14

SELF-IGNORANCE

Self-ignorance comes in many different varieties, some more interesting than others. Suppose you have been eating too many desserts and your weight has been creeping upwards. You have put on weight but you neither know nor believe you have. This is a form of self-ignorance – physical self-ignorance- but not one that excites much philosophical interest. The natural view is that having a physical property like being overweight or weighing sixty kilos is one thing, knowing or believing that you have such properties is another. It isn’t in the nature of such properties that if you have them you know or believe you have them, and there is no reason to suppose that you can’t be ignorant, or plain wrong, about your weight. Neither error nor ignorance is ruled out, and neither calls for any special explanation.

When it comes to our mental properties, philosophers influenced by Descartes tell a very different story. They agree that you can be in pain without knowing that you are in pain. Maybe you don’t have the concept of pain. You can believe you are wearing socks without knowing that you believe you are wearing socks. Maybe you don’t have the concept of belief. However, Cartesians think if you have the necessary concepts, and there isn’t anything wrong with you, then if you are in pain you must know you are in pain, and if you believe you are wearing socks then you must know that you believe you are wearing socks. On this view, core mental properties are necessarily or constitutively self-intimating: it belongs to the very nature of such properties that if you are rational and have the necessary conceptual resources you can’t have them without believing and knowing that you have them. If this is correct then certain forms of mental self-ignorance are ruled out; there are truths about you mental life you can’t fail to know unless there is something wrong with you, whatever that means. Indeed, when it comes to knowledge of your own standing attitudes, you might think that not only can you not fail to know what they are, you can’t fail to know why they are as they are.

Not all mental properties are like this. Suppose you think of fastidiousness as a mental property, at least to the extent that being fastidious is a matter of what you care about or what bothers you. No one would suppose that fastidiousness is constitutively self-intimating, and it doesn’t take any special effort or ingenuity to conceive of a fastidious person failing to know that, or why, he is fastidious. Nevertheless, if you are a Cartesian you might still believe you are uniquely well placed to know your own character traits, as well as your own emotions and values. Knowledge of these things might not be unavoidable in the way that knowledge of your own beliefs is unavoidable, but is still straightforwardly attainable.

What I have been describing as the Cartesian view is a form of what I’m going to call optimism about human self-knowledge. At the end of chapter 12, I mentioned three questions about self-ignorance:

1. How prevalent is it?
2. What are its sources?
3. To what extent can it be overcome?

Optimists about self-knowledge think that self-ignorance isn’t and can’t be prevalent, at least in relation to a designated range of mental properties: if you are rational and not conceptually impoverished then you can’t fail to know your own sensations, mental actions, and standing attitudes. You normally know why your attitudes are as they are, and know much, if not all, of what there is to know about your own character, values and emotions. Your abilities and what makes you happy are perhaps more elusive but optimists see no reason in principle why you couldn’t also acquire these forms of substantial self-knowledge.

Optimists take a dim view of the sources of self-ignorance. For example, Tyler Burge contrasts knowledge of our own thoughts with perceptual knowledge. He argues that a person can be perceptually wrong without there being anything wrong with him. In this domain brute errors - ones that do not result from any carelessness, malfunction, or irrationality on our part - are possible because the objects of perception are independent of our perceptual awareness of them. In contrast ‘all matters where people have special authority about themselves are errors which indicate something wrong with the thinker’ (Burge 1994: 74). Optimists take such matters to include standing attitudes such as beliefs, hopes, and desires. They maintain that not knowing what you want, hope or believe is a clear indication that there is something wrong with you, and that only irrationality or some kind of cognitive malfunction can explain your self-ignorance. Ignorance in relation to your own character traits is a different matter. This form of self-ignorance need not have its source in irrationality or cognitive malfunction but might still be an indication of lesser defects such as carelessness or laziness.

Insofar as irrationality or cognitive malfunctions are the sources of self-ignorance the only way to overcome self-ignorance is to correct or overcome these defects. But if you are irrational there may be little you can do about it, and it’s also conceivable that some cognitive malfunctions can’t easily be fixed. If that is how things turn out, then forms of self-ignorance that indicate something wrong with the thinker might be incurable. In this respect, optimists are pessimists. The more abnormal you regard self-ignorance as being, the more you interpret it as a sign that something is wrong with the thinker, the harder it becomes to still regard it as curable. Still, the overall picture remains optimistic in the following sense: optimists see self-knowledge in its various forms as the norm, and self-ignorance as a deviation from the norm, as an indication that something has gone wrong.

My aim in this chapter is to argue for a more pessimistic view of self-knowledge. On this view some self-ignorance, even with respect to one’s own standing attitudes, is inevitable and quite normal for humans. Not knowing what you want, believe or hope doesn’t indicate that you are irrational or malfunctioning, any more than not knowing your own character indicates that you are irrational or malfunctioning. All it indicates is that you are a fully paid up and normal member of the species homo sapiens. If Descartes is the arch optimist then Nietzsche is the arch pessimist, as indicated by this dramatic passage:

We are necessarily strangers to ourselves, we do not comprehend ourselves, we have to misunderstand ourselves, for the law “each is furthest from himself” applies to all eternity (Genealogy of Morals).

This is a bit of an exaggeration. A sensible pessimism should refrain from claiming that self-knowledge is impossible, or that we are necessarily strangers to ourselves. You can think that a degree of self-ignorance is normal for humans without also thinking that self-knowledge is impossible. However, this leaves it open whether the respects in which we are strangers to ourselves are, as Nietzsche would no doubt say, the important ones. The trick is to strike the right balance between human self-knowledge and human self-ignorance, and to explain our self-ignorance without taking it as proof that there is something wrong with us. You don’t have to be irrational or mentally defective to be self-ignorant and, as we will see in the next chapter, some self-ignorance might actually be good for us.

Here is my plan for this chapter: I’ll start by looking at the extent to which we can and do fail to know our attitudes, where this is a matter of failing to know what we want, believe, and so on. I will distinguish motivational and non-motivational explanations of self-ignorance and show how the inferentialism I’ve been defending in this book delivers a straightforward non-motivational explanation of self-ignorance. If you want to explain how self-ignorance is possible, you can hold, but don’t have to hold, that it is motivated. It’s better to think of self-ignorance as a natural by-product of the mechanisms of self-knowledge: given how we go about acquiring self-knowledge, it is entirely predictable that we don’t always know our own minds, even if there is nothing wrong with us. It remains to be seen how easily, on this view, self-ignorance can be overcome.

Next, I will look at knowledge of our own attitudes in the sense of knowing why they are as they are. This is what Nietzsche is really interested in, and I will suggest that optimism in this domain is even less defensible than optimism about knowledge of our own attitudes in the sense of knowing what they are. Again, inferentialism has no difficulty explaining our self-ignorance in respect of why we believe what we believe, want what we want, and so on. The key is to distinguish epistemic and non-epistemic explanations of our attitudes. Going back to our friend Oliver the conspiracy theorist, an epistemic explanation of his beliefs about 9/11 will talk about his reasons or evidence. A non-epistemic explanation will concentrate on character traits such as his gullibility. Nietzsche thinks that non-epistemic explanations of our attitudes are not only the deepest explanations but also the most inaccessible. We will see if he is right about this.

Finally, I’ll look at the extent to which we are self-ignorant in relation to such things as our character, values and emotions. The more substantial a variety of self-knowledge the easier it is to accept that one can lack it. Ignorance of one’s own character is straightforward, and the interesting question is not whether this form of self-ignorance is possible or but rather what explains it. Ignorance in relation to one’s own values and emotions is a little harder to understand but inferentialism has something to say about each case. The take home message of this chapter is that self-ignorance is inevitable for humans and can be tough to overcome. Instead of denying that self-ignorance is possible what philosophers should be doing to trying to explain it.

How can you fail to know what you believe? If you believe the present government will be re-elected mustn’t you know that this is what you believe? However, before getting too excited about the seeming inevitability of self-knowledge in this case, it would be worth asking how representative the example is. When it comes to other attitudes, it’s as easy as pie to conceive of our failing to know what there is to know: you can want something without knowing that you want it, hope for something without knowing that you hope for it, fear something without knowing that you fear it, and so on. Here is a nice example of ignorance of what one hopes:

I believe that I do not hope for a particular result to a match; I am conscious of nothing but indifference; then my disappointment at one outcome reveals my preference for another. When I had that hope I was in no position to know that I had it (Williamson 2000: 24).

Even when it comes to one’s own beliefs, it’s not that difficult to imagine someone having a belief they don’t realize they have: for example, perhaps it’s clear from what you say and do that you do in fact believe that the present government will be re-elected, but you have never explicitly thought about the government’s election prospects and do not have the belief that you believe the present government will be re-elected. Even if, in this case, you don’t actually know what you believe, it might be said you are at least in a position to know you believe the government will be re-elected: all you have to do is think about what you believe. But even this isn’t guaranteed to produce self-knowledge. Maybe the government is so odious that you are unable to admit to yourself that you believe it will be re-elected; you have a determinate belief about its election prospects you don’t know you have, and aren’t even in a position to know you have, as your path to knowing what you believe is blocked by a psychological obstacle. The obstacle in this case might be embarrassment or despair.

This suggests the following picture: suppose you have a particular attitude A, and the question is whether you know that you have A. Let’s agree that you can’t know you have A if you don’t believe you have A. There is a difference between not believing that you have A and believing that you don’t have A. In mild cases of self-ignorance, the sense in which you don’t know that you have A is simply that you lack the second-order belief that you have A. Let’s call this ‘mere’ self-ignorance. However, there is also the possibility that you have A but believe that you don’t have A: for example, you hope for a particular outcome to a match but mistakenly believe that you don’t hope for that outcome. Alternatively, you don’t have A but you mistakenly believe that you have A: you don’t really want another martini but believe you do. In contrast with ‘mere’ self-ignorance, the last two examples are ones in which you are mistaken; you don’t just lack a true second-order belief about whether you have A, you have a false second-order belief. Let’s call these cases of self-deception. If you mistakenly believe that you don’t have an attitude which you do in fact have then this is what Shoemaker calls ‘negative self-deception’ (2009: 35). Positive self-deception happens where you believe you have an attitude which in reality you don’t have.

In explaining mere self-ignorance and self-deception motivational approaches appeal to motivational factors. For example, if believing that you have a particular attitude A would be unpleasant or anxiety-inducing then you will be motivated not to believe that you have A. But how can you not believe you have A, or believe that you don’t have A, if all the evidence points to your having A? Motivational explanations of self-deception suggest that your desire not to subject yourself to psychic discomfort motivates you to forget, misconstrue or ignore the evidence. Looking the other way when confronted by evidence that you have A, or simply forgetting the evidence, are strategies your psyche pursues, usually unconsciously, in order to minimize its own discomfort. If successful, these strategies cause you to be self-ignorant or self-deceived and thereby maximize your psychic well-being.

One problem with this explanation of self-ignorance is that it is limited in scope. It’s just not plausible that all or even most cases of self-ignorance are ones in which the unknown attitude is unpleasant or anxiety-provoking. For example, you are wrong about whether you want another martini even though recognizing that you do wouldn’t cause you any psychic distress. More generally, it’s not plausible that every case in which you are mistaken about what you want or hope or believe can be explained on the basis that you are trying to protect yourself from pain or anxiety. Sometimes you are ignorant or mistaken about your attitudes without any ulterior psychological motive. In such cases a different account of self-ignorance is needed, either one that refers to different motivational factors or that doesn’t explain self-ignorance in motivational terms.

In addition to questions about the scope of motivational accounts of self-ignorance there are also questions about the mechanisms or processes by which what might be called ‘self-protective’ self-ignorance is achieved. I’ve talked vaguely about looking the other way when confronted by evidence that you have an anxiety-provoking attitude but perhaps the most influential account of motivated self-ignorance is associated with Freud. This says that self-ignorance is specifically the result of repression. Is there any evidence for this view? The issue here is not whether it is always right to explain self-ignorance by reference to repression but whether, in light of the empirical evidence, it is ever right. This issue has been taken up by Timothy D. Wilson and Elizabeth W. Dunn, who maintain that an empirical demonstration of repression would have to show that:

1. People are motivated to keep thought, feelings, or memories outside of awareness.
2. The attempt to keep material out of awareness is itself an unconscious process.
3. People succeed in removing the undesired material from consciousness.
4. The material, once removed from consciousness, still exists in memory and continues to influence people’s thoughts, feelings, or behaviour.
5. The material is recoverable if the repressive forces are removed.

Dunn and Wilson conclude that though ‘a patchwork of studies depicts a mental architecture that would allow repression to occur’, it’s also the case that ‘no single study has demonstrated all the necessary criteria to establish the existence of repression definitively’ (2004: 17.4). At best there is ‘suggestive evidence’, and sceptics about the motivational approach will not see this as an adequate basis for constructing a theory of self-ignorance.

There is obviously much more to be said about all this but you might wonder whether it’s even necessary to explain self-ignorance in motivational terms. Motivational approaches get into trouble over questions about their scope and empirical support but these questions can be avoided if it’s possible to provide a non-motivational explanation of self-ignorance. A non-motivational approach needn’t deny that self-ignorance is sometimes motivated, perhaps in the way that Freud describes, but insists on a core account of self-ignorance which makes no mention of motivational factors. Such factors might have a part to play but they are not the fundamental basis on which self-ignorance is to be explained. If an account along these lines can be made to work then we don’t need to worry so much about the quality of the evidence for repression, or the fact that we are sometimes ignorant of attitudes which wouldn’t bother us if we knew of their existence.

What would a non-motivational explanation of self-ignorance look like? Suppose that when all goes well self-knowledge comes about as a result of theory-mediated inferences from internal or other evidence. This is inferentialism about self-knowledge, and one of the attractions of this view is that it points to a straightforward, non-motivational account of self-ignorance. If you have an attitude A then you will fail to know that you have A in any of the following circumstances:

1. You haven’t performed the necessary inference from the evidence you have.
2. You lack the necessary evidence.
3. You have all the evidence you need but draw the wrong conclusion about whether you have A because:
4. You reason poorly.
5. You misinterpret the evidence.
6. You have a defective theory about the relationship between your evidence and your attitude.

The result of (a) and (b) will be ‘mere’ self-ignorance. The other scenarios may result in your having a false belief about your attitude – self-deception rather than mere self-ignorance. But there is no mention of repression or of your self-ignorance being motivated by a desire not to subject yourself to psychic distress. Your failure to perform the necessary inference or your misinterpretation of the evidence might be motivated but needn’t be.

To see how this all works in practice, let’s go back to the example of Katherine from chapter 11. In that example, Katherine wonders whether she wants another child and answers this question in the affirmative on the basis of ‘internal promptings’ – memories, emotions, imaginings, and so on. However, it’s easy to conceive of Katherine being aware of the same internal promptings but not reflecting on their significance and not self-attributing the desire for another child. Maybe she has too much else on her mind and never gets as far as inferring that she wants another child. She wants another child but doesn’t (yet) know or believe she wants another child. This is an example of mere self-ignorance. Or suppose that she does in fact want another child but has yet to encounter the internal promptings that would enable her to infer that she wants another child. This is still mere self-ignorance, but a reflection of Katherine’s lack of evidence rather than her failure to perform the necessary inference.

Suppose, on the other hand, that Katherine has the evidence that she wants another child but judges that she doesn’t want another child. How can she be negatively self-deceived about her desire for another child? In Lawlor’s example, Katherine is described as noting an emotion that could be envy when an acquaintance reveals her pregnancy but there is always the possibility of misinterpretation or of Katherine failing to grasp the significance of her own envy on hearing of her acquaintance’s pregnancy: she doesn’t see that her envy is evidence of her desire for another child. Her misinterpretation of the evidence and failure to grasp its true significance explain her mistaken belief that she doesn’t want another child.

Another example from chapter 11 can be used to make the same point. In this case, a subject is paid a dollar to perform a long and boring laboratory task, and infers that he must actually have enjoyed performing the task because he discounts monetary inducement as the major motivating factor. It is completely clear in this case that the subject infers his attitude on the basis of a theory about why he performed the task. But of course there is no guarantee that his theory is correct. In this case he didn’t enjoy the task but his defective theory leads him to infer that he did enjoy it. This is now a case a positive self-deception, that is, the self-attribution of an attitude one does not or did have. In Wilson’s terminology, the theoretical route to self-knowledge can lead to ‘self-revelation’ but it can also result in ‘self-fabrication’, where you mistakenly infer the existence of an attitude that was not or is not actually present. The misattribution might be motivated, but could also simply be the result of the so-called ‘fundamental attribution error’, whereby ‘people underestimate the effects of external factors on their behaviour’ and ‘misattribute their actions to an internal state’ (Dunn & Wilson).

The inferentialist thinks that, in principle, you can run the same kind of story to make sense of the misattribution of any attitude. You can misattribute a belief, a hope, a fear, or an emotion like jealousy because you jump to the wrong conclusion about your state of mind. This needn’t be a conscious process, any more than correctly inferring your attitude needs to be a conscious process. Both self-knowledge and self-deception can be, and normally are, the result of automatic transitions rather than deliberate reflection. However, as long as you think of self-knowledge as the product of theory-mediated inferences, you are effectively building into your account the possibility of self-ignorance and self-deception. This is a strength rather than a weakness of inferentialism since self-ignorance and self-deception clearly are possible, and inferentialism explains with minimum fuss and without appealing to motivational factors how they are possible: theories can be defective, evidence can be lacking or misleading, and inferences are not guaranteed to be correct.

Do cases of positive or negative self-deception indicate that there is, as Burge puts it, ‘something wrong with the thinker’? Not if the point of this is to suggest that self-deception must be the result of carelessness, malfunction or irrationality. Of these three possibilities the easiest to deal with is irrationality. On the narrow conception of irrationality which I’ve been relying on in this book, irrationality in the clearest sense occurs when a person’s attitudes fail to conform to her own judgements, when ‘a person continues to believe something (continues to regard it with conviction and take it as a premise in subsequent reasoning) even though she judges there to be good reason for rejecting it’ (Scanlon 1998: 25). But when the subject in the Festinger-Carlsmith experiment believes that he must have enjoyed the boring task, or when Katherine believes that she doesn’t want another child, there is no irrationality in this sense. They both believe what they believe they have good reason to believe, and there is no conflict between their attitudes and their judgements. Carelessness isn’t the issue either: the subject in the Festinger-Carlsmith experiment isn’t proceeding without due care and attention when he concludes, on reflection, that he must have enjoyed the tedious task; he’s just wrong. What about the idea that this kind of error indicates a malfunction? Again, that’s not obvious. A malfunctioning device is one that doesn’t work as it should but a person who self-attributes attitudes after due consideration of the evidence is in one sense operating just as he should. The fact that some of his self-attributions are misattributions doesn’t indicate a malfunction unless there mere making of a mistake indicates a malfunction. If that were so then no errors would come out as ‘brute’; they would all indicate something wrong with the thinker.

Burge thinks that a person can be perceptually wrong without there being anything wrong with him; perceptual errors can be ‘brute’. What I have just been arguing is, in effect, that errors about your own attitudes can also be brute errors. When you make mistakes about your own attitudes you aren’t misperceiving them but you may be misinterpreting them. Just because you occasionally misread what you believe, hope, or want, that doesn’t necessarily mean that there is something wrong with you. This is a reflection of the fact that such objects of self-knowledge are independent of our knowledge of them, just as the objects of perceptual awareness are independent of our awareness of them. Inferentialism sees self-knowledge as a process of self-discovery which allows for the possibility of blameless mistakes. Gross or frequent errors about you own attitudes are a different matter. They would indicate something wrong with you, but so of course would gross or frequent perceptual errors.

It should be obvious that on this account of self-knowledge our standing attitudes are not ‘constitutively’ or necessarily self-intimating. There would be no question of explaining how self-ignorance is possible if you can’t have a particular belief, desire, or other standing attitude without knowing that you have it. Inferentialism makes it clear that and how self-ignorance is possible, and thereby removes any basis for going along with the thesis that our standing attitudes are necessarily self-intimating. This thesis is utterly implausible quite apart from what inferentialism implies. Throughout this book I’ve operated with a dispositionalist account of belief and other attitudes: to believe that P is to be disposed to think that P, to act as if P is true, to use P as a premise in reasoning, and so on. Merely having the dispositions associated with believing that P is no guarantee that you know or believe that you have them, just as believing that you have the relevant dispositions is no guarantee that you have them. Neither ignorance nor error is ruled out, and self-ignorance is possible even if the dispositions you need in order to count as believing that P include the disposition to self-ascribe the belief that P. If you believe that P, and the question arises whether you believe that P, then other things being equal you will judge that you believe that P but it doesn’t follow that you believe that you believe that P prior to the question arising. Suppose you believe that the government will be re-elected. The thought that this is what you believe might never have crossed your mind, and if it did cross your mind you might find it hard to admit to yourself. Yet your other dispositions might leave no room for doubt that this is what you believe.

In addition to the question whether you can believe that P without knowing that you believe that P there is the question whether you can believe that P without knowing why you believe that P. In addition to the question whether you can want that P without knowing that you want that P, there is the question whether you can want that P without knowing why you want that P. Pessimists see no difficulty here. They think it’s obvious that self-ignorance in the ‘knowing why’ sense is possible, and perhaps even unavoidable. The interesting question here is not how self-ignorance is possible but how self-knowledge is possible, that is, how it’s ever possible for you to know why you believe the things you believe, want the things you want, and so on. Inferentialism has a simple answer to this question: you can sometimes infer why your attitudes are as they are. However, your inferences can lead you astray, and at other times you may find yourself stuck for an answer. It might be obvious to you that you want to have lunch now because you are hungry but it might be far from obvious to Katherine why she wants another child. The possibilities of self-deception and confabulation are endless and self-ignorance is always on the cards.

Opposed to this form of pessimism is a form of optimism which says that insofar as our attitudes are the product of reasoning we are in a position to know, by reflecting on our reasons, why they are as they are. Here is Matthew Boyle’s statement of this view:

[I]f I reason “P, so Q” this must normally put me in a position, not merely to know that I believe that Q, but to know something about why I believe Q, namely, because I believe that P and that P shows that Q…. successful deliberation normally gives us knowledge of what we believe and why we believe it (2011: 8).

In principle you can run the same line for any attitude that you reason yourself into. If you form the desire to go to Italy for the summer after considering the pros and cons then you are in a position to know that you want to go to Italy for the summer and why you want to do that. Once again, the source of your self-knowledge is deliberation: deliberation can give you knowledge of what you want, and why you want it, to the extent that your desire is the result of deliberation.

You might wonder how much optimists and pessimists really disagree. Here is a way of splitting the difference between the two positions: some of our attitudes arise as a result of deliberation but some do not. If you reason “P, so Q” then you are in a position to know why you believe Q but this strategy won’t work if you believe that Q without having deliberated. So maybe we should say that successful deliberation gives us knowledge of what we believe and why we believe it as long as we are talking about beliefs formed by deliberation. If we haven’t deliberated then deliberation can’t be what gives us knowledge of what we believe and why we believe it, and these are the cases in which self-ignorance is genuinely on the cards.

Neither optimists nor pessimists are likely to be impressed by this attempt to split the difference between them. Optimists will argue that even if you haven’t actually reasoned your way to Q you can still be asked why you believe that Q. This is a request for your reasons, and in giving your reasons you will be revealing why you have that belief. It doesn’t matter whether you have actually deliberated your way to Q. What matters is that you have reasons for your belief, and that you can give them if challenged. Pessimists will insist that once you grasp what they mean by ‘knowing why you believe that Q’, it will be apparent that even if you have reasoned your way from P to Q you might still not be in a position to know why, in the relevant sense, you believe that Q. When it comes to knowing why your attitudes are as they are, there are different levels of explanation, some more superficial than others. In some cases, only reflection or reasoning that is external from your reasoning from P to Q can tell you why, in the deepest sense, you believe that Q. Even then, there is no guarantee of self-knowledge.

This is all too abstract and the best way of making it more concrete is to go back to the case of OLIVER from chapter 1. You will remember that Oliver is the conspiracy theorist with a 9/11 obsession. He insists that the collapse of World Trade Center towers on 9/11 was caused by explosives planted by government agents rather than by aircraft impacts. He thinks that the 9/11 Commission Report was part of a grand conspiracy to deceive the public and that, to coin a phrase, “the truth is out there”. He focuses on these propositions:

P – Aircraft impacts couldn’t have caused the collapse of the twin towers, and eye witnesses heard explosions just before the collapse of each tower, some time after the planes struck

Q – The collapse of the twin towers was caused by explosives rather than by aircraft impacts.

Oliver believes there is good evidence for P, and reasons from P to Q. P doesn’t entail Q but (as Oliver sees it) strongly supports Q. So Oliver’s reasoning is of the form “P, so Q”, though the “so” is not the “so” of entailment. Now ask Oliver why he believes that Q. He will be more than happy to tell you. He believes that Q because he believes that P and that P shows that Q.

Should we accept Oliver’s explanation? Suppose we flesh out the story a little: it turns out that Oliver loves conspiracy theories. He has conspiracy theories about the assassination of JFK, alien landings in New Mexico, and all manner of other things. He is biased to believe such theories and to disbelieve official accounts. He is generally gullible and has a poor grasp of logic, statistics, and probability. He is jumps to conclusions and has little sense of his own cognitive limitations. These are all statements about what might be called Oliver’s “epistemic character”. Bearing all this in mind, let’s ask again: why does Oliver believe that Q? At this point, it’s hard not to think that Oliver’s own explanation in terms of the logical or evidential relations between various things he believes about 9/11 is extremely superficial. The problem with Oliver is that he has a crazy view of what happened on 9/11, and the deep explanation of this fact is an explanation in terms of his epistemic character. Oliver can talk all he likes about how the various things he believes about 9/11 fit together. Perhaps, in a certain sense, they do all fit together, but that doesn’t mean that what he believes is true, or that describing his conception of the relationship between P and Q is enough to explain why he believes Q. It also needs to be added that Oliver has the beliefs he has about 9/11 because he is the way he is. This explanation of Oliver’s beliefs in terms of his character is based on reasoning, but reasoning that is external to Oliver’s own reasoning from P to Q; rather, it is reasoning from evidence about Oliver’s epistemic character to his beliefs about 9/11.

Oliver’s epistemic character comes out in lots of different ways. His view of P is one example: he thinks aircraft impacts couldn’t have caused the towers to collapse because he has read statements to that effect on 9/11 conspiracy websites. He attaches insufficient weight to studies which refute such statements. He doesn’t consider obvious alternative explanations of the sounds witnesses are supposed to have heard on the day, and accepts that they heard explosions, in a sense that implies the presence of explosives. His interpretation of the alleged “evidence” for P and Q manifests a range of epistemic character defects, and his beliefs about 9/11 only really make sense to us in light of these character defects. These defects, if they are genuine character defects, will affect his thinking on other topics besides 9/11 but may well be accentuated in this case by his desire for an explanation that is, as it were, proportionate in its scale and complexity to the scale of what happened on 9/11.

To explain a belief in terms rational linkages to other beliefs or to supporting evidence is to explain it in epistemic terms. Such explanations, which Ward Jones labels ‘epistemically rationalizing doxastic explanations’ (2002: 220), explain by showing that the target belief was brought about by a process which should lead to a true belief. Oliver’s own explanation of his beliefs is epistemically rationalizing. The suggested explanation in terms of epistemic character defects is not epistemically rationalizing. In the terminology I used in chapter 2, it is an undermining non-epistemic explanation, in the sense that the belief it explains would be threatened if Oliver were to accept the explanation. As Jones puts it, ‘if and when I become convinced, rightly or wrongly, that the right explanation for a belief is non-epistemic, then the grip of that belief will be loosened’ (2002: 223). If this is correct then Oliver is the one person who can’t believe that he only believes Q because of an epistemic character defect; he can’t believe this while continuing to believe that Q.

But is it true that Oliver only believes that Q because of his character defects? Oliver does, after all, reason from P to Q, and thinks that he believes that Q because he believes that P. Who are we to say he is wrong about this? There are two issues here: one is whether, in general, we have privileged access to why we believe what we do. The other is whether, aside from any considerations of privileged access, we are entitled in this particular case to dismiss Oliver’s own account of his belief that Q. Before tackling these questions head on it would be worth taking a look at the closely related discussion in a famous paper by Nisbett and Wilson. First-person explanations of one’s standing attitudes aren’t their main concern, but pessimists will interpret what Nisbett and Wilson argue as directly applicable to such explanations, and as vindicating both pessimism and inferentialism.

One of Nisbett and Wilson’s central findings is that ‘people may have little ability to report accurately on their cognitive processes’ (1977: 247). In particular, people are not at all good at detecting influences on their evaluations, choices, or behaviour. In one study, people were asked to evaluate four identical pairs of nylon stockings. There was a pronounced left-to-right position effect, with the right-most pair being preferred over the left-most by a factor of almost four to one. However, ‘when asked about the reasons for their choices, no subject ever mentioned spontaneously the position of the article in the array’ (1977: 243-4). It is difficult not to think of this as a case of self-ignorance: people were actually being influenced in their evaluations by positional factors of which they had no knowledge. They knew which pair they preferred but not why. Another study showed that people are increasingly less likely to assist others in distress as the number of witnesses or bystanders increases. Yet the subjects seemed ‘utterly unaware of the influence of other people on their behaviour’ (1977: 241). In every example of this kind, there are significant influencing factors to which people are blind, even to the extent of vehemently denying that such factors could have been influential when this possibility is raised by the experimenter.

What is going on? What accounts for our self-ignorance in such cases, and how do we ever get it right when it comes to explaining our own choices and evaluations? Nisbett and Wilson’s hypothesis is very much in line with inferentialism:

We propose that when people are asked to report how a particular stimulus influenced a particular response, they do so not by consulting a memory of the mediating process but by applying or generating causal theories about the effects of that type of stimulus on that type of response. They simply make judgements, in other words, about how plausible it is that the stimulus would have influenced the response (1977: 248).

People give defective explanations when they rely on dubious assumptions about the link between stimulus and response. Even correct reports are ‘due to the incidentally correct employment of a priori causal theories’ (1977: 233). Either way, the question ‘Why did you prefer/ choose/ do that?’ is answered by means of a theory-mediated inference. If you have a false belief about why you chose as you chose then you are, to this extent, self-ignorant, and your self-ignorance is, as inferentialism predicts, the result of a faulty inference.

How does this apply to explanations of one’s own standing attitudes? Let’s start with explanations of one’s own desires. Suppose you are a lapsed smoker and that you suddenly and unexpectedly find yourself with the desire for a cigarette. It’s a long time since you last wanted to smoke and you ask yourself why you want to smoke now. You have been under quite a bit of stress recently – you are writing a book on self-knowledge, perhaps- and you convince yourself that that’s why you want to smoke. In fact, that has nothing to do with it; you actually want to smoke because you have just watched a film in which the sympathetic lead character does a lot of stylish smoking. You have a theory about why you want to smoke but your theory is no good; the true explanation of your desire is much more prosaic than you realize. In other cases, the true explanation might be less prosaic. For example, Nietzsche speculates that our desires are explained by the presence of certain drives, such as the drive to sociality, to knowledge, to fight, to sex, and to avoid boredom. Be that as it may, discovering the true explanation of a desire doesn’t necessarily “undermine” the desire. The realization that you only want to smoke because you have just seen a particular movie doesn’t extinguish your desire for a cigarette or make it any less intense.

Beliefs are different. If Oliver were ever to be persuaded that he only believes that Q because of epistemic character defects then that will presumably loosen the grip of his belief. However, I have envisaged Oliver as insisting that he believes that Q because he believes that P, and that P shows that Q. My question was, ‘Who are we to say he is wrong about this?’. It doesn’t matter whether P does show that Q; what matters is whether Oliver thinks that P shows that Q and that that’s why he believes that Q. We can now see how to respond to this: the first thing to say that beliefs are like other attitudes in respect of our knowledge of why we have them. Just as we are sometimes wrong about why we want the things we want, or do the things we do, there is no guarantee that our beliefs about why we believe what we believe are correct. Having said that, it’s also plausible that Oliver’s reasoning “P, so Q” is part of the explanation for his believing that Q. The issue is whether it is the entire explanation or the deepest explanation, since one and the same belief can be explained in different ways and at different levels. Oliver wouldn’t believe Q if he didn’t believe P but it’s also true that he would believe neither P nor Q if he weren’t biased to believe conspiracy theories.

Let’s agree, then, that there is one explanation of Oliver’s belief that Q in terms of his reasoning and another explanation in terms of his epistemic proclivities. Both explanations have something going for them, but in what sense is the latter a “deeper” explanation? The thought is that it is deeper in the sense that it places Oliver’s reasoning in this instance in the context of his reasoning about other related matters. The explanation in terms of his epistemic character gives us an insight into the person that Oliver is, whereas merely talking about his inferential transitions in isolation doesn’t do that; it doesn’t explain why a particular claim or transition which in reality has little going for it is appealing to Oliver. What he has is a world view, and that is what the non-epistemic explanation enables us to understand. You can argue about whether talk of one explanation being “deeper” than another is defensible, but the point about self-ignorance doesn’t turn on the use of that terminology. The crux of the matter is that non-epistemic factors seem to be playing an important role in Oliver’s thinking, and their role is unacknowledged by Oliver. His self-ignorance is curable but at a price: in principle he could infer that non-epistemic factors are playing an important role in his thinking, but accepting that this is so would require him to rethink his beliefs about 9/11.

The pessimism I’ve been defending on the basis of OLIVER is a moderate rather than an extreme Nietzschean pessimism. Nietzsche argues that ignorance of our own attitudes in the ‘knowing why’ sense is incurable. My pessimist allows that you can sometimes infer why your attitudes are as they are, but insists on the possibility of self-ignorance even where your reasoning is as simple as “P, so Q”. Nietzsche seems to think that non-epistemic factors are what always what explain your attitudes but this isn’t something moderate pessimism needs to say. Imagine you are a master logician who reasons “P, so Q” without being influenced by anything other than the fact that P genuinely entails Q. In such cases of “pure” or “pristine” deliberation, non-epistemic factors might indeed be playing no role, and Boyle might be right that what puts you in a position to know why you believe Q is your successful deliberation. Even so, the assumption that non-epistemic factors are playing no role is justified, to the extent that it is, by an implicit theory of you, an implicit theory of the kind of consideration that is or is not likely to be influencing your thinking in the case at hand.

It’s worth adding that the purity of the master logician’s thinking is rarely replicated in real life. For most of us, most of the time, reasoning is a messy business; it’s a matter of drawing less than certain conclusions from less than perfect evidence. The range of factors which can influence “impure” reasoning or attitude-formation is bewilderingly large, which is why there is always at least the possibility that one’s thinking is being influenced by factors that are beyond one’s ken. The conclusions we come to are a reflection of the weight to attach to one kind of evidence over another, one theory over another. It would be nice to think that our weightings are appropriately grounded, and no doubt they sometimes are. But when there is a bias to believe, there is a corresponding bias to attach undue weight to some kinds of evidence and to discount others. The self-ignorance that pessimism describes is ultimately a reflection of how bad we are at detecting such contortions and distortions. If you don’t know that you are selectively privileging certain kinds of evidence then you don’t know why you believe the things you believe on the basis of that evidence.

Not knowing why your attitudes are as they are is one respect in which you might lack substantial self-knowledge. Corresponding to other varieties of substantial self-knowledge are other varieties of self-ignorance: ignorance of your character, value and emotions. Ignorance of one’s character is easy to explain, and some of it may well be motivated. We all like to think well ourselves, and this can lead us to be self-deceived about our character traits. Aside from motivated self-deception there is also the possibility of that you are ignorant of aspects of your own character because you lack the necessary conceptual resources or fail to grasp the relevance of certain kinds of evidence for the purposes of assessing your character: for example, you have evidence that you are fastidious but fail to infer you are fastidious.

Ignorance of your own values sounds more mysterious. How can you value equality without realizing it? If valuing equality is a matter of desiring to desire equality then there is no mystery: you can desire to desire something without realizing it because such desires are not self-intimating. It might come out in your treatment of others and your political and other preferences that you value equality, but you might not be sufficiently self-aware to grasp that an underlying concern with equality is what organizes your thinking across a wide range of social and political issues. Knowing your own values is, as I argued in chapter 12, a matter of interpreting your patterns of desire and thought on the basis of an understanding of what is, and what is not relevant to having certain values rather than others. As long as knowledge of one’s values is viewed as a substantial cognitive achievement, as a form of self-insight, it has to be allowed that it is a form of self-insight that it is possible for a person to lack.

Ignorance of your own emotions is straightforwardly possible in the case of complex emotions like love. Marcel infers he loves Albertine from his suffering on hearing she has left but suppose that Albertine had decided to stick around, or that Marcel didn’t hear news of her departure. In either case, he would still have loved her but not known that he loved her. When it comes to what might be regarded as less complex emotions, such as fear, it might seem harder to conceive of the possibility of self-ignorance. If self-knowledge of simple emotions is non-inferential, then self-ignorance in these cases can’t be the result of flawed inferences. The inferentialist’s reply is to argue that even knowledge of supposedly simple emotions like fear is inferential, or at least has a significant inferential component, and that self-ignorance in these cases can therefore be explained along inferentialist lines: you can be afraid without realizing it because you haven’t reflected, or you infer that what you are feeling is something other than fear. Since the dividing line between different emotions isn’t always sharp, there is always the possibility self-ignorance due to the subject misidentifying one kind of emotion for another.

The idea that knowledge of simple emotions is inferential might seem far-fetched but has empirical support. There is the old study by Valins and Ray which describes how subjects infer their level of fear of snakes from false information about changes in their heart rate. The snake-phobic subjects in the experiment were played recordings of what they believed falsely were their own heart beats. Then they were shown various slides, including slides of snakes. The snake slides weren’t accompanied by any change in their apparent heart rate, from which the phobic subjects apparently inferred that they weren’t as afraid of live snakes as they had previously thought. As a result, they were more willing to approach live snakes. This case is interesting because not only is the level of fear inferred, but the inference changes the actual level of fear. It is also suggestive that the inference is an inference from bodily data: given the connection between simple emotions and bodily changes (flushing, blushing, changes in heart rate and temperature), it comes as no surprise that such changes are often the basis on which a person interprets, and thereby knows, his own emotions.

In this chapter I set out to answer these three questions about self-ignorance:

1. How prevalent is it?
2. What are its sources?
3. To what extent can it be overcome?

I’ve concentrated on 2, on the idea that self-ignorance sometimes results from motivational factors, and sometimes from other factors, such as insufficient evidence, misinterpretation of the evidence, failure to perform the necessary inferences, and so on. To the extent that I have identified some sources of self-ignorance I have explained how self-ignorance is possible, but explaining how something is possible is different from demonstrating that it’s prevalent, or even actual. So the question remains: how prevalent is self-ignorance? For all that I’ve said optimism is still an option: couldn’t you think that self-ignorance is possible and explicable along inferentialist lines, but that in reality humans aren’t actually self-ignorant?

You could think this but it wouldn’t be a very sensible thing to think. Suppose you are convinced that self-ignorance is caused by a mixture of motivational and non-motivational factors. In that case, the more common these factors are the more prevalent one would expect the resulting self-ignorance to be. The prevalence among humans of the factors that cause us to be self-ignorant is, at least to some extent, an empirical matter. It’s an empirical question how prone we are to misinterpreting the behavioural and psychological evidence for our own attitudes, or to what extent we are capable of avoiding various kinds of bias in thinking about our own characters. No doubt there are psychological studies that bear on these questions, but you don’t have to have read these studies to realize that the cognitive vices which lead to self-ignorance are far from rare or unusual; reading great novels and talking to your friends would do just as well. Optimists who question whether self-ignorance is prevalent must either deny that the cognitive vices I have been describing in this chapter are prevalent or deny that these vices result in self-ignorance. Neither denial is remotely plausible.

Having said that, it must also be admitted that there is something odd about discussing the prevalence of self-ignorance among humans, as if all humans are the same in this respect. We aren’t equally reflective or sophisticated. Some of us reason better than others and engage in self-inquiry more than others. There are character traits you can only know you have if you have certain concepts which not all humans have. You might learn about yourself by reading great literature but we don’t all have the time, energy, or inclination to read Proust. The point of saying this isn’t to suggest that only the clever or educated can avoid self-ignorance. The truth is that no human can avoid being self-ignorant to some degree because the factors which lead to self-ignorance are so powerful and pervasive. All the same, individual differences do affect the degree as well as the type of self-ignorance individuals suffer from. We aren’t all the same.

To what extent, and by what means, can self-ignorance be overcome? One way of approaching this is, at least initially, as a practical question: assuming you are as self-ignorant as the next man or woman, what can you do to overcome your self-ignorance? Once we have a list of practical steps we can assess their chances of success and thereby estimate the extent to which it might be possible for us to overcome the self-ignorance to which all humans are liable. This practical approach is in keeping with the suggestion in chapter 5 that an account of self-knowledge for humans might be expected to provide guidance to those of us who seek self-knowledge. I described this as ‘self-knowledge for human in the guidance sense’, and it’s reasonable to think that guidance to those who seek self-knowledge should include guidance as to the most effective ways overcoming of self-ignorance.

Sometimes overcoming self-ignorance requires no special measures because there is no obstacle that needs to be overcome. Before she has thought about it Katherine didn’t know she wanted another child. When she wonders whether she wants another child it might be obvious to her that she does, and there need be nothing that blocks this realization. There is a smooth transition in this case from self-ignorance to self-knowledge but no ‘overcoming’ of self-ignorance except in the sense that Katherine comes to know something about herself she didn’t previously know. It’s more natural to talk about a person ‘overcoming’ self-ignorance when there is an obstacle to self-knowledge or when special cognitive effort is required. This suggests that we should be concentrating on substantial self-knowledge and on practical steps for overcoming self-ignorance with regard to one’s own character, emotions, and so on.

Suppose that the obstacles which prevent you from acquiring substantial self-knowledge are inattention, poor reasoning, or misinterpretation of the evidence. In that case, it might seem that the way to overcome self-ignorance is to pay attention, reason better and be careful not to misinterpret the evidence. These are all improvements you might achieve by thinking ‘slow’ rather than ‘fast’. The suggestion is that careful, patient and slow self-inquiry is the key to overcoming self-ignorance, and that the more careful and patient you are the more likely you are to avoid self-ignorance. If, on the other hand, the source of your self-ignorance is motivational, then the key is to recognize that this is so. You need to be open to the idea that there may be truths about yourself you have difficulty seeing because they are unpalatable or anxiety-provoking. Acquiring self-knowledge in such cases is a matter steeling yourself, and making sure that you self-inquiry is as honest as possible, with as little wishful thinking as possible.

There is something to this, but less than meets the eye. Focusing your attention on your own character and emotions might end up distorting the very psychological facts you are trying to uncover. Self-enquiry can be self-defeating, especially if it turns into self-obsession, and the vision of someone spending a lot of time and energy in pursuit of self-knowledge is in any case not especially an attractive one. Slow thinking in the context of self-enquiry might help you to avoid certain types of illusion about yourself, but if your self-ignorance results from false assumptions or a poor background theory then thinking slowly on the basis of such assumptions or such a theory isn’t necessarily going to help. However hard you try, you might find it impossible not to self-attribute to feelings and attitudes you don’t have.

Other practical measures for overcoming self-ignorance are no less problematic. What about seeing ourselves through the eyes of others? This is less solipsistic than the project of overcoming self-ignorance through isolated self-enquiry but Wilson and Dunn point out that we aren’t god at detecting how other people view us when their views are different from our own: ‘rather than taking an objective look at how other people view them and noticing the fact that this view might differ from their own, people often assume that other people see them the way they see themselves’ (2004: 17.16). As for observing your own behaviour and tackling one’s self-ignorance on that basis there is always the danger of this resulting in ever more sophisticated fabrications rather than self-revelation.

This adds up to a pessimistic view of the prospects for overcoming self-ignorance. It’s not that there is nothing you can do to tackle the most challenging forms of self-ignorance; no doubt the practical steps I have described are helpful to some extent but it’s important not to exaggerate their prospects of success. The worst form of self-ignorance is ignorance of one’s own self-ignorance, and overcoming such second-order self-ignorance isn’t so much a matter of engaging in prolonged self-inquiry as approaching questions about the extent which self-ignorance can be overcome in a spirit of humility: the unknown unknowns about the self need to become known unknowns. There is just no getting away from the fact that substantial self-knowledge is often hard to get, and that we have less of it than many of us we like to think in our more optimistic moments. We need to be realistic, and that means acknowledging the full extent to which human beings can be, and frequently are, as opaque to themselves as they are to each other.

Should we care about the pervasiveness and intractability of self-ignorance? To the extent it’s possible to overcome some of our self-ignorance by therapy, self-inquiry or some other effortful means is it worth the effort? That depends on the value of self-knowledge. It’s easy to see why some forms of self-knowledge, such as knowledge of your own abilities, has practical value but what about knowledge of your own attitudes or character? What possible use is that? If most self-knowledge is of little value – practical or otherwise- then pessimism about the prospects of overcoming self-ignorance is something we can happily live with. Yet both philosophers and non-philosophers tend to assume that self-knowledge is valuable, and that more is better than less. The next question is whether they right about this.