

PREFACE

One of the downsides of writing about self-knowledge is having to explain yourself to the philosophically uninitiated. Flights and dinner parties are usually where this happens, and the problem you face isn't the usual one. It's not that self-knowledge strikes people as being too boring or weird or *recherché* to be worthy of philosophical attention. Far from it. Self-knowledge as a philosophical topic sounds, as one tires of being told, fascinating. Indeed, it's just the sort of subject which non-philosophers expect philosophers to be interested in.

The disappointment sets in when the time comes to try to explain what philosophers these days mean by "self-knowledge". You might try saying this: self-knowledge, as many philosophers nowadays understand it, is first and foremost knowledge of one's own states of mind, that is, knowledge of such things as one's own beliefs, desires, and sensations. In the interests of clarity, you might add that the states of mind that are at issue here needn't be particularly deep or elusive or important. Suppose you believe that you are wearing socks and know that this is what you believe. What philosophers of self-knowledge typically focus on is this kind of seemingly trivial or easy self-knowledge. Of course they don't claim that this is all there is to self-knowledge. There is also knowledge of your deepest desires, hopes and fears, knowledge of your character, emotions, abilities and values, and knowledge of what makes you happy. These are examples of what you might call substantial self-knowledge, but you have to admit that substantial self-knowledge, for all its undoubted human interest, isn't where the philosophical action is. What philosophers find interesting isn't how you can know your own character or abilities but much more mundane examples of self-knowledge such as your knowledge that you believe you are wearing socks or that you want to have ice cream for pudding.

This is usually news to non-philosophers. They imagine philosophy tries to answer the Big Questions, and “How do you know that you believe you are wearing socks?” doesn’t sound like a Big Question. They don’t get why you would be more interested in trivial than in substantial self-knowledge, and are surprised that you would want to describe knowledge of your more mundane beliefs as “self-knowledge”. So what? The usual move at this point is to argue that what is philosophically interesting can’t be decided by what non-philosophers think is interesting, and that a big part of what makes knowledge of one’s more mundane states of mind philosophically interesting is that it is epistemologically special or distinctive. It’s unlike substantial self-knowledge or ordinary worldly knowledge since it isn’t based on evidence. To know that you are wearing socks you need evidence that you are wearing socks (you can feel them on your feet) but to know that you believe you are wearing socks you don’t need evidence that you believe you are wearing socks. To know that you have a particular character trait, say fastidiousness, you need behavioural evidence from which you can infer that you are fastidious. In contrast, you don’t normally infer your own beliefs from behavioural or, for that matter, any other evidence. But if knowledge of your own mundane beliefs and desires is, in this sense, immediate, then it would be natural to ask how such epistemologically immediate self-knowledge is possible. This is the question which many philosophers of self-knowledge are trying to answer. They don’t worry about how substantial self-knowledge is possible, because it isn’t special.

I don’t buy this, and that is partly why I decided to write this book. For a start, even if substantial self-knowledge isn’t epistemologically distinctive, that doesn’t make it unworthy of philosophical attention. There is more to philosophy than epistemology. In any case, the story I’ve been telling exaggerates the epistemological distinctiveness of trivial or mundane self-knowledge. I claim that this kind of self-knowledge is normally based on evidence and is inferential. This is currently a deeply unpopular view among philosophers and one of my

aims here is to defend it. In addition, there is also very little to be said in favour of the assumption that inferences from behavioural evidence are the sole or even the primary basis on which one knows such things as one's own character and values. As I hope to show, the different varieties of substantial self-knowledge are much more epistemologically interesting than is commonly supposed, and this means that there are sound philosophical reasons for taking this kind of self-knowledge seriously.

Anyway, it's not just about philosophical interest or importance of self-knowledge. There is also its human importance. The self-knowledge which most reflective human beings tend to regard as important is substantial self-knowledge, and they expect philosophy to have something to say about its nature, scope, and value. I don't think that this is an unreasonable expectation. Philosophy needs to pay some attention to forms of self-knowledge that seem humanly important, and this means engaging with substantial self-knowledge. As it happens, I believe that the value of all forms of self-knowledge is something that people often overestimate, but philosophy had better have something to say about the value of self-knowledge, as well as about why the kinds of self-knowledge which many of us value are so hard to get. The other side of substantial self-knowledge is self-ignorance, and this is another neglected topic about which I have tried to say something.

The desire to engage with humanly important varieties of self-knowledge is one of the factors which led me to write this book, but not the only factor. Another factor is the difficulty I have in recognizing real human beings in philosophical descriptions of how self-knowledge is acquired. The philosophical account of self-knowledge which I have the most problems with and which is in many ways the target of this book is what has come to be called Rationalism about self-knowledge. In its simplest form, Rationalism says that you can know whether you do believe a given proposition P by reflecting on whether you ought rationally to believe that P. You assume that normally what you believe is what you ought

rationally to believe, and this assumption is what enables you to know that you believe that P by knowing that you ought rationally to believe that P. The same goes for your desires, hopes, fears and other attitudes. In each case, knowing what your attitude ought rationally to be puts you in a position to know what your attitude is.

This strikes me as a strange idea, not least because it is often much easier to know that you have a given attitude than to know whether you ought rationally to have it. In such cases Rationalism substitutes a much harder question for a much easier one. For example, it might be obvious to me that I would rather spend the morning relaxing rather than writing, but very far from obvious whether this is how I ought rationally to want to spend the morning. In any case, the assumption that our attitudes are normally as they ought rationally to be seems a bit optimistic if it is real human beings we are talking about. Just as behavioural economists distinguish between homo sapiens and homo economicus, the idealised rational agent of so much economic theorizing, so I think it's helpful to distinguish real humans as we know them from homo philosophicus, the idealized subject of so much philosophical theorizing.

Homo philosophicus is a model epistemic citizen. His attitudes are as they ought rationally to be, his beliefs are properly based on the evidence available to him, and when he encounters good reasons to change any of his current attitudes he changes them. As I see it, Rationalism is an account of self-knowledge for homo philosophicus rather than for humans. Homo philosophicus can know his beliefs by reflecting on what the reasons require him to believe, but for human beings matters are far less straightforward. Human beings are far from being model epistemic citizens and the extent to which our attitudes are out of line with our reasons makes it hard for us to know what our attitudes are by reflecting on what they ought rationally to be. What we should be looking for is an account of self-knowledge for humans as they actually are, and Rationalism about self-knowledge isn't it. What I'm after is an alternative to Rationalism which doesn't underestimate our cognitive failings and limitations.

I call the respects in which humans are unlike homo philosophicus the Disparity, and my account of the Disparity draws heavily on the work of behavioural economists and social psychologists, including Daniel Kahneman, Richard Nisbett, Lee Ross and Timothy Wilson. Kahneman's book Thinking Fast and Slow has been an especially significant influence on my thinking. It has recently become fashionable for behavioural economists to regard the Disparity as proving that humans are irrational. However, Kahneman refrains from drawing this conclusion and so do I. I think it's unhelpful to think in these terms, and that not being homo philosophicus doesn't make human beings irrational. Acknowledging the Disparity isn't about convicting us of irrationality but of trying to be realistic in what we say about how we reason and come to know ourselves.

I wrote this book during a year of research leave funded by the Mind Association. It's a huge honour to have been elected to a Mind Senior Research Fellowship, and I'd like to take this opportunity to thank Mind for its support. Thanks also to Peter Momtchiloff at OUP and to his two readers, Lucy O'Brien and David Finkelstein. Lucy kindly read the chapters as I was writing them, and discussions with her in the early stages of the project helped me to clarify my thinking. I had the terrible idea of calling the book Reality Check: Self-Knowledge for Humans, and she persuaded me to stick to Self-Knowledge for Humans. I also wrote chapter 12 in response to a question she raised. David's comments on the whole draft were also incredibly helpful, and I made a number of changes in response to them. I also need to thank John Campbell for a superb set of comments, to which I haven't really been able to do justice, and Naomi Eilan for her valuable input and encouragement. Wayne Waxman also sent me many useful comments and questions. Discussions with Deborah Ghate helped me to come up with an account of the value of self-knowledge in the final chapter, and I was also helped by conversations over lunch with Bill Brewer.

I've given talks and lectures based on ideas from the following chapters in many places, including: Barcelona, Bonn, Chicago, Copenhagen, Edinburgh, Fribourg, London, Luxembourg, Porto, Reading, St. Andrews, Stirling, Stuttgart, Sussex, and Zurich. Thanks to the audiences on those occasions for showing up and asking difficult questions. I presented a draft of the book to an MA class at Warwick in 2013. The quality of discussion and student presentations was excellent, and I got a lot out of it. I won't attempt to name the students who came but I'm grateful to all of them.

The style of this book is conversational, and I've made a conscious effort to say things in writing more or less as I would say them if I were speaking. I've also tried to do justice to what makes self-knowledge such an interesting topic, and to address at least some of the questions about self-knowledge which non-philosophers ask. Writing a boring book on a boring topic is one thing but writing a boring book on an interesting topic is inexcusable. I have the sense that I've only scratched the surface of some of the issues I discuss here, and I would like to say much more in future about the value of self-knowledge. The Disparity is another topic about which there is more to be said. For moment, however, the pages that follow will have to do.