's considered view may in fact involve only the weaker claim. If so, he is not a mixed theorist and I have no very fundamental disagreement with him (although see the end of the present section)..

knowledge, such as Burgean critical reasoning, seem to be explained *by* self-knowledge, rather than explanatory of it.

Once we reject the claim that self-knowledge is necessary for rational agency, it seems unmotivated to tie self-knowledge so closely to the actual activity of reasoning. It seems more natural to hold that what grounds self-ascriptions of conscious episodes, including judgments, is what is involved in enjoying those conscious episodes. Rational agency is still important to self-knowledge, in at least two respects. Firstly, the warrant for self-ascriptions is explained in part by general capacities that the subject has as a rational agent. And secondly, rational agency will contribute to the account of self-knowledge of certain states and episodes insofar as it is involved in the occurrence of those states and episodes. By adopting this sort of view, we can account for the important connections that Moran and Burge have identified, without supposing that those connections are epistemically explanatory. That is the sort of view I have proposed.

7.3 Conclusion

Self-knowledge is grounded in the nature of conscious states and episodes, some of which involve rational agency, and depends on certain conditions of rational agency; self-knowledge in turn helps to ground critical reasoning, self-determination and personhood. That is the picture I have argued for. On this sort of account we can capture the important insights of Burge and Moran regarding the role of self-knowledge, without needing to use those insights in a transcendental argument in the epistemology of self-knowledge. We can offer a more satisfying account of why self-ascriptions are knowledge.

Let me finish by recapping the route this thesis has taken, and saying something about the broader philosophical project within which it is located.

I framed a two-part explanatory problem for the thesis: why are self-ascriptions of conscious states and episodes knowledge, and why do they have the special features of security, authority, first-person privilege, and so on? I argued that such self-ascriptions are indeed knowledge, with genuine epistemic credentials. And I argued that the attempt to explain the special features independently of those epistemic credentials fails. I then addressed the most historically salient epistemic account of self-knowledge—introspectionism—and argued that it mischaracterises the unique perspective each person has on his or her own conscious states and episodes. That suggested that self-ascriptions are based, not on experiences *of* conscious states and episodes, but on those states and episodes themselves.

I then took a detour to put in place an epistemological framework in which I could present my account. I argued that judgments can be knowledge in virtue of being based on reasons, even when those judgments do not meet certain internalist conditions on warrant.

I argued that, when a subject consciously entertains a content, she thereby has a reason to self-ascribe that content. Judgments made for that reason can be knowledge, I argued, partly in virtue of the subject's grasp of the first-/third-person distinction—this grasp being a practical capacity, fundamental to our thought, that is manifested in certain sorts of reasoning that are more primitive than self-knowledge. I claimed that self-knowledge of the *type* of conscious state or episode you are enjoying can be explained separately. I considered two cases, arguing that self-knowledge of perceptual experiences depends on the directness of perceptual experience, and that self-knowledge of judgment depends on the nature of judgment as a rational action. In each case, I claimed that the relevant explanatory feature of the type makes a difference to what it is rational for the subject to do, even though it is not itself a reason that the subject has.

Finally, I showed that my account can explain the central role of self-knowledge in our nature as persons. I argued that it is overall more satisfying than an account that uses that central role as a resource in explaining self-knowledge.

This is, at best, only a small fraction of a complete picture of self-knowledge. But it does suggest certain directions for further work.

The resources I drew on, in explaining self-knowledge of perception and of judgment, promise to contribute to the explanations of self-knowledge of various other types of conscious state and episode. It is plausible, I think, that there is a distinctive way in which states of affairs are given in memory, parallel to the distinctive way in which states of affairs are given in perception. This distinctive way can help explain how a subject knows she is *remembering* something to be the case. There are various types of mental action—visualising, calculating, directing your attention to something, and many others—that, like judgment, involve control, and knowledge of which depends in part on the nature of control.

A further task would be to explain self-knowledge of various attitudes, such as belief and desire, that are dispositional rather than occurrent. Some philosophers claim that self-knowledge of such attitudes depends on occurrence (or potential occurrence) of the conscious episodes that manifest them.[13] The claims of this thesis are congenial to such an approach. Perhaps you know that you *believe* that *p*, for example, by knowing about the

13 Peacocke (1998, 1999) is an example.

episodes of *judgment* that manifest that belief.

Similarly, perhaps other psychological states, such as emotions, are known partly through their manifestations in conscious thought and experience. If so, then the account of this thesis has something to contribute to the explanation of self-knowledge of such psychological states.

All of this, however, will still leave unexplained many other sorts of self-knowledge, including self-knowledge of sensations, and self-knowledge of the phenomenological aspects of our thought and experience. It will also leave us in need of a fuller account of how the epistemology of self-knowledge ties in with the metaphysics of consciousness, subjectivity and the mental.

Perhaps when these needs are met we will begin to have an adequate understanding of the unique relationship each person has with his or her own mind.

REFERENCES

Allen, C. (1997). “Animal Cognition and Animal Minds.” Philosophy and the Sciences of Mind: Pittsburgh-Konstanz Series in the Philosophy and History of Science. P. Machamer and M. Carrier. Pittsburgh, Pittsburgh University Press. **4**.

Anscombe, G. E. M. (1957). Intention. Oxford, Blackwell.

Armstrong, D. M. (1981). The Nature of Mind. Brighton, Harvester.

Ayer, A. J. (1936). Language, Truth, and Logic. London, V. Gollancz.

Bar-On, D. (2004). Speaking my Mind: Expression and Self-Knowledge. Oxford, Clarendon Press.

Bar-On, D. and D. C. Long (2001). "Avowals and first-person privilege." Philosophy and Phenomenological Research **62**(2): 311-335.

Bermúdez, J. L. (1998). The Paradox of Self-Consciousness. Cambridge, Mass., MIT Press.

Boghossian, P. (1989). "Content and Self-Knowledge." Philosophical Topics **17**: 5-26.

Bonjour, L. (1978). "Can Empirical Knowledge Have a Foundation." American Philosophical Quarterly **15**(1): 1-13.

Brewer, B. (1999). Perception and Reason. Oxford, Clarendon Press.

Broome, J. (1999). "Normative requirements." Ratio-New Series **12**(4): 398-419.

_____. (2007). "Does Rationality Consist in Responding Correctly to Reasons?" Journal of Moral Philosophy **4**(3): 349-74.

_____. (forthcoming). "Is Rationality Normative?" Disputatio.

Burge, T. (1979). "Frege and the Hierarchy." Synthese **40**: 265-81.

_____. (1993). "Content Preservation." Philosophical Review **102**(4): 457-488.

_____. (1996). "Our Entitlement to Self-Knowledge." Proceedings of the Aristotelian Society **96**: 91-116.

_____. (1999). "A Century of Deflation and a Moment about Self-Knowledge." Proceedings and Addresses of the American Philosophical Association **73**(2): 25-46.

_____. (2003). "Perceptual Entitlement." Philosophy and Phenomenological Research **67**(3): 503-548.

_____. (2004). “Postscript to 'Frege and the Hierarchy'.” Truth, Thought, Reason: Essays on Frege. Oxford, Oxford University Press.

Campbell, J. (1994). Past, Space, and Self. Cambridge, Mass., MIT Press.

Cassam, Q. (1997). Self and World. Oxford, Clarendon.

Chalmers, D. (2002). “The Content and Epistemology of Phenomenal Belief.” Consciousness: New Philosophical Essays. Q. Smith and A. Jokic. Oxford, Oxford University Press.

Chapuis, N. and C. Varlet (1987). "Short Cuts by Dogs in Natural Surroundings." Quarterly Journal of Experimental Psychology Section B-Comparative and Physiological Psychology **39**(1): 49-64.

Chisholm, R. (1981). The First Person. Minneapolis, University of Minnesota Press.

Chrisman, M. (submitted). "Expressivism, Truth and (Self-) Knowledge."

Clements, W. A. and J. Perner (1994). "Implicit Understanding of Belief." Cognitive Development **9**(4): 377-395.

Dancy, J. (2002). Practical Reality. Oxford, Oxford University Press.

Davidson, D. (1987). "Knowing One's Own Mind." Proceedings and Addresses of the American Philosophical Association **61**(441-58).

Dretske, F. (1999). "The Mind's Awareness of Itself." Philosophical Studies **95**(1-2): 103-124.

_____. (1981). Knowledge and the Flow of Information. Oxford, Blackwell.

Dunn, R. (1998). "Knowing What I'm About To Do Without Evidence." International Journal of Philosophical Studies **6**(2): 231-252.

Evans, G. (1982). The Varieties of Reference. Oxford, Clarendon Press.

Flavell, J. H. (1986). "The Development of Childrens Knowledge About the Appearance Reality Distinction." American Psychologist **41**(4): 418-425.

Flavell, J. H., E. R. Flavell, et al. (1987). "Young Children's Knowledge about the Apparent-Real and Pretend-Real Distinctions." Developmental Psychology **23**(6): 816-22.

Frankfurt, H. (1982). Freedom of the Will and the Concept of a Person. Free Will. G. Watson. Oxford, Oxford University Press.

Frith, C. D. (1992). The Cognitive Neuropsychology of Schizophrenia. Hove, L. Erlbaum.

Frith, C. D. and D. J. Done (1989). "Experiences of Alien Control in Schizophrenia Reflect a Disorder in the Central Monitoring of Action." Psychological Medicine **19**(2): 359-363.

Gazzaniga, M. S. (1995). "Principles of Human Brain Organization Derived from Split-Brain Studies." Neuron **14**(2): 217-228.

Geach, P. T. (1957). Mental Acts : their Content and their Objects. London, Routledge &

Paul.

Gertler, B. (2001). "Introspecting Phenomenal States." Philosophy and Phenomenological Research **63**(2): 305-328.

Ginet, C. (1968). "How Words Mean Kinds of Sensations." Philosophical Review **77**(1): 3-24.

Goldman, A. I. (1986). Epistemology and Cognition. Cambridge, Mass., Harvard University Press.

Greco, J. (2002). "Virtues in Epistemology." The Oxford Handbook of Epistemology. P. Moser. Oxford, Oxford University Press.

_____. (2004). "Virtue Epistemology." Stanford Encyclopedia of Philosophy Retrieved July, 2007, from http://www.seop.leeds.ac.uk/entries/epistemology-virtue/.

Haigh, S. N. and E. J. Robinson (in press). "What Children Know about the Source of their Knowledge without Reporting it as the Source." European Journal of Developmental Psychology.

Hurley, S. L. (2001). "Overintellectualizing the Mind." Philosophy and Phenomenological Research **63**(2): 423-431.

Husserl, E. (1982). Ideas pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book; General Introduction to a Pure Phenomenology. Dordrecht, Kluwer.

Jacobsen, R. (1996). "Wittgenstein on Self-Knowledge and Self-Expression." Philosophical Quarterly **46**(182): 12-30.

James, W. (1976). "The Experience of Activity." Essays in Radical Empiricism. Cambridge, Massachusetts, Harvard University Press.

Lehrer, K. (1974). Knowledge. Oxford, Clarendon Press.

Locke, J. (1700). An Essay Concerning Humane Understanding. London, Churchil.

Lycan, W. G. (1996). Consciousness and Experience. Cambridge, Mass., MIT Press.

_____. (2003). Dretske's Ways of Introspecting. Privileged Access: Philosophical Accounts of Self-Knowledge. B. Gertler. Aldershot, Ashgate.

Martin, M. F. (2001). "Epistemic Openness and Perceptual Defeasibility." Philosophy and Phenomenological Research **63**(2): 441-448.

_____. (1998). "An Eye Directed Outward." Knowing Our Own Minds. C. Wright, B. C. Smith and C. MacDonald. Oxford, Oxford University Press.

_____. (2002). "The Transparency of Experience." Mind & Language **17**(4): 376-425.

McDowell, J. (1998). “Response to Crispin Wright.” Knowing Our Own Minds. C. Wright, B. C. Smith and C. MacDonald. Oxford, Oxford University Press.

Menzel, E. W. (1973). "Chimpanzee Spatial Memory Organization." Science **182**: 943-5.

Moran, R. (2001). Authority and Estrangement: an Essay on Self-Knowledge. Princeton, Princeton University Press.

_____. (2003). "Responses to O'Brien and Shoemaker." European Journal of Philosophy **11**(3): 402-419.

_____. (2004). "Responses to Heal, Reginster, Wilson, and Lear." Philosophy and Phenomenological Research **69**(2): 455-72.

Nisbett, R. E. and T. D. Wilson (1977). "Telling More Than We Can Know: Verbal Reports on Mental Processes." Psychological Review **84**: 231-59.

O'Brien, L. (2003). "Moran on Agency and Self-Knowledge." European Journal of Philosophy **11**(3): 375-390.

_____. (2003). On Knowing One's Own Actions. Agency and Self-Awareness. J. Roessler and N. Eilan. Oxford, Oxford University Press.

_____. (2005). "Self-Knowledge, Agency and Force." Philosophy and Phenomenological Research **71**(3): 580-601.

_____. (2007). Self-Knowing Agents. Oxford, Oxford University Press.

O'Neill, D. K. and S. C. F. Chong (2001). "Preschool Children's Difficulty Understanding the Types of Information Obtained through the Five Senses." Child Development **72**(3): 803-815.

Peacocke, C. (1983). Sense and Content: Experience, Thought and their Relations. Oxford, Clarendon.

_____. (1992). A Study of Concepts. Cambridge, Mass., MIT Press.

_____. (1996). "Entitlement, Self-Knowledge and Conceptual Redeployment." Proceedings of the Aristotelian Society **96**: 117-58.

_____. (1998). “Conscious Attitudes, Attention and Self-Knowledge.” Knowing Our Own Minds. C. Wright, B. C. Smith and C. MacDonald. Oxford, Oxford University Press.

_____. (1999). Being Known. Oxford, Oxford University Press.

_____. (2001). "Does Perception have a Nonconceptual Content?" Journal of Philosophy **98**(5): 239-264.

_____. (2003). “Action: Awareness, Ownership and Knowledge.” Agency and Self-Awareness. J. Roessler and N. Eilan. Oxford, Oxford University Press.

_____. (2004). The Realm of Reason. Oxford, Clarendon Press.

_____. (2005). "'Another I': Representing Conscious States, Perception and Others." Thought, Reference and Experience: Themes from the Philosophy of Gareth Evans. J. L. Bermúdez. Oxford, Oxford University Press.

_____. (2007). "Mental Action and Self-Awareness (I)." Contemporary Debates in the Philosophy of Mind. J. Cohen and B. McLaughlin. Oxford, Blackwell.

_____. (forthcoming). "Mental Action and Self-Awareness (II): Epistemology." Mental Action. L. O'Brien and M. Soteriou. Oxford, Oxford University Press.

Perry, J. (1986). "Thought Without Representation." Proceedings of the Aristotelian Society (Supplementary Vol.) **60**: 263-83.

Plantinga, A. (1993). Warrant: the Current Debate. Oxford, Oxford University Press.

Pritchard, D. (2005). Epistemic Luck. Oxford, Clarendon Press.

Pryor, J. (2000). "The Sceptic and the Dogmatist." Nous **34**: 517-49.

_____. (2001). "Highlights of Recent Epistemology." British Journal for the Philosophy of Science **52**(1): 95-124.

Robinson, E. J., S. N. Haigh, et al. (forthcoming 2007). "Children's Working Understanding of the Knowledge Gained from Seeing and Feeling." Developmental Science.

Roessler, J. and N. Eilan (2003). Agency and Self-Awareness: Issues in Philosophy and Psychology. Oxford, Clarendon Press.

Ruffman, T., W. Garnham, et al. (2001). "Does Eye Gaze Indicate Implicit Knowledge of False Belief? Charting Transitions in Knowledge." Journal of Experimental Child Psychology **80**: 201-224.

Ryle, G. (1949). The Concept of Mind. London, Hutchinson's University Library.

Scanlon, T. (1998). What We Owe to Each Other. Cambridge, Mass., Harvard University Press.

Schaffer, J. (2005). "Contrastive Knowledge." Oxford Studies in Epistemology **1**: 235-71.

Searle, J. R. (1983). Intentionality : an Essay in the Philosophy of Mind. Cambridge, Cambridge University Press.

Shoemaker, S. (1968). "Self-Reference and Self-Awareness." Journal of Philosophy **65**(19): 555-67.

_____. (1994). "Self-Knowledge and Inner Sense." Philosophy and Phenomenological Research **54**(2): 249-314.

Smith, M. (1987). "The Humean Theory of Motivation." Mind **96**(381): 36-61.

Sosa, E. (1980). "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge." Midwest Studies in Philosophy **5**: 3-25.

Soteriou, M. (2005). "Mental action and the epistemology of mind." Nous **39**(1): 83-105.

Sperry, R. W. (1985). "Consciousness, Personal Identity, and the Divided Brain." The Dual Brain: Hemispheric Specialization in Humans. D.F. Benson and E. Zaidel. New York, Guilford.

Stampe, D. W. (1987). "The Authority of Desire." Philosophical Review **96**(3): 335-381.

Sturgeon, S. (2000). Matters of Mind : Consciousness, Reason and Nature. London, Routledge.

Velleman, J. D. (1989). Practical Reflection. Princeton, Princeton University Press.

_____. (1996). "Self to Self." Philosophical Review **105**: 39-76.

_____. (forthcoming). "What Good is a Will?" Action in Context. A. Leist and H. Baumann. Berlin/New York, de Gruyer/Mouton.

Wedgwood (2002). "Internalism Explained." Philosophy and Phenomenological Research **65**: 349-69.

Williams, B. A. O. (1978). Descartes: the Project of Pure Enquiry. Harmondsworth, Penguin.

_____. (1981). "Internal and External Reasons." Moral Luck. Cambridge, Cambridge University Press.

Wittgenstein, L. (1953). Philosophical Investigations. Oxford, Blackwell.

_____. (1980). Remarks on the Philosophy of Psychology. Oxford, Blackwell.

Wright, C. (1991). "Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention." Meaning Scepticism. K. Puhl. Berlin/New York, de Gruyter.

_____. (1998). "Self-Knowledge: the Wittgensteinian Legacy." Knowing Our Own Minds. C. Wright, B. C. Smith and C. MacDonald. Oxford, Oxford University Press.