13

KNOWING YOURSELF

What does philosophy have to say about substantial self-knowledge? A depressingly popular answer to this question among philosophers of self-knowledge is: very little. When I talked about substantial self-knowledge in chapter 3 I included knowledge of such things as one’s character, values, emotions, and abilities. You would think that these are the kinds of knowledge that are at issue in the oracle’s injunction to “know yourself”. If you aren’t a professional philosopher then substantial self-knowledge is probably what you think “self-knowledge” is, and you might be taken aback to discover that when philosophers write books and articles about self-knowledge they often start by acknowledging both the existence and importance of substantial self-knowledge but then say that they aren’t going to talk about it. Here’s a typical example of this approach from the preface of an otherwise excellent book on self-knowledge by Peter Carruthers:

Disappointingly for some readers, this book isn’t about the sort of self-knowledge that has traditionally been thought to be part of wisdom. This includes knowledge of one’s abilities and limitations, one’s enduring personality characteristics, one’s strengths and weaknesses, and the mode of living that will ultimately make one happy. Everyone allows that knowledge of this kind is hard to come by, and that having more of it rather than less of it can make all the difference to the overall success of one’s life. Moreover, it is part of common sense that those close to us may have a better idea of these things than we do ourselves. Instead, this book is about a kind of self-knowledge that nearly everyone thinks is easy to come by, almost to the point of triviality. This is the knowledge we have of our own current thoughts and thought processes, which are generally believed to be transparently available to us through some sort of introspection (Carruthers 2011: xi).

But if substantial self-knowledge can make all the difference to the overall success of one’s life, and is hard to come by then shouldn’t philosophy be interested in it? Aren’t philosophers usually interested knowledge that is hard to come by? Why spend so much time and energy on knowledge of our own current thoughts if this kind of knowledge is so easy to come by?

In chapter 4 I suggested the following answer to these questions: substantial self-knowledge tends to be neglected in philosophical discussions of self-knowledge because their focus is usually the epistemology of self-knowledge, and it’s widely assumed that substantial self-knowledge isn’t epistemologically interesting, even though it might be interesting in other ways. It isn’t epistemologically interesting because it isn’t epistemologically distinctive. Specifically, it lacks the epistemological privileges (first-person authority, immediacy, etc.) of what I have been referring to as trivial self-knowledge, such as your knowledge that you believe that you are wearing socks. From the standpoint of epistemology, your knowledge that you believe that you are wearing socks is more interesting, because more distinctive, than your knowledge of your own character. Substantial self-knowledge should be of interest to moral philosophers but not to epistemologists.

One way of developing this line of thinking would be to go back to the Asymmetry which came up in the last chapter. According to the Asymmetry, self-knowledge is different in kind and manner from knowledge of others. The way I know you believe you are wearing socks is by inference from what you say and do but that isn’t how I know that I believe I am wearing socks. However, when it comes to knowledge of my character what I have to go on is no different from what I have to go on when drawing conclusions about your character, or what you have to go on in drawing conclusions about my character. What I know about my character is, like what I know about your character, based on evidence, and the evidence can only be behavioural evidence: I go on what I say and do in my own case, just as I go on what you say and do when it comes to figuring out your character.

Suppose we describe the view that substantial self-knowledge is based on behavioural evidence as behaviourism about substantial self-knowledge, or just behaviourism for short. If behaviourism is correct then it’s understandable that philosophers like Carruthers don’t find substantial self-knowledge epistemologically interesting. It’s not that there are no interesting philosophical questions about knowledge based on behavioural or other evidence. If that were so there would be no problem of other minds or problem of the external world. However, the thought is that there are no special questions about substantial self-knowledge, questions that don’t also arise about any other evidence-based knowledge. It’s true that philosophers are often interested in knowledge that is hard to come by, but the sense in which knowledge of one’s own character can be hard to come by is no different from the sense in which many other kinds of empirical knowledge can be hard to come by.

Needless to say, I haven’t laid all this out simply in order to agree with it. Aside from any doubts one might have about the epistemological distinctiveness of trivial self-knowledge the main problem with the approach I’ve just outlined is that it presupposes a simple-minded and impoverished conception of substantial self-knowledge. Behaviourism about substantial self-knowledge is, to put it mildly, a fairly crude view. As well as failing to take account of subtle and interesting differences between different kinds of substantial self-knowledge, it paints a picture of substantial self-knowledge which simply doesn’t ring true. You can’t lump together all substantial self-knowledge and dismiss it with the remark that it’s all based on behavioural evidence. No doubt some of it is based on behavioural evidence but a lot of it isn’t. There is much more to be said, and philosophers need to say it.

So what’s the alternative to behaviourism? The obvious alternative is inferentialism of the sort I was discussing in the last chapter. Inferentialism says that inference is a basic source of self-knowledge for us. The inferences which give us intentional self-knowledge are, or include, theory-mediated inferences from internal promptings. Could it be that a lot of substantial self-knowledge also has its source in such inferences? Suppose it does. Wouldn’t that collapse the distinction between substantial and other self-knowledge? No. Going back to my list of ten characteristics of substantial self-knowledge (see chapter 3, pp. 35-6), it could still be the case that inferential knowledge of such things as one’s character and values more clearly satisfies more of these conditions than my knowledge that I believe that I’m wearing socks. For example, greater cognitive effort might be required to detect one’s own character, and there may be obstacles to knowing in this case are which are unlikely to be obstacles to knowing that I believe I am wearing socks. Inferentialism about substantial self-knowledge doesn’t collapse the distinction between substantial and other self-knowledge, though it does support the suggestion that the difference is only one of degree.

Should we be inferentialists rather about substantial self-knowledge, and are there any forms of substantial self-knowledge that inferentialism can’t handle? These are the questions I want to address in this chapter, and the way I propose to address them is to take a close look at three examples of substantial self-knowledge. These relate, respectively, to knowledge of one’s character, knowledge of one’s values and knowledge of one’s emotions. After making the case for inferentialism in connection with each of these forms of self-knowledge I will then examine the following objections to this approach:

1. Inferentialism makes substantial self-knowledge out to be more of an intellectual achievement than it really is. Especially when it comes to knowledge of one’s own emotions, the worry is that inferentialism is a form of what Nussbaum refers to as ‘intellectualism’, and that intellectualism is no good.
2. Inferentialism ignores the extent to which self-attributions of character, values and emotions constitute one’s character, values and emotions. Inferentialism makes us out to be more passive in relation to such things than is actually the case.
3. There are major sources of substantial self-knowledge which inferentialism can’t account for because it fails to acknowledge the role of insight in the acquisition of substantial self-knowledge. For example, novels and films can give you insights into your own character, values and emotions, and the source of your substantial self-knowledge when that happens isn’t inference.

All three objections have something going for them, and a defensible inferentialism needs to be able to accommodate the insights they embody. This brings me to the position I want to defend. I’d like to suggest that when it comes to accounting for substantial self-knowledge, a sophisticated and inclusive inferentialism is the best bet. There are, as I’ve said, interesting differences between different kinds of substantial self-knowledge but inferentialism provides the most fruitful and flexible framework for thinking about these differences, as well as about what different forms of self-knowledge have in common. But even if you aren’t convinced by inferentialism I do hope by the end of this chapter to have convinced you of something else: as well as being something that humans actually care about, substantial self-knowledge, in its different forms, is much more philosophically interesting than you might have thought on the basis of the dismissive attitude of so many philosophers. If, as a philosopher and a human being, you are interested in self-knowledge then you really should be interested in substantial self-knowledge; there is no excuse for only trying to account for trivial self-knowledge and its supposed privileges.

Knowledge of one’s own character is my initial test case, and the first problem we run into is that there are serious philosophers who think that there is no such thing as character or character traits. If there is no such thing as character then there is no such thing as knowledge of one’s character, in which case there is nothing to be said about how one knows one’s own character. All that needs to be explained is why so many people still believe that there is such a thing as character. This isn’t really the place for a full-blown discussion of scepticism about the existence of character traits. At the same time, the writings of character sceptics like John Doris and Gilbert Harman can hardly be ignored, given how much I have been making in this book about knowledge of one’s character as a form of substantial self-knowledge. So before turning to the epistemological issues, something needs to be said about character scepticism. Apart from anything else, character scepticism turns on a certain view of what character is, and the nature of character is something we need to get clear about anyway.

As Harman defines them, character traits are relatively long-term, stable dispositions to act in distinctive ways. We ordinarily suppose that people differ in character and also that a person’s character helps explain at least some things he or she does. Harman thinks ‘there is no reason at all to believe in character traits as ordinarily conceived’ (2000: 223), and that the way to explain our behaviour is in terms of situational factors rather than character. He bases this view on the work of social psychologists who argue that observers wrongly infer that actions are due to distinctive character traits of agents rather than relevant aspects of the situation. For example, in the notorious Milgram experiment people were asked to administer increasingly powerful electric shocks to unseen “victims” who gave incorrect answers to various questions, or who refused to answer. All subjects, regardless of individual character, were willing to go to at least 300 volts, and fully 65% were prepared to deliver the maximum shock of 450 volts, past the label “Danger: Severe Shock”. Why was this? Not because of some shared character defect but because of the specifics of the situation. Harman concludes that the attribution of character is explanatorily redundant and therefore unjustified.

It’s certainly plausible that in the extreme circumstances of the Milgram experiment it isn’t easy to explain subjects’ actions by reference to their character traits, unless destructive obedience is a character trait. But the fact that character traits don’t explain everything we do doesn’t mean that it isn’t right to explain some of what a person does by reference to his or her character. It’s also worth pointing out that character traits are not just dispositions to act in certain ways. Consider fastidiousness as a character trait. The dictionary definition of ‘fastidious’ is ‘very careful about accuracy and detail’ and ‘concerned about cleanliness’. Synonyms include ‘meticulous’, ‘fussy’, ‘pernickety’, ‘overcritical’ and ‘difficult to please’. Being concerned about accuracy and detail or difficult to please aren’t just, or even primarily, dispositions to act a certain ways; fastidiousness is the underlying state of mind. A fastidious person is one who acts as he acts because he thinks in a certain ways, cares about certain things, and has particular desires and emotions. If you are fastidious then you tend to be bothered by things that wouldn’t bother you if you weren’t fastidious. Moreover, you can be disposed to act as a fastidious person would act even you aren’t fastidious; perhaps you have other motives for being disposed to act in these ways. A fastidious person is not just someone who behaves fastidiously, but one whose fastidious behaviour is a reflection of, and prompted by, a particular set of concerns, desires and emotions.

Now consider a fictional character we can call Woody. Here are some things we know about Woody: he is meticulous in his work, and his office and desk are always tidy. When he goes to bed he folds his clothes carefully, and he is disturbed by domestic disorder. He is in perpetual conflict with his teenage children over the state of their bedrooms. They are tidy by normal teenage standards but Woody is overcritical and nit-picking about even trivial lapses. Suppose we now wonder: when Woody is at work why does he spend so much time filing and labelling documents? The obvious answer is: because he is so fastidious. This looks like a perfectly reasonable and indeed informative explanation of his behaviour in terms of one of his character traits. If you don’t know Woody then I’ve just told you something which should make his behaviour intelligible to you. His behaviour would still be intelligible to you, but in a different way, if I told you that Woody files and labels because he is afraid of his boss. On a given occasion there might be situational factors that help explain his behaviour but you are unlikely to get very far if you attempt to explain all his complaining, tidying and nit-picking by reference to such factors. After all, this isn’t a Milgram-type scenario. We are trying to explain a pattern of behaviour in a range of different contexts, and we would be depriving ourselves of a valuable explanatory resource if we don’t say the obvious thing about Woody’s filing and labelling: he does it because he is fastidious. Reference to Woody’s character isn’t explanatorily redundant.

Assuming there is such a thing as character, the next question is: how do you know your own character traits? If character traits are just dispositions to act in certain ways then it’s understandable that behavioural evidence should be regarded as the only basis on which it is possible for one to know one’s character traits. What’s more, the basis on which I’m able to discern my character traits would then be no different from the basis on which you are able to discern them. The reason that, in reality, we aren’t stuck with behaviourism is that, as I have suggested, character traits aren’t merely dispositions to act. If they are dispositions at all they are ‘dispositions to have prevailing desires and emotions of particular sorts’(Velleman 2007: 243), though even this is an over-simplification. Consider Woody again. How does he know he is fastidious? Since being fastidious is partly a matter of what you care about and what bothers you, for him to know that he is fastidious he would have to know, among other things, what he cares about and what bothers him. It would be strange to suppose that he only knows what he cares about or what bothers him on the basis of behavioural evidence. But it also wouldn’t be right to say that he knows these things on the basis of no evidence. So the challenge is to give an account of Woody’s self-knowledge which avoids both extremes.

Here is how Woody might come to know that he cares about such things as tidiness and attention to detail, and that he is bothered by their absence: when he imagines the state of his teenagers’ bedrooms he is conscious of feeling a mixture of dismay and irritation. The same mixture of dismay and irritation is prompted by the recollection that he didn’t have time to tidy his desk when he finished work yesterday, and he is conscious of a desire to put things right as soon as possible. When he thinks about what needs to be done tomorrow, he focuses on what he sees as the need to restore order. He knows that his work colleagues aren’t nearly as meticulous as he is, and is conscious of thinking thoughts along the lines of “if you want something done right, do it yourself”. On the basis of his thoughts, imaginings and emotions Woody is in a position to conclude that he cares about cleanliness and attention to detail. In ‘The Importance of What We Care About’, Frankfurt writes that a person who cares about something ‘identifies himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits depending upon whether what he cares about is diminished or enhanced’ (1998: 83). Saying that Woody identifies himself with tidiness might seem a little excessive, but he is certainly ‘vulnerable’ to its absence; he is vulnerable to it in the sense that he is disturbed by it.

As I have described it, Woody’s knowledge that he cares about tidiness and attention is inferential. In the terminology of chapter 11, he infers from various ‘internal promptings’ that he cares about these things, in a way that is not very different from the way that Lawlor represents Katherine as inferring she desires another child. This makes Woody’s knowledge that he is fastidious doubly inferential. Just because he knows that he cares about tidiness and attention to detail it doesn’t follow that he knows, or is even in a position to know, that he is fastidious; he might not have the concept fastidious, or it might never cross his mind that he is fastidious. Even if it does cross his mind, he might wonder whether he cares enough about tidiness and attention to detail to make him fastidious. Or, in his more reflective moments, he might wonder whether he is merely fastidious or has obsessive compulsive disorder. We can imagine Woody running through these possibilities and finally concluding that he is indeed a fastidious person. Assuming that this conclusion is justified on the basis of the evidence that is available to him, it counts as a piece of hard-earned self-knowledge. He infers that he has certain psychological characteristics and infers his character from these characteristics.

It’s worth emphasizing that Woody’s self-knowledge is substantial in my terms; it is fallible, there may be a range of obstacles to its acquisition, it tangles with his self-conception and is open to challenge. In addition, his self-knowledge is corrigible, indirect, and requires cognitive effort. He can’t acquire it by using the Transparency Method but relies instead on evidence. I have represented Woody’s evidence as psychological, which isn’t to say that there isn’t also a role for behavioural evidence. He might appeal to behavioural evidence in support of his self-attribution of fastidiousness, but in the nature of the case his evidence isn’t mainly or primarily behavioural. And that is why there remains something of an asymmetry between how Woody knows he is fastidious and how someone else knows that Woody is fastidious. His friends can only go on what he says and does but Woody also has access to his internal promptings. He still has to interpret these promptings if they are to give him self-knowledge, and he can also make them available to others by telling them about it. Nevertheless, his self-knowledge is different from the knowledge that other people have of his character.

Turning, next, to knowledge of one’s values, in chapter 3 I talked about whether and how one knows that one is not a racist. In that discussion, I emphasized the role of behaviour, on the basis that not being a racist isn’t just a matter of espousing racial equality. I said that it is also a matter of whether you put your money where your mouth is, that is, a matter of how you behave with people from other races. Although this supports the idea that the evidence that bears on whether you are a racist is behavioural evidence, it ignores key questions about the values that underpin what you say and do. It’s easy to imagine someone whose behaviour and dispositions to act are unimpeachable but who is still an instinctive racist. An instinctive racist is someone who, as Taylor puts it, ‘only feels a sense of moral solidarity with members of his own race’ (1985: 61). Figuring out whether you are an instinctive racist isn’t just a matter of reflecting on your behaviour. It’s also a matter of how you think and feel, so here is another case in which inference from internal promptings plays a key role in the acquisition of substantial self-knowledge. As Taylor points out, it’s all too easy to imagine an instinctive racist saying he knows that race shouldn’t make any difference but that ‘he does not feel it’ (ibid.).

Another example: in one of his early diaries the British Labour politician Tony Benn asks “Am I a socialist?”. The answer to the question might have been obvious to him in later years but when he asked the question the answer wasn’t obvious to him. It’s natural to view Benn’s question as a question about his values, and not just a question about his beliefs, about whether he believed a list of propositions which expressed the core tenets of socialism. To be a socialist in the relevant sense is to have certain values and concerns, and to think like a socialist, that is, to be disposed to analyse and explain historical and political events along socialist lines. If, like David Lewis, we say that valuing something is desiring to desire it then someone who has socialist values is someone who desires to desire such things as equality and social justice. To know that you are a socialist would be to know your relevant second-order desires, and that’s no easy task. Nor is it a straightforward matter to determine whether you ‘think’ like a socialist. However, what does seem clear is that you aren’t going to be able to determine your values just on the basis of behavioural evidence. It’s more a matter of interpreting your patterns of desire and thought on the basis on an understanding of what is, and is not, relevant to having certain values rather than others.

Does this mean that behavioural evidence has no part to play in coming to know your values? That obviously depends on how we understand the notion of ‘behavioural evidence’. The example of a person figuring out whether he is a socialist is trickier than the example of a person figuring out he is a racist. The difference is that we have a much clear notion of racist behaviour than of ‘socialist behaviour’. There is, of course, the way you live your life, and that might be what putting your money where your mouth is might come to in this case. But even this isn’t a straightforward matter. Consider Friedrich Engels, described by one recent biographer as a ‘raffish, high-living, heavy-drinking devotee of the good things in life’ (Hunt 2009). It’s hard to make the case that Engels lived his life in the way that a socialist might be expected to live his life but even harder to make the case that he wasn’t a socialist. The whole idea of knowing your own or anyone else’s values on the basis of behavioural evidence is so problematic because the relationship between a person’s values and his ‘behaviour’ is much more complicated than behaviourism suggests. It should go without saying, however, that in so far as you do genuinely know your values, your self-knowledge is about as ‘substantial’ as self-knowledge can be, and that inference is still the means by which you acquire it.

My last example of substantial self-knowledge is knowledge of one’s emotions. There are many different emotions, and little hope of accounting for all emotional self-knowledge in the same way. One specific form of emotional self-knowledge that has attracted philosophical attention is knowledge of one’s own love. In ‘Love’s Knowledge’, Nussbaum discusses what might be called Proust’s epistemology of love. Nussbaum interprets Proust as suggesting in Remembrance of Things Past that ‘knowledge of the heart must come from the heart’ (1990: 262). The contrary view, which she calls ‘intellectualism’, says that knowledge of whether one loves another person can best be attained by a ‘detached, unemotional, exact intellectual scrutiny of one’s condition, conducted in the way a scientist would conduct a piece of research’ (ibid.). But from an inferentialist perspective this is a false dichotomy: the sense in which knowledge of the heart comes from the heart is that it is, or can be, derived from impressions such as those which, in Proust’s novel, tell Marcel that he loves Albertine. The specifics means by which love’s knowledge is derived from such impressions is inference or self-interpretation. These are ‘intellectual’ processes which only deliver emotional knowledge given the appropriate background ‘theory’. If inferentialism is ‘intellectualism’ it isn’t the pernicious intellectualism which Nussbaum caricatures. But before getting to that, let’s take a closer look at the passage from Proust which Nussbaum discusses.

The passage begins with an announcement: “Mademoiselle Albertine has gone!”. Just before hearing the announcement Marcel had assumed that Albertine’s departure would be a matter of indifference to him, or even what he wished for after comparing ‘the mediocrity of the pleasures’ she afforded him with the ‘richness of the pleasures’ she prevented him from realizing. The news of Albertine’s departure changes everything, and Marcel observes:

How much further does anguish penetrate in psychology than psychology itself! A moment before, in the process of analysing myself, I had believed that this separation without having seen each other again was precisely what I wished…. I had…. concluded that I no longer loved her. But now these words: “Mademoiselle Albertine has gone”, had produced in my heart an anguish such that I felt I could not endure it much longer…. I had been mistaken in thinking that I could see clearly into my own heart (III. 425-6).

In Nussbaum’s terminology, what led Marcel astray to begin with was his intellectualism, his conviction that, when it came to knowledge of his own heart, he was ‘like a rigorous analyst’, leaving nothing out of account. Now he knows better, and it is his anguish which reveals the truth to him. His newly acquired self-knowledge – that he loves Albertine- is self-knowledge ‘through’ suffering.

Nussbaum tries to make sense of Proust by bringing in the Stoic notion of a cateleptic impression. Cateleptic impressions are impressions which, by their own internal character, certify their own veracity and ‘drag us to assent’ (1990: 265). In these terms Marcel’s anguish is a cateleptic impression. It isn’t simply a route to knowing, it is knowing:

The suffering is itself a piece of self-knowing. In responding to a loss with anguish, we are grasping our love. The love is not some separate fact about us that is signalled by the impression; the impression reveals the love by constituting it. Love is not a structure of the heart waiting to be discovered (1990: 265-6).

Marcel’s love for Albertine is constituted by his suffering in the sense that, ‘while he was busily denying that he loved her, he simply was not loving her’ (1990: 268); love denied isn’t exactly love. Intellectualism tells us that our passions and feelings are ‘unnecessary to the search for truth about any matter whatever (1990: 262-3) but love’s knowledge is a problem for this view. For ‘to try to grasp love intellectually is a way of not suffering, not loving – a practical rival, a stratagem of flight’ (1990: 268-9).

There’s no denying the seductiveness of Nussbaum’s account of love knowledge but is it any good? Consider her insistence that Marcel’s suffering isn’t just a route to knowing. On an inferentialist reading, that is precisely what his suffering is. Marcel’s anguish does not itself constitute knowledge of anything but it can be the basis of self-knowledge. For a start, love is only one possible explanation of Marcel’s anguish, and there are plenty of others. For example, anguish can also be induced by the departure of a person on whom one is dependent but doesn’t love. Perhaps the two kinds of anguish are different, but what is to prevent one kind of anguish from being mistaken for another? When Marcel concludes, on the basis of his suffering, that he loves Albertine it is because he interprets his suffering as signalling love for Albertine. If his interpretation is correct then he is in a position to infer, and thereby know, that he loves Albertine. The inference is mediated by an interpretation of his suffering that is grounded in his understanding of the relationship between this kind of suffering and romantic love. His route to self-knowledge here is inference, whereas the basis of his self-knowledge is suffering. Here ‘basis’ means ‘evidence’, but suffering is obviously not the only evidence of love; there is also joy. As for love not being a structure of the heart waiting to be discovered that is exactly what love can be. What Marcel discovers is a pre-existing emotional fact about himself, and it’s not an objection to this view that he didn’t believe he loved Albertine before he heard the announcement. It’s no more plausible that love denied is not love than that jealousy denied isn’t jealousy or that depression denied isn’t depression.

This form of inferentialism about emotional self-knowledge doesn’t have to say that feelings are redundant in the search for love, any more than it has to say that feelings have nothing to do with love itself. To love someone is, among other things, to be disposed to feel a certain way about them, and that’s why actually feeling that way is such good evidence that you love them. These feelings do not give us access to truths that could in principle be grasped in other ways, for example by intellect alone, but it doesn’t follow that they constitute love’s knowledge or that this knowledge isn’t inferential. This is the crux of inferentialism’s middle way between extreme intellectualism and Proust’s account: feelings of joy, anguish, or suffering provide us with access to truths about ourselves which may not be accessible in other ways, but the access is inferential, and relies on the role of feelings as signs of one’s underlying emotional state.

I’ve now examined three examples of substantial self-knowledge and interpreted each example along inferentialist lines. In chapter 3 I gave several other examples of substantial self-knowledge, and the inferentialist’s hypothesis – which admittedly can’t be conclusively proved without going through all of them - is that in every case you obtain self-knowledge by means of a theory-mediated inference from internal evidence, sometimes supplemented by behavioural evidence. You have an implicit theory or background understanding of what it would be for you to have a particular attribute (character trait, value, emotion, or whatever) and infer that you have the attribute in question because you interpret the evidence available to you as evidence that you have it. The Asymmetry, such as it is, is an asymmetry between the evidence available to you and to others, not an asymmetry between inferential and non-inferential knowledge. In one sense Marcel’s knowledge that he loves Albertine is different from your knowledge that he loves Albertine- your knowledge isn’t derived from suffering- but neither of you knows other than on the basis of evidence. Different kinds of substantial self-knowledge are supported by, and depend on, different kinds of evidence, but in none of the examples I have considered does behaviourism look remotely plausible.

Where does this leave my insistence that there are subtle and interesting differences between different forms of substantial self-knowledge? If all substantial self-knowledge turns out to be inferential what becomes of these differences? The answer to this question is much the same as the response I gave to the earlier suggestion that inferentialism about substantial self-knowledge collapses the distinction between substantial and other self-knowledge. The answer is that there is plenty of scope for exploring differences between different varieties of self-knowledge in an inferentialist framework; indeed, inferentialism provides a fruitful and flexible framework for thinking about these differences. For example, as soon as you think of different kinds of substantial self-knowledge as supported by different kinds of evidence you need to explain in more detail which types of evidence are relevant in each case, and why. It’s one thing to say that substantial self-knowledge is fallible and corrigible but it’s natural to suppose that you are less fallible and more authoritative about your emotions than about your character. The obstacles and challenges are different in each case, as is the level of what I’ve been calling cognitive effort. There is endless scope for further reflection and analysis, and it is for philosophy to engage in this further reflection and analysis. You would have to have a wretchedly impoverished conception of philosophy to think that it has nothing of interest to say about these issues.

Before moving on to consider some potential objections to inferentialism, there is one further implication of this approach that needs to be brought out. I’ve suggested that, as far as inferentialism is concerned, the difference between substantial and trivial self-knowledge is a difference in degree. The difference between your knowledge that you are fastidious and your knowledge that you believe you are wearing socks is the difference between obviously and unobviously inferential knowledge rather than the difference between inferential and non-inferential knowledge. Your knowledge is manifestly and consciously inferential in the one case but not in the other. If this is what we say then we need to be careful how we bring in the notion of substantial self-knowledge. The natural way to do this is by example but examples are easily misunderstood. From the fact that love is an emotion, and that Marcel’s knowledge of his love for Albertine is substantial self-knowledge, it’s easy to conclude that knowledge of one’s emotions is always substantial. But do we really want to say that a parent knows that she loves her child by inference or that it requires cognitive effort to determine that you love your child? These rhetorical questions open up the possibility that, in principle, knowledge of some of one’s emotions might belong on the less substantial end of the spectrum. If that is the case, then it’s probably best to avoid introducing the notion of substantial self-knowledge by giving examples. No doubt knowledge of one’s emotions, values and abilities is often or even typically substantial, but inferentialism can leave open the possibility that sometimes it isn’t. If it turns out that something like knowledge of one’s character can only ever be substantial then that’s another interesting difference between varieties of self-knowledge.

To bring matters in this chapter to a close, let’s go back to the three serious-sounding objections to inferentialism listed above. The first was that inferentialism makes substantial self-knowledge out to be more of an intellectual achievement than it really is. I have already talked about Nussbaum’s concerns about intellectualism but there is a much simpler way of explaining what the problem is: substantial self-knowledge is something that is available to most mature adults; it isn’t only available to the educated or philosophically sophisticated. In that case, it might seem totally inappropriate to think of substantial self-knowledge as based on theory-mediated inferences from internal and other evidence, especially if the inferences are supposed to be conscious and explicit. Ordinary people don’t have the ‘theories’ which this account requires.

There is a more and a less concessive response to this concern. The more concessive response would be to point out that it’s important not to read too much into talk of theory-mediated inferences. Someone who knows, and therefore believes, he is generous or in love has to have some conception of what it would be to be generous or in love. In this sense they must have a ‘theory’ of generosity or romantic love but the theory doesn’t have to be all that sophisticated, and their grasp of the theory may well be implicit. Perhaps they would be hard pushed to articulate their theory of romantic love but that doesn’t mean that they aren’t in fact relying on a ‘theory’ in taking themselves to be in love. To have substantial self-knowledge you must have justified beliefs about yourself, and that means understanding what is evidence for what. This leads naturally to the less concessive response. This response bites the bullet in insisting that the acquisition of substantial self-knowledge is an intellectual achievement, and that your level of intellectual sophistication does affect the quantity and quality of substantial self-knowledge that is available to you. Obviously no one wants to be accused of elitism but fact is that substantial self-knowledge isn’t evenly distributed across the population. For example, not every fastidious person knows that he is fastidious; you only know that you are fastidious if you know what fastidiousness is, and what counts as evidence of fastidiousness. Not knowing these things is a barrier to knowing that you are fastidious, and if you do know you are fastidious that is only because you have and use the necessary intellectual resources. In this case only conscious reflection of quite a sophisticated sort can deliver self-knowledge. Of course this isn’t to say that a Rolls Royce intellect guarantees substantial self-knowledge. However clever or knowledgeable you are there is always the possibility of repression or self-deception. Still, the more you know, the more you are in a position to know about yourself.

The next objection to inferentialism says that it makes substantial self-knowledge out to be more passive than it really is, and underestimates our role in creating or constituting the attributes of which we have substantial self-knowledge. There are better ways of making this point than to go along with Nussbaum’s view that denying that one has a particular emotion amounts to one not having that emotion. For example, in his classic paper ‘Self-Interpreting Animals’, Charles Taylor observes that: ‘human life is never without interpreted feeling; the interpretation is constitutive of the feeling’ (1985: 63). He gives the example of remorse as a feeling or emotion which involves a certain level of articulation. A feeling of remorse ‘implies our sense that our act was wrong’ (ibid), even if we can’t initially articulate what is wrong about what we have done. In such cases we may try to understand further, and if we succeed our feeling may change: ‘the remorse may dissipate altogether, if we come to see that our sense of wrong-doing is unfounded, or it may alter in other ways, as we come to understand what is wrong’ (ibid.). The articulation transforms the emotion, which means that we aren’t mere passive recorders of our emotions. There is both an element of creation and an element of discovery, and the toughest challenge for an account of emotional self-knowledge is to do justice to both elements. Or so Taylor implies.

A lot of what Taylor says sounds right, but is that a problem for inferentialism? There would be a problem if inferentialism has to deny that interpretations or articulations play any part in constituting our emotions but that’s plainly not the case. Here is how inferentialism can accommodate Taylor’s insight: suppose you are trying to figure out how you feel about something you have just done, and aren’t sure whether it is remorse or embarrassment. You ask yourself whether what you did was wrong and your initial thought is that it was. You see that your feeling incorporates your sense that what you did was wrong and you conclude that what you are feeling is remorse. To conclude that what you feel is remorse is to interpret your feeling as one of remorse, but you aren’t making it up as you go along. You take it that what you feel is remorse because of your sense that what you did was wrong; to put it another way, you infer from your sense that what you did was wrong that what you feel is remorse. So far so good for inferentialism.

Next, you ask yourself whether on reflection what you did really was wrong, and the more you think about it the less convinced you are that it was. As a result, your feeling alters. The fact that you no longer have the sense that what you did was wrong implies that what you feel is no longer remorse. Maybe you don’t feel anything at all, or what you feel is quite different. Again, inferentialism has no problem with any of this. Your evolving sense that what you did wasn’t wrong both makes it the case that what you now feel isn’t remorse and is evidence from which you can infer that what you feel isn’t remorse but something else. There is no problem for inferentialism because your thoughts about the moral status of your action can play a dual role: they can both alter an emotion and reveal to you what emotion it is. Just because you interpret an emotion a certain way on the basis of your thought, it doesn’t follow that those thoughts can’t partially constitute your emotion. By the same token, just because a particular thought contributes to making your emotion the particular emotion it is, it doesn’t follow that you having that thought isn’t evidence from which you can correctly infer that you have just that emotion. Inferentialism doesn’t imply that we are merely passive recorders if our emotions if what that means is that what we think can’t shape what we feel. In fact, it is because what we think can and does shape what we feel that it is evidence of what we feel.

The last objection to inferentialism says that there are major source of substantial self-knowledge it doesn’t account for because of its failure to acknowledge the role of insight in the acquisition of substantial self-knowledge. Reading a novel or seeing a movie can give you an insight into your own character and emotions but knowledge by insight isn’t inferential. Here’s an unflattering example: suppose that I’ve never thought of myself as a cold fish but I read Anna Karenina and can see myself in Karenin. He is unfeeling, unromantic, and cold. These are not epithets I would willingly apply to myself but now I see that temperamentally I’m not very different from him. The chances are that I will be dismayed by this realization, but perhaps I also find myself identifying with Karenin. Whatever my reaction, it is tempting to claim that this new insight into my character is a piece of substantial self-knowledge, and also that the source of my self-knowledge here is the novel itself. I don’t infer that I am like Karenin; I see that I am like him. The question is whether inferentialism can account for this.

There are several things to say about this. First, it’s false that Anna Karenina is the source of my self-knowledge. It isn’t the novel that tells me that I am cold but my reflection on the novel (cf. Hetherington). What is more, this reflection presupposes self-knowledge. I can only see myself in Karenin because I notice how we resemble each other, and I can only notice how much I resemble Karenin if I already know something about how I am. What reading the novel does is to make certain aspects of how I am salient to me and help me to conceptualize these aspects. If I see Karenin as cold and recognize that I am in the relevant respects like him then the inescapable conclusion is that I’m a cold person. This inescapable conclusion is the conclusion of an inference whose premises include statements about me and about Karenin. What the inference provides me with is still ‘self-insight’ but self-insight, like ordinary seeing, is inferential: I see that I’m like Karenin because I infer that I am like him.

Here’s another way of reaching the same conclusion: suppose, somewhat improbably, that I am aware of identifying with Karenin as I read Anna Karenina. There is a lot to be said about what is involved in identifying with a literary character but suppose we take this notion for granted for present purposes. Just because I identify with Karenin it doesn’t follow that I am like him, let alone that I know that I am like him. Still, I might wonder what the fact that I identify with him tells me about myself. It’s hard to get away from the notion that identifying with a character like Karenin is somehow self-revealing but it’s only self-revealing if I reflect on my identification with him and have a plausible story to tell about what my identification with Karenin reveals. If I have such a story then I can perhaps draw certain conclusions about my own character but any such conclusions are inferred from my reactions to the character of Karenin. Once again, what gives me self-knowledge is a theory-mediated inference.

Even if I’m wrong about this and literature turns out to be a sui generis source of non-inferential substantial self-knowledge it still wouldn’t follow that inferentialism is no good; it wouldn’t follow that inference isn’t a basic source of substantial self-knowledge even if there are other sources of substantial self-knowledge. However, as I say, I don’t believe that novels are a sui generis source of non-inferential substantial self-knowledge. Let’s agree, then, that inferentialism is in good shape, both relation to substantial and insubstantial self-knowledge, and proceed on that basis. None of the three objections to inferentialism I have discussed is successful, and there is no obvious alternative to inferentialism. If you still think that philosophy needn’t bother with substantial self-knowledge, or that it doesn’t have anything interesting or useful to say about the epistemology of substantial self-knowledge, then your conception of philosophy is very different from mine.

Where do we go from here? As well as delivering an account of the epistemology of self-knowledge one might also expect philosophy to have something to say about is value. It would hardly be worth spending so much time thinking about self-knowledge unless there are reasons for thinking that it is valuable. Whether there are such reasons will be the topic of chapter 14. However, we aren’t quite done with the epistemology of self-knowledge because it’s no good having an account of self-knowledge unless you also have an account of self-ignorance. Self-ignorance is a genuine phenomenon, and something that humans go to a great deal of trouble - and expense - to overcome. In light of that fact, it would be reassuring if philosophy has answers to some obvious questions about self-ignorance: for example, how prevalent is it, what are its main sources, and to what extent can it be overcome by us? I think that these are questions to which inferentialism suggests answers, so now would be a good time to say some more about self-ignorance.