11

SELF-KNOWLEDGE AND INFERENCE

It’s time to cut to the chase. So far in this book I’ve spent a lot of time criticizing a range of standard approaches to self-knowledge but the question raised by these criticisms is: do you have any better ideas? Beating up on other views is easy but the real challenge is to replace them with something better. In this chapter I’m going to defend what I’ve been calling inferentialism about intentional self-knowledge, the view that knowledge of our own beliefs, desires, and other attitudes is first and foremost a form of inferential knowledge. In Chapter 1 I identified the following sources question about self-knowledge:

(SQ) What are the sources of self-knowledge for humans?

Inferentialism says that inference is a key source of intentional self-knowledge for humans. Whether inferentialism regards any of our intentional self-knowledge as non-inferential is a question I will come back to later in this chapter. In the next chapter, I’ll answer the sources question in relation to substantial self-knowledge. Again, inference will turn out to play a key role.

I think it’s fair to say that inferentialism hasn’t exactly been a popular approach to self-knowledge is recent years. Many currently influential discussions of self-knowledge start out with a statement to the effect that intentional self-knowledge is normally ‘immediate’, that is, non-inferential. Philosophers who defend inferentialism – Ryle is usually mentioned in this context- are then berated for defending a patently absurd view. The assumption that intentional self-knowledge is normally immediate – this is the immediacy premise that came up in the last chapter- is rarely defended; it’s just seen as obviously correct. In contrast, I think that the immediacy premise is not obviously correct, and that inferentialism is a live option. So I need to do two things in this chapter: make a positive case for inferentialism and show why the usual objections to this view don’t work.

Here’s a simple statement of inferentialism: suppose you know that you have a certain attitude A and the question arises how you know that you have A. In the most straightforward case you know that you have A in so far as you have access to evidence that you have A and you infer from your evidence that you have A. As long as your evidence is good enough and your inference is sound you thereby come to know that you have A. On this account, the idea that self-knowledge is inferential is closely related to the idea that it is based on evidence. Moran writes that ‘the basic concept of first-person awareness we are trying to capture is that of awareness that is not based on evidence, behavioural or otherwise’ (2001: 11). The concept of first-person awareness – or self-knowledge- which inferentialism is trying to capture is that of awareness or knowledge that is based on evidence, behavioural or otherwise.

It will save time and help to prevent various kind of misunderstanding if I make a few things clear at the outset:

1. The attitudes I’m talking about are “standing” rather than “occurrent”. Standing attitudes remain in existence when you are asleep; they aren’t mental events like judging or deciding (cf. Carruthers 2011: xi). It’s controversial whether a belief can ever be occurrent but when I talk about belief I’m talking about beliefs understood as standing states. Ditto for desires, hopes, and so on.
2. I take it that if E is evidence for some proposition P then E makes it more likely or probable that P is true. Evidence can be, but needn’t be, conclusive. When inferentialism says that self-knowledge is based on, or inferred from, evidence it remains to be seen what kind or kinds or evidence are at issue. One kind of evidence is behavioural but there are other possibilities; you can discover your own standing attitudes on the basis of your judgements, inner speech, dreams, passing thoughts and feelings. These are all potentially varieties of evidence but they aren’t behavioural evidence.
3. Inference can be, but needn’t be, conscious. This came up in the last chapter, in connection with the idea that perception involves inference. To repeat what I said there: it isn’t a knockdown argument against representing perception as inferential to point out that we aren’t normally conscious of inferring when we perceive. There might be strong theoretical grounds for seeing perceptual knowledge or self-knowledge as inferential regardless of whether they involve any conscious inference. It is also worth adding that sometimes the claim that a particular type of knowledge is inferential is a claim about the various conscious or unconscious psychological processes that underpin it. At other times the claim is primarily epistemological rather than psychological; assuming that you can’t know that P unless you are justified in believing that P, saying that your knowledge that P is inferential might be a way of making a point about its justificational structure: your knowledge that P is inferential if your justification for believing that P comes in part from your having justification to believe other, supporting propositions (see Pryor 2005).
4. If E is your evidence that you have a particular attitude A, and you are know that you have A by inference from E, then you need access to E. What kind of access to E? That depends on the nature of your evidence. If your evidence is behavioural then your access to E might have to be perceptual, but this won’t be true if your evidence consists of your passing thoughts or inner speech. Inferentialism is not committed to regarding our access to these things as inferential. I will more to say about this later in this chapter.
5. The inferences that lead to self-knowledge are normally mediated by the subject’s implicit grasp of what is sometimes called a ‘theory of mind’. You only infer from E that you have A because you take E to be evidence (at least in your own case) for A. Taking E to be evidence for A is an implicit theoretical commitment of yours. Here’s an example: suppose that you judge that you ought rationally to believe that P and infer from this that you do in fact believe that P. You can only get from a claim about what you ought rationally to believe to a claim about what you actually believe on the assumption that you actually believe what you ought rationally to believe. The role of this assumption about how your mind works – Finkelstein calls it the Rationality Assumption- is to mediate the transition from ‘ought’ to ‘is’.

In view of what I’ve been arguing so far in this book inferentialism might seem a surprising position for me to be defending. I’ve criticized rationalism for operating with an over-optimistic and unrealistic conception of how humans think; in effect, the charge is that rationalists over-intellectualize human beings and regard as us far more intellectually virtuous and capable than we are. For example, humans often believe things on the basis of little or no actual evidence; and even when they have evidence for their beliefs they frequently don’t consult it. Doesn’t this make it unlikely that self-attributions of propositional attitudes are evidence-based? Insofar as inferentialism represents humans as coming to know their own attitudes by theory-mediated inference from evidence isn’t it just as guilty as rationalism of over-intellectualizing human self-knowledge?

 Well, it all depends on how notions like ‘theory’ and ‘inference’ are understood. Inferentialism is only in danger of over-intellectualizing self-knowledge on what is itself an over-intellectualized understanding of these notions. Take inference as an example: there is the case in which inferring something from something else is a careful and reflective piece of ‘slow’ rather than ‘fast’ thinking. The inferences that give us self-knowledge can be like this but needn’t be. They can be, and often are, automatic, effortless, and barely conscious. They are ‘fast’ rather than ‘slow’, and the assumptions that underpin them are usually implicit rather than explicit. Describing these assumptions as amounting to a ‘theory’ of mind is misleading if the implication is that in order to know your own attitudes you need to be able to do the philosophy of your own situation. Nothing beyond the intellectual reach of normal human adults is required. It’s also worth pointing out that our inferences, even our inferences about our own states of mind, aren’t always good, and these are cases in which we will end up making mistakes about our own attitudes. Such mistakes are possible and inferentialism explains how they are possible. It also explains how, when we do know our own attitudes, we know them: we know them by inference.

Another worry about inferentialism might be that it goes against my insistence that when it comes to explaining human self-knowledge there is no magic bullet, no one source that is capable of accounting for all our intentional self-knowledge. Isn’t inference a single source? There are two things to say about this: first, saying that inference is a key source of intentional self-knowledge for humans doesn’t mean that there aren’t other sources. Later I’ll consider the possibility that memory is a source of intentional self-knowledge, and that when you know what you think about something by remembering what you think about it your self-knowledge isn’t inferential. The second point is that ‘inference’ as I understand it is such a broad category and covers so many different things that inferentialism hardly amounts to a ‘magic bullet’ explanation of human self-knowledge. For example, if you come to know your own attitudes by using the Transparency Method your self-knowledge is inferential but your self-knowledge is also inferential if it is based on awareness of your behaviour or passing thoughts. Inferentialism is an inclusive doctrine that keeps many doors open. What it does is to help structure our thinking about self-knowledge, and put paid to various myths about self-knowledge that have grown up in recent years, the primary myth being that intentional self-knowledge is normally ‘immediate’.

What is the case for inferentialism? I’m going to outline three arguments in favour. Then I’ll discuss and respond to a range of more or less standard objections to inferentialism. The first argument for inferentialism, which I call the argument by elimination, goes back to Boghossian’s trilemma which came up in chapter 10. Here’s the trilemma again: knowledge of our own attitudes can only be:

1. Based on inference.
2. Based on inner observation.
3. Based on nothing.

I’ve already rejected 2 and 3 so that leaves 1 as the only remaining option. It had better be the case that we can know our own attitudes by inference because there is no viable alternative to inferentialism. We can argue about what self-knowledge is an inference from but rejecting 1 would leave us in the unhappy position of having to say that we do not know our own minds. If you are convinced that there are decisive objections to 2 and 3, and that scepticism about self-knowledge isn’t a serious option, then by default 1 has to be right: inferentialism is the only game in town.

Like all arguments of this form, the argument by elimination for inferentialism is only as strong as the case for thinking that:

1. All the alternatives have genuinely been eliminated.
2. There aren’t problems with remaining option that are just as serious as the problems with all the other options.

With regard to (i), I have already argued at length against 2 and won’t repeat these arguments here. The basic objection to 3 is this: self-knowledge can’t be based on nothing unless it is cognitively insubstantial, like knowing that I am here now. However, as Boghossian points out, there are plenty of indications that self-knowledge is not cognitively insubstantial: one can decide how much attention to pay to one’s thoughts, some adults are better than others at reporting on their inner states, and self-knowledge is fallible and incomplete. It’s natural to understand the difference between getting it right and failing to do so with regard to our own attitudes as ‘the difference between being in an epistemically favourable position with the relevant evidence – and not’ (Boghossian 1998: 167). On this account of the fallibility and incompleteness of self-knowledge the door is wide open to viewing self-knowledge not just as cognitively substantial but as inferential. That leaves (ii). If, as I’ll be arguing below, there aren’t decisive objections to inferentialism then the argument by elimination is in good shape. The position, then, is this: there aren’t decisive objections to inferentialism, there are decisive objections to the alternatives, so let’s all be inferentialists.

It’s one thing to think that inferentialism must be right but we also need to understand how knowledge of our own attitudes can be inferential. This brings me to my next argument for inferentialism, which I’ll call the argument by example. This argument builds on the idea that self-knowledge is cognitively substantial by giving examples of how human beings come to know their own attitudes by inference. Having an abstract guarantee that self-knowledge must be inferential is one thing but inferentialism will be much more concrete and secure if it is possible to come up with realistic examples of inferential self-knowledge. Such examples will serve to demystify inferentialism by showing it to be grounded in how we actually seek and acquire knowledge of our attitudes.

I’ve already mentioned one somewhat surprising example of inferentialism in action. If you come to know that you have an attitude A on the basis that you ought rationally to have A then your self-knowledge is by inference: with the help of the Rationality Assumption you infer what your attitude is from what it ought to be. I say that this is a surprising example because influential proponents of the Transparency Method (e.g. Moran) represent it as a way of acquiring ‘immediate’ self-knowledge. I suggested at the end of chapter 9 that TM is quite consistent with inferentialism, though it’s a further question to what extent humans actually rely on TM to know their own attitudes. Given all the problems with TM I’ve been discussing in this book it seems likely that TM is a relatively peripheral source of inferential intentional self-knowledge, at least when it comes to knowledge of such things as our hopes, desires, and fears.

For a more compelling example of humans acquiring self-knowledge by inference we need look no further than Krista Lawlor’s paper ‘Knowing What One Wants’. I want to spend some time on this paper because I’m very much is sympathy with Lawlor’s approach. Lawlor rightly observes that ‘too little attention has been paid to the experience of getting (and trying to get) self-knowledge, especially of one’s desires’ (2009: 56). She focuses on desires that are not easy to know about, such as the desire for another child. She gives the detailed example of Katherine who feels there is a fact of the matter about her desire for another child but struggles to know the answer to the question “Do I want another child?”. Notice how odd it would be for Katherine to answer this question by asking herself whether she ought rationally to want another child. She might ask herself this question if she conceives of herself as homo philosophicus but if she is a well-adjusted human being there are lots of other things she can and will do to answer her question:

Katherine starts noticing her experiences and thoughts. She catches herself imagining, remembering, and feeling a range of things. Putting away her son’s now-too-small clothes, she finds herself lingering over the memory of how a newborn feels in one’s arms. She notes an emotion that could be envy when an acquaintance reveals her pregnancy. Such experiences may be enough to prompt Katherine to make a self-attribution that sticks. Saying “I want another child”, she may feel a sense of ease or settledness (Lawlor 2009: 57).

If her self-attribution sticks, if she experiences a sense of ease when she says “I want another child”, then she has an answer to her question. She has evidence that she wants another child. If her self-attribution doesn’t stick then there are further things she can do in pursuit of self-knowledge. She can concentrate on her imaginings and try to replay imaginings about having another child. Her imaginings and fantasies are further data from which she can infer that she does, or does not, want another child.

If Katherine concludes on the basis of her feeling, imaginings and emotions that she wants another child it would be reasonable to describe her self-knowledge as inferential. What tells her what she wants is what Lawlor calls ‘inference from internal promptings’, which is in turn a form of ‘causal self-interpretation’ (2009: 48-9). Inference from internal promptings is, as Lawlor points out, a ‘routine means by which we know what we want’ (2009:60) and the resulting self-knowledge is a cognitive accomplishment. Clearly, there’s a lot at stake for Katherine when she asks whether she wants another child, but more prosaic examples can be dealt with in much the same way. The waiter asks you if you would like a pre-dinner cocktail and you ask for a glass of champagne. The minute you say the words “I’d like a glass of champagne” you realize that what you actually want is a vodka martini. It’s possible that you have changed your mind but it’s also possible that your reaction to placing your order tells you something about what you really wanted all along. This might be called the “suck it and see” route to self-knowledge: if you can’t work out what you want, commit to an option and it might become apparent what you want. In principle you could run the same line for cases of belief: you say you believe the present government will be re-elected but the minute you say the words you realize they don’t ring true. To say that something does or doesn’t feel right or ring true is to draw attention to the way that conscious experience or phenomenology can have an evidential role in relation to one’s attitudes. Beliefs and desires aren’t feelings but what you feel can sometimes tell you what you believe or desire.

A worry you might have about cases like Katherine is that they aren’t representative. The thought would be that although you can on occasion figure out your desires by inference from internal promptings you normally “just know” what you want without any inference. Lawlor feels the force of this in her discussion. She contrasts her view with the view that our desires are self-intimating and concedes that sometimes our desires are so ‘simple’ that ‘the idea that the desire is self-intimating is very plausible’ (2009: 56). To the extent that our desires are self-intimating our knowledge of them is neither inferential nor a cognitive accomplishment. Lawlor is relaxed about the existence of self-intimating desires because she is happy to concede that ‘there are many ways to discover one’s wants’ (2009: 60). However, the issue isn’t whether inference from internal promptings is a way to know one’s desires but whether it is the normal way. This is what Lawlor’s opponents will deny, and the existence of Katherine-type cases is neither here nor there as far as this issue is concerned.

In fact, the inferentialist’s position is far stronger that Lawlor’s discussion suggests. The objection to inferentialism is that ‘in normal cases’ one’s desires are self-intimating because ‘knowing one’s desire is not a matter of successfully finding out about or discovering desires that one has through cognitive effort’ (2009: 55). There are several things the inferentialist can say in response. Here are three:

1. Just because a desire is self-intimating it doesn’t follow that you (the subject of the desire) don’t know of its existence by inference. To say that a desire or other attitude A is self-intimating is to say that if you have the relevant concepts then you can’t have A without knowing that you have it. This doesn’t explain how you know you have A and leaves the epistemological issues wide open; on the face of it you could think that you can’t fail to know that you have A but that inference from internal promptings is the means by which you know you have A. Anyway, it’s not clear why we should view simple desires, such as the desire for something cool to drink, as self-intimating in the first place. Just because ‘many times, knowing what one wants is easy’ (2009: 56) it doesn’t follow that it’s in the nature of such desires that you can’t have them without knowing that you have them.
2. In “normal” cases in which we seemingly know our own desires without inference it is open to the inferentialist to maintain that we know our desires by unconscious inference. This goes along with the idea, which I’ve already mentioned, that ‘the cognitive processes of inference can be very swift, barely rising to consciousness (2009: 65). What is good about the idea that inference is always involved is that you then don’t have to think that different desires have different epistemologies. The difference between easily and not so easily known desires is a difference in degree, and it would be in keeping with this to suppose that knowledge of our own desires is always inferential. The difference between having to work out what you want and “just knowing” what you want isn’t the difference between desires your knowledge of which is inferential and desires your knowledge of which is non-inferential. It is the difference between obviously and unobviously inferential knowledge.
3. Many of the supposed examples of “immediate”, non-inferential self-knowledge actually have little to do with self-knowledge. Saying to the waiter that you would like something cool to drink is in the first instance a way of ordering something cool to drink and not a statement about your state of mind. The question whether your knowledge of what you want is inferential or non-inferential doesn’t arise because your statement isn’t expressive of self-knowledge. By the same token, many statements of the form “I believe that P” are, according to what Moran calls the Presentational View, nothing more than ‘the speaker’s way of presenting the embedded proposition P’ (2001: 70-1). If you say “I believe that P”, the way to contradict you is to deny P, not to deny that you believe that P. Where “I believe that P” is genuinely a statement about your state of mind, and the question arises whether you do in fact believe that P, inferentialism claims that you know that you believe that P by conscious or unconscious inference. Exceptions are cases where your statement doesn’t fall within the scope of the Presentational View and your knowledge of your state of mind is in no sense inferential. Later I will argue that there are such cases but that they put no pressure on the idea that inference is a key source of intentional self-knowledge for humans.

That’s just about as much as I want to say about the argument by example. If you are the kind of philosopher who takes it for granted that intentional self-knowledge is normally immediate or that, as Boghossian claims, ‘knowledge of one’s own mental states has to be non-inferential’ (1998: 153) then this argument should help you. It’s telling that Katherine is recognizably a human being addressing a humanly important question in a human way. Given the extent to which the philosophical literature has been dominated by examples of allegedly immediate intentional self-knowledge it’s good to be reminded that a lot of perfectly ordinary intentional self-knowledge is manifestly inferential, and that intentional self-knowledge that isn’t manifestly inferential could still be inferential. Whatever the merits of arguing that cases like Katherine aren’t “normal”, the very least they do is raise questions about the tendency to take the immediacy of intentional self-knowledge as a datum.

The final argument for inferentialism points in the same general direction and is really just an extension of the argument by example. I call it the argument by experiment. This draws attention to the experimental evidence for inferentialism. Needless to say, the evidence isn’t conclusive, but should certainly give pause to people who think that inferentialism is a crazy view. As an illustration of the craziness of inferentialism philosophers often quote the following passage from Ryle’s The Concept of Mind:

The sorts of things that I can find out about myself are the same as the sorts of things that I can find out about other people, and the methods of finding them out are much the same (1949: 149).

Since I find out what other people think by observing what they say and do Ryle seems to be saying that I find out what I think by observing what I say and do. Byrne says that this view ‘can appear obviously absurd’ (2012:1) and defends Ryle on the grounds that it isn’t what he really thought. In contrast, Davidson doesn’t accuse Ryle of defending an apparently or even actually absurd view but he does insist that ‘Ryle was wrong’ because ‘it is seldom the case that I need or appeal to evidence or observation in order to find out what I believe’ (1998: 87). The clear implication is that while it’s not ruled out that you might rely on behavioural evidence to find out what you believe such cases are far-fetched and unusual.

As I’ve emphasized inferentialism isn’t committed to the view that the evidence from which you infer your own attitudes is behavioural rather than, say, psychological. Still, it’s worth considering the role of behavioural evidence in intentional self-knowledge in the light of the work of social psychologists such as Daryl Bem. Bem is a proponent of what he calls ‘self-perception theory’ (SPT), which he describes as follows:

Self-perception theory was initially formulated, in part, to address empirically certain questions in the “philosophy of mind”…. When an individual asserts, “I am hungry”, how does he know? Is it an observation? An inference? Direct knowledge? Can he be in error or is that impossible by definition? How does the evidential basis for such a first-person statement (or self-attribution) differ from the evidential basis for the third-person attribution, “He is hungry”?

This passage is in one respect misleading because it suggests that SPT is concerned with self-attributions of sensation, whereas in fact its focus is the self-attribution of attitudes, including beliefs. Bem mentions Ryle’s view as a philosophical precursor of SPT, and it’s clear why he does so given the two ‘postulates’ of SPT. The first states that: ‘individuals come to “know” their own attitudes, emotions, and other internal states partially by inferring them from observations of their own overt behaviour and/or the circumstances in which this behaviour occurs’ (1972: 5). The second states: ‘to the extent that internal cues are weak, ambiguous, or uninterpretable, the individual is functionally in the same position as an outside observer, an observer who must necessarily rely upon those same external cues to infer the individual’s inner state’ (1972: 5). These postulates are distinctly Rylean but Bem, unlike Ryle, appeals to experimental evidence rather than conceptual analysis to support his theory. Could it be, then, that Ryle’s ‘behavioural’ inferentialism can be defended on the basis of such evidence?

The experimental evidence to which Bem refers includes the experiment described by Festinger and Carlsmith in their famous 1959 paper on ‘Cognitive Consequences of Forced Compliance’. In the classic Festinger-Carlsmith experiment, 60 students were randomly assigned to one of three conditions: in the $1 condition, the subject was first required to perform long, repetitive laboratory tasks and then paid $1 to tell a waiting fellow student that the tasks were enjoyable and interesting. In the $20 condition subjects were paid $20 to do the same thing. Control subjects simply engaged in the repetitive tasks. The results showed that the subjects paid $1 evaluated the tasks as significantly more enjoyable than subjects who had been paid $20. The latter subjects didn’t express attitudes significantly different from those expressed by the control subjects. What is going on here? Bem writes:

[S]elf-perception theory considers the subject in such an experiment as simply an observer of his own behaviour. Just as an outside observer might ask himself, “What must this man’s attitude be if he is willing to behave in this fashion in this situation?” so too, the subject implicitly asks himself, “What must my attitude be if I am willing to behave in this fashion in this situation?” Thus the subject who receives $1 discards the monetary inducement as the major motivating factor for his behaviour and infers that it must reflect his actual attitude; he infers that he must have actually enjoyed the tasks. The subject who receives $20 notes that his behaviour is adequately accounted for by the monetary inducement, and hence he cannot extract from the behaviour any information relevant to his actual opinions; he is in the same situation as a control subject insofar as information about his attitude is concerned (1972: 16-17).

There are many other experiments that point in the same general direction: in every case you have a subject who, in keeping with the postulates of SPT, knows his own opinion or attitude by inference from his own behaviour. The subject doesn’t “just know”.

Clearly, none of this amounts to a proof of inferentialism, not least because SPT is itself a controversial view among social psychologists. However, Bem’s discussion does call into question some of the bolder pronouncements of philosophers about the Rylean view. It is neither absurd nor far-fetched to suppose that subjects in forced compliance scenarios come to know their own attitudes by inference from their behaviour. Nor can it be assumed without further argument that it is very rare or unusual for us to self-attribute attitudes on the basis of behavioural evidence. According to Bem’s second postulate we rely on such evidence when internal cues are weak, ambiguous, or uninterpretable. If we take such internal cues to include internal promptings in Lawlor’s sense then it looks like an empirical question whether it is or is not unusual for such cues to be deficient in one way or another. However, whether we rely on internal promptings or behavioural evidence to determine our own attitudes, the resulting self-knowledge remains inferential and based on evidence. Bem’s work is helpful because it provides empirical support for the view that, while the evidence for self-attributions needn’t be behavioural, it certainly can be.

To sum up, I have described inferentialism as the view that inference from evidence is a key source of intentional self-knowledge for humans. One kind of evidence is behavioural, and the argument by experiment supplies experimental evidence that we sometimes come to know our own attitudes on the basis of our own behaviour. This is a serious problem for anti-inferentialists who say that behavioural evidence is never required or relevant when it comes to self-attributions of attitudes. It’s less of a problem for anti-inferentialists who only claim that we seldom know our own attitudes on the basis of behavioural or other evidence but it’s not clear why evidence-based self-knowledge should be deemed to be exceptional or unusual. I’ll come back to this issue shortly. Meanwhile, the position at which we have arrived in this: given the distinction between ‘internal’ and ‘behavioural’ evidence, it’s always an option for cautious inferentialists who don’t want to be tarred with the same brush as Ryle to retreat to the view that ‘internal’ evidence – judgements, inner speech, dreams, passing thoughts and feelings- is the only evidence on the basis of which we self-attribute attitudes. The argument by experiment suggests that inferentialists needn’t be so lily-livered. They should obviously refrain from saying that we can only know our own attitudes on the basis of behavioural cues but they can plausibly insist that the evidence we rely on in self-attributing attitudes includes behavioural evidence. There is experimental evidence for this, just as there is more informal but still compelling evidence of our reliance on internal cues for intentional self-knowledge. Combining the argument by elimination, by example and by experiment it’s hard to avoid the conclusion that inferentialism, in the moderate and inclusive form I’ve set out, is not only not a non-starter but a coherent, intuitively plausible, and well-supported alternative to the view that intentional self-knowledge is, if not impossible, then based on inner observation or based on nothing.

Why, then, has inferentialism had so few takers until recently? Because although it’s true that some critics have been far too quick to dismiss it, there are serious challenges and objections to inferentialism that still persuade many philosophers that it is not the way to go even if, as I’ve tried to show, it’s a position that deserves to at least be taken seriously. So it’s now time to deal with these objections and show that they don’t refute inferentialism. Here’s a quick summary of the objections I am going to take on:

1. Inferentialism can’t account for the epistemic asymmetry (‘the Asymmetry’) between knowledge of oneself and knowledge of others.
2. On an ‘internalist’ conception of epistemic justification inferentialism generates a vicious regress.
3. There are obvious counterexamples to inferentialism, cases which make it clear that inference is neither required nor relevant for self-knowledge.
4. Although inferentialism accounts for a certain form of self-knowledge, it doesn’t account for ordinary self-knowledge; inferential self-knowledge is alienated self-knowledge.

These are all potentially serious objections and some of them have some merit. Still, if we are careful about what inferentialism is and is not committed to then there is no need to lose any sleep over any of them.

Here’s a nice, clear statement of the Asymmetry from the preface to Moran’s book Authority and Estrangement:

[F]or a range of central cases, whatever knowledge of oneself may be, it is a very different thing from knowledge of others, categorically different in kind and manner... It is not necessary to say that the mind of another person is “essentially” hidden from me, in order to acknowledge that this person knows, and comes to know, his own thoughts and experiences in ways that are categorically different from how I may come to know them (2001: xxxi).

If I know the mind of another person by inference, and how I know my own mind is different in ‘kind and manner’ from how I know the mind of another person, the obvious conclusion is that in unexceptional cases I don’t know my own mind by inference.

The idea that inferentialism is no good because it is at odds with the Asymmetry is the most serious and certainly the most influential objection to this view. The options available to inferentialism are accommodation, denial, or some mixture of the two. The best bet is some mixture of the two, but with a greater emphasis on denying the existence of the Asymmetry than on attempting to accommodate it. Inferentialism can accept that there is an asymmetry between knowledge of oneself and knowledge of others but not the difference that Moran is talking about. In Alex Byrne’s terminology, our access to our own attitudes is ‘peculiar’, but not as peculiar as is commonly supposed.

To see how far inferentialism can accommodate the Asymmetry let’s go back to the case of Katherine. She wants to know whether she wants another child, and we can suppose that her best friend Melissa also has an interest in the question “Does Katherine want another child?”. Katherine answers this question by inference from internal promptings but Melissa is in no position to work out what Katherine wants on the same basis. Melissa can only find out what Katherine wants on the basis of what Katherine says and does rather than on the basis of what she feels and imagines. In this sense their methods of finding out are not the same even by the inferentialist’s lights; if you are an inferentialist you don’t have to think that, with respect to the question whether Katherine wants another child, Katherine and Melissa are in exactly the same boat epistemologically speaking, since they have access to different types of evidence.

On this account, the ‘asymmetry’ between knowledge of oneself and knowledge of others boils down to a difference in the kinds of evidence that are available in the two cases.

Although this is a significant difference, it’s not a difference in ‘kind and manner’. Even if Katherine and Melissa have access to different kinds of evidence, their ‘ways of knowing’ are the same to the extent that they both know what Katherine wants by drawing inferences from the evidence available to them. The difference in evidence (if there is one) isn’t the kind of difference that justifies talk of an epistemic Asymmetry. As for whether Katherine’s access to her own desires is ‘peculiar’, that depends on whether she knows about her desires in a way that is available to no one else. Suppose she discovers her desire for another child on the basis of her passing thoughts, feelings, inner speech, and dreams. In one sense her way of knowing is peculiar, since no one else has direct access to her passing thoughts, feelings, inner speech, and dreams. But in another sense there is nothing peculiar about Katherine’s way of knowing since Melissa can also infer what Katherine wants, and Katherine can make her ‘internal’ evidence available to Melissa simply by telling her about it.

The clear implication of this discussion is that inferentialism fails to accommodate the Asymmetry; even if Katherine and Melissa aren’t in exactly the same boat epistemologically speaking, their boats aren’t sufficiently different to give Moran the kind of Asymmetry he is looking for. This is therefore the point at which inferentialism needs to shift from half-baked accommodation to full-blown denial: instead of pretending to accommodate the Asymmetry it should question its existence. The issue isn’t whether there is an asymmetry between self-knowledge and knowledge of others – even inferentialism can accept this- but the kind of Asymmetry that rules out inferentialism. It’s no use saying that the Asymmetry is a primitive datum which doesn’t need to be argued for because it’s so obvious. As far as inferentialism is concerned it’s not obvious that there is an Asymmetry, and there is nothing more boring than two philosophers reiterating that their opposing views are ‘obvious’. Nor is it much of an improvement for proponents of the Asymmetry to go on about how a ‘categorical’ difference between self-knowledge and knowledge of others is built into the naïve conception of self-knowledge. Even if this is so, so what? Couldn’t the naïve conception be mistaken, especially in view of the trouble we get into when we deny that self-knowledge is inferential? Anyway, it’s doubtful whether we are naively committed to the existence of the Asymmetry, as distinct from the existence of an asymmetry.

The alternative to claiming that the Asymmetry is obvious is to argue for its existence, and the most promising argument is what might be called the argument from authorship. This says that we are the authors of our own attitudes, and that that is why our access to them must be non-inferential (as well as non-observational). Now we have two questions: in what sense are we the authors of our own attitudes, and why does being the ‘author’ of an attitude imply non-inferential knowledge of its existence? On the first question, the idea is that your beliefs, desires and other attitudes don’t come over you in the way that a headache might come over you. When all goes well your attitudes are the product of rational deliberation and a reflection of your judgements or decisions. Stuart Hampshire gives the example of a man who does not know whether he wants to go to Italy and who has to stop and think whether he does. If, from his initial state of uncertainty, ‘he moves to a conclusion which amounts to his now knowing what he wants, or to his now knowing what his attitude is, his process of thought is properly characterized as deliberation’ (1979: 289). It is his business whether he wants to go to Italy and it is for him to decide. When he decides after thinking about it that he wants to go, he is the author of his own desire in the sense that he has formed the desire ‘as the conclusion of a process of thought’ rather than come across it as a ‘fact of his consciousness’. In this respect, he is unlike Katherine, who merely ‘finds herself’ wanting or not wanting another child rather than taking control of her desire.

When you form a desire as the conclusion of a process of thought what you are doing is making up your mind, and in making up your mind you know your mind (cf. Moran 2011: 95). Since you are the author of your own attitude you don’t need to rely on evidence to know it. Your knowledge is epistemically and practically immediate: you are able to know what you think without relying on any evidence, observation or inference because your attitude is your responsibility. In contrast, I have to rely on evidence in order to know what you think because my relationship to your attitude isn’t authorial. This is the sense in which there is a difference in kind and manner between self-knowledge and knowledge of others. In brief: we have non-inferential access to our own minds because we make them up. We don’t have non-inferential access to other minds because we don’t control them in the same way. This is the asymmetry (and, indeed, the Asymmetry) that inferentialism does not and cannot account for.

The argument from authorship trades on what I referred to in chapter 9 as Activism. Activism says that we know own minds by actively shaping their contents, and that ‘making something puts one in a special cognitive relation to what one has made’ (Gaukroger 1986: 29). The argument from authorship adds that this special cognitive relation is such that, where you are the author a particular attitude A, you can know without inference or observation that you have A. Obviously this does nothing to show that we have non-inferential access to those of our attitudes which we haven’t authored. With respect to such attitudes, the argument from authorship leaves it open that there is no difference in kind and manner between first-person and third-person access. Even more problematically, it isn’t even clear that making something means that you know of its existence and nature without inference or observation. Consider this example: you work in souvenir a factory where your job is to make miniature replicas of a famous building, say Buckingham Palace. Your replicas are supposed to have a specific number of windows – say 27- and you set about producing replicas with the requisite number of windows. Do you know that the replicas you have made, or are making, have 27 windows? Clearly, the fact that you have made or ‘shaped’ the replicas that doesn’t make you infallible about the actual number of windows they have – you might have miscounted. You might still know without looking that they have 27 windows each, but only insofar as you know how many windows they are supposed to have, and also have general grounds for thinking that by and large when you try to make something with 27 windows it turns out to have 27 windows. This enables you to infer that your replicas have 27 windows but your knowledge that they have 27 windows is broadly inferential; it’s certainly not ‘based on nothing’.

No doubt there are all sorts of disanalogies between the sense in which we ‘shape’ our own attitudes and the sense in which we make things in factories but the question remains whether and how it is possible to get from ‘we shape our own attitudes’ to ‘we know our own attitudes non-inferentially’. It’s certainly not obvious that this is a legitimate move, and the epistemology of maker’s knowledge is in any case pretty obscure. Given the choice between saying that ‘maker’s knowledge’ is based on inference, observation or nothing, the best bet is the first option; at any rate, nothing has so far been said to rule this option out. Activists will no doubt want to say that the knowledge they have in mind is sui generis and doesn’t fall into any of Boghossian’s categories but it’s hard to know what to make of this without a positive account of the supposedly unique epistemology of maker’s knowledge. It’s worth adding also that we are now a very long way from the idea of the Asymmetry as a premise or datum from which philosophical reflection on self-knowledge is to start. A better approach is to avoid making contentious assumptions at the outset about the relationship between self-knowledge and knowledge of others, and focus instead on how we actually come to know our own minds and other minds. When we do that it becomes apparent that there is no knockdown argument against inferentialism from the Asymmetry. If anything, it’s the other way round: since a lot of ordinary self-knowledge is clearly inferential, and so is a lot of our knowledge of others’ minds, it would seem to follow that the Asymmetry is yet another philosophical myth.

What about the objection that inferentialism about self-knowledge generates a vicious regress, at least on an internalist view of justification? Boghossian writes: ‘on the assumption that all self-knowledge is inferential, it could have been acquired only by inference from yet other known beliefs. And now we are off on a vicious regress’ (1998: 155). Suppose you believe you are wearing socks and believe that you believe you are wearing socks. On an internalist view of justification the latter belief is epistemically justified only if you grasp the fact that it bears some appropriate relation to your other beliefs. However, these other beliefs must also be ones you are justified in believing you have, and this is what threatens a regress. The only way to avoid a vicious regress is to suppose that it’s possible to know the content of some mental states non-inferentially. Specifically, it follows from the regress argument that ‘not all knowledge of one’s beliefs can be inferential’ (1998: 156).

It’s easy to see where the regress argument goes wrong. The assumption is that if self-knowledge is inferential it must have been acquired from inference from other known beliefs, but there is no reason to accept that assumption even if you are an internalist about epistemic justification. For example, Katherine infers her desire for another child not from other beliefs but from internal promptings that aren’t standing attitudes. The passing thoughts, emotions and imaginings from which her infers her desire aren’t beliefs, and there is no reason why she couldn’t infer (some of) her beliefs on the same basis. There’s no question that her intentional self-knowledge is inferential but it isn’t inferred from other known beliefs. This isn’t to say that you can’t infer one belief or standing attitude from another, only that this isn’t how it has to go. Notice also that there is no conflict with internalism in either of its two standard forms. One kind of internalism (‘Accessibilism’) says that the epistemic justification of your beliefs is determined by things to which you have some sort of special access. Another kind of internalism (‘Mentalism’) says that your beliefs are only justified by things that are internal to your mental life. None of this is a problem for Katherine, on the assumption that she has ‘special access’ to her own inner speech, passing thoughts and imaginings. Such ‘internal promptings’ are about as internal to her mental life as anything could be.

It might seem that this only pushes the problem a stage further back. Inferentialism’s response to the regress is to say that you can infer your standing attitudes from your internal promptings, but what’s the story about access to internal promptings? Doesn’t inferentialism have to think that this is also inferential, in which case there is still a regress? Well, no. As I have emphasized, inferentialism is specifically a view about knowledge of our own standing attitudes. Just because knowledge of standing attitudes is inferential that doesn’t mean that all other self-knowledge is also inferential. So inferentialism doesn’t have to claim that we only have inferential knowledge of our internal promptings. The alternative is to say that we have non-inferential access to our own inner speech, fantasies, judgements, etc. and that, given the appropriate theory of mind, we are then able to infer our own standing attitudes on this basis. Clearly, even if it is open to inferentialism to adopt such a hybrid approach to self-knowledge it’s a further question whether it should adopt it. That depends, in part, on whether the hybrid approach can account for our supposedly non-inferential access to the internal promptings from which we infer our standing attitudes. I will have more to say about this in chapter 12.

The next objection to inferentialism says that there are obvious counterexamples that make it clear that inference is neither required nor relevant for self-knowledge. Some of the supposed counterexamples clearly don’t work. Here’s one from Boghossian:

You think: even lousy composers sometimes write great arias. And you know, immediately on thinking it, that this is what you thought. What explanation can the Rylean offer of this? The difficulty is not merely that, contrary to appearance and the canons of epistemic practice, he has to construe this knowledge as inferential. The difficulty is that he has to construe it as involving inference from premises about behaviour that you could not possibly possess (1998: 152).

Boghossian’s target here is that the view that knowledge of our occurrent thoughts is inferred from premises about behaviour. This is not inferentialism about self-knowledge as I interpret it, and is also not Ryle’s view. You don’t have to think that knowledge of our own occurrent thoughts is inferential in order to think that knowledge of our standing attitudes is inferential. When it comes to self-knowledge of standing attitudes, it is not ‘contrary to the canons of epistemic practice’ to construe this knowledge as inferential, whether the inferences in question are from behaviour or from other evidence.

Another potential counterexample to inferentialism turns on the role of memory in self-knowledge. Consider Katherine again. She wonders whether she wants another child and infers from various internal promptings that she does. Days later someone asks her whether she wants another child and she already knows the answer to this question. The answer, let’s suppose, is ‘yes’. She doesn’t have to rethink the question, since she already knows she wants another child. In order to answer the question all she has to do is to retrieve from memory what she already knows. This is significant because remembering that she wants another child does not require her to infer that she wants another child; the source of her self-knowledge in this case is not inference but memory, and she knows what she wants because of the role of memory in the preservation of her self-knowledge.

If in turns out that memory is a source of non-inferential intentional self-knowledge this wouldn’t be a disaster for inferentialism. Inferentialism says that inference is a basic source of intentional self-knowledge for humans. This could be true even if some of our self-knowledge, including intentional self-knowledge, is non-inferential. It’s worth pointing out, however, that it’s not obvious that Katherine’s memory-based knowledge of her desire for another child is non-inferential. Suppose someone were to object that Katherine only knows that she wanted another child when she last thought about it, and that it doesn’t follow from this that she now wants another child or that she now knows that she wants another child. The obvious thing to say in response to such sceptical doubts is that while it’s quite true that these things don’t follow – she could have changed her mind- it’s reasonable for Katherine to take it, in the absence of any evidence to the contrary, that her desires haven’t changed. In effect, she infers that what she wanted when the question last arose is what she still wants, and this is how she knows that she (still) wants another child. Her self-knowledge is inferential, even though based on memory.

The lesson is this: if you are sympathetic to inferentialism about self-knowledge, and are presented with a supposedly obvious counterexample to this view, you have quite a range of options at your disposal. One is to show that the counterexample isn’t relevant because it is attacking something to which inferentialism isn’t actually committed; this is the way to take care of counterexamples such as Boghossian’s. If that doesn’t work, then another option to show that the example is one in which the subject’s intentional self-knowledge is inferential. It’s always possible that what seems to be an example of non-inferential self-knowledge is in reality an example of unobviously inferential knowledge, unobvious because the inferential element might be unconscious or only implicit. In such cases, the justification for regarding the subject’s self-knowledge as inferential might itself be inferential, a case of inference to the best explanation. Finally, there remains the option of conceding that the counterexample is genuine but arguing that inferentialism can make space for it because saying that inference is a basic or key source of intentional self-knowledge for humans doesn’t require you to say that there is no non-inferential intentional self-knowledge. Inferentialism is only in trouble if there are counterexamples which can’t be dealt with in any of these three ways, and there is no reason to think that such counterexamples exist.

The final objection to inferentialism says that inferential self-knowledge is alienated rather than ordinary self-knowledge, and that inferentialism doesn’t account for ordinary, ‘unalienated’ self-knowledge. This is Moran’s objection to inferentialism. He argues that a person lacks ordinary self-knowledge ‘if he can only learn of his belief through assessment of the evidence about himself’ (2001: 67). Even if the evidence from which you infer your attitudes is highly reliable, and includes ‘internal’ as well as behavioural evidence, you will still only end up with ‘theoretical’ or ‘attributional’ self-knowledge. Whereas self-knowledge in the ordinary sense is a ‘specifically first-person phenomenon’ (2001: 2), attributional self-knowledge is ‘the expression of an essentially third-personal stance towards oneself’ (2001: 106). Ordinary self-knowledge is knowledge of attitudes you can identify with and rationally endorse, but attributional self-knowledge can be of attitudes you don’t identify with, and whose reasons are opaque to you. In such cases, you only have alienated self-knowledge, and the reason it is alienated is that it is third-personal.

An example might help. Suppose you have evidence that our old friend Oliver has a particular attitude A. You can infer that A is Oliver’s attitude even if you find A repellent or incomprehensible. Suppose A is the belief that the 9/11 attacks were the work of government agents rather than al Qaeda terrorists. You find it impossible to identify with, or endorse, this belief; indeed, you find Oliver’s belief absurd, but you still recognize that it is what Oliver believes. The belief you are justified in attributing to him is not the one that is supported by overwhelming evidence of al Qaeda’s responsibility for the 9/11 attacks but rather the one supported by psychological and behavioural evidence relating to Oliver. But if inferentialism is correct then you could find yourself with evidence that you have a belief or other attitude that is as difficult for you to identify with as Oliver’s belief. You will then find yourself in the absurd position of having to admit that you have an attitude which you can’t make any sense of. Your relation to your own attitude in this case will be no different from your relation to Oliver’s attitude; you will be as deeply alienated with respect to your own attitude as you are with respect to Oliver’s attitude.

The reply to this objection to inferentialism is straightforward: just because inferential or attributional self-knowledge can be of attitudes you don’t identify with and can’t endorse, it doesn’t follow that this kind of self-knowledge has to be alienated in these ways. The mere fact that you self-ascribe an attitude on inferential grounds doesn’t make the attitude alienated or impervious to reason. You can infer on behavioural or psychological grounds that you have a particular belief or desire and yet have no difficulty endorsing or identifying with that belief or desire. It’s also worth pointing out that the rationalist alternative to inferentialism can itself result in a form of alienation. Depending on the kind of person you are, you might find it hard to identify with attitudes you only take yourself to have because you assume that they are the attitudes you ought rationally to have.

Katherine is a good illustration of all this. Suppose she infers from the internal and behavioural evidence available to her that she wants another child. Just because this is how she knows that she wants another child, that needn’t prevent her from recognizing the desire as fully her own. Indeed, the fact that the desire is manifested in her emotions, dreams, and fantasies might actually enhance her sense of the desire as fully her own, an expression of who she is at this point in her life. If this is the case, then the fact that she comes to know her desire on the basis of these same emotions, dreams and fantasies can hardly make her self-knowledge alienated. It’s true, of course, that inferring from her internal promptings that she wants another child might not be the end of the matter for Katherine. There is a further question she can ask, namely, is there good reason to have, or to want to have, another child? Acquiring her self-knowledge by inference doesn’t make it impossible for her to ask this further question, or for her to change her mind in response to the arguments for and against having another child. If she does change her mind at this point then she will need evidence to know that this is what has happened. On the other hand, suppose her desire for another child survives her recognition that there are good practical reasons for her not to have one. That still doesn’t make her desire alienated; it might instead be viewed as an indication of the strength of her identification with the desire. As Harry Frankfurt points out, ‘some of the desires with which a person is most deeply and significantly identified, and from which it is nearly impossible for him to become alienated, are not based on any thought about what is good to be pursued’ (2002: 223). There is no tension or conflict in such cases between being committed to the attitude and knowing inferentially, on the basis of evidence, that you have it.

The point about rationalism and alienation is this: rationalists are keen on the idea that unalienated attitudes are ones that are answerable, and knowable by reference to, to rational considerations. For example, Moran suggests that a person’s unalienated desires are those that are ‘guided by the direction of his thought about what is desirable’, so that ‘he takes the general question of what he wants… to be the expression of his sense of what he has best reason to pursue in this context’ (2001: 117-8). Now consider the following variation on the case of Katherine: suppose that she has no children and asks herself whether she now wants a child. She has never thought of herself as interested in having children and has always been comfortable with the idea of not having any; she has never been envious of her female friends with children and doesn’t see herself as cut out for motherhood. However, she worries that she will one day regret not having children, and convinces herself on a variety of grounds that she ought to have a child. Suppose that she now attributes to herself the desire for a child on the basis that it is what she ought rationally to want. One possibility is, of course, that it’s just false that she wants a child, however convinced she is that she ought to want one. It might turn out, however, that at some level she does now want a child; maybe this is the effect of her rehearsing to herself the reasons in favour. Still, there remains a sense in which this new desire, even though a genuine expression of her sense of what she has best reason to pursue in this context, isn’t a genuine expression of her. It doesn’t fit her self-conception and might take a lot of getting used to. It feels to her like an alien or alienated desire precisely because it is grounded in reason rather than in her sense of who she is.

If this is right then rationalism is on shaky grounds when it accuses inferentialism of only explaining alienated self-knowledge: inferential self-knowledge needn’t be alienated, and rationalism about self-knowledge isn’t immune to worries about alienation. This should come as no surprise since, as I’ve argued in previous chapters, rationalism in any case only delivers inferential self-knowledge. For the purposes of this chapter than the second: the key point is that inferential self-knowledge needn’t be alienated. This takes care of the ‘objection from alienation’ to inferentialism, and shows that this objection is no more effective in undermining inferentialism about self-knowledge than all the other objections I’ve discussed. Inferentialism is alive and kicking, and remains the only game in town. Much better to admit that intentional self-knowledge is inferential than to claim that it is observational or based on nothing. But what about substantial self-knowledge? Is this inferential too? Some of it clearly is, since some intentional self-knowledge is also substantial. Katherine’s self-knowledge is a case in point. However, there are many other types of substantial self-knowledge, and the question for the next chapter is: what are the sources of these other kinds of substantial self-knowledge?