9

LOOKING OUTWARDS

Way back in chapter 1 I talked about the rationalist approach to self-knowledge. Rationalism tries to account for our knowledge of our beliefs, desires, fears, and other such ‘intentional’ states. The central idea of rationalism is that using the Transparency Method is, for us, a basic source of ‘intentional’ self-knowledge, that is, knowledge of our own beliefs, fears, desires etc. A lot of what I’ve been calling ‘substantial’ self-knowledge isn’t accounted for by TM but that’s not an objection to rationalism, any more than it’s an objection to rationalism that it doesn’t try to explain how we know our own sensations: it’s neither necessary nor possible to account for all self-knowledge in one fell swoop, and it would be a good start to at least have a successful account of how we know our own beliefs, desires etc. This brings us to the main question I want to address in this chapter: is rationalism any good at least as an account of intentional self-knowledge?

In chapter 1 I mentioned several worries about rationalism about self-knowledge, one of which is that it runs into trouble with the Disparity: maybe homo philosophicus can work out whether he believes that P by asking whether he ought rationally to believe that P but it’s not clear that that’s going to cut it for homo sapiens. This led to a discussion of the extent of the Disparity but in this chapter I want to leave that discussion behind. My question here is whether rationalism can live with the Disparity, that is, whether you can accept that there is an extensive Disparity and still maintain that TM is for us a basic source of intentional self-knowledge. I have referred to this combination of views as ‘compatibilism’, and one way of starting to make the case for compatibilism is to question my understanding of rationalism. It might be argued, in particular, is that what I’ve been calling ‘rationalism’ is a rather crude and simple-minded version of that doctrine – call it Simple Rationalism- and that there is a more subtle form of rationalism which can deal with the Disparity much better than Simple Rationalism. For reasons which will become clearer in a while I will refer to this more subtle form of rationalism as Activism.

Whichever kind of rationalist you are, one thing you are quite likely to think is that intentional self-knowledge is epistemically privileged. Some rationalists think that intentional self-knowledge is infallible or at least authoritative: your beliefs about you own beliefs and other attitudes can’t be mistaken or, even if they can be, there’s a presumption that they aren’t mistaken. Other rationalists emphasize what they think of as the ‘immediacy’ of intentional self-knowledge, the fact that this kind of knowledge is normally non-inferential and not based on evidence. So there are two questions for rationalism, whether Simple or Activist: one is whether, bearing in mind the Disparity and the other difficulties I mentioned in chapter 1, it explains how we can know our own beliefs, desires, and other attitudes. The other is whether, on a rationalist construal, intentional self-knowledge comes out as epistemically privileged knowledge. Rationalism is in trouble if the answer to either question is ‘no’.

This brings me to the position I want to defend in this chapter: although I think that it is possible for us to know some of our attitudes by using TM, I doubt that the self-knowledge we gain by using TM has all the epistemic privileges rationalism is after, or that this kind of self-knowledge is especially basic or fundamental. In fact, TM is one of a range of pathways to intentional self-knowledge, all of which can only deliver indirect and evidence-based self-knowledge. There is also a connection with what I was saying in chapter 3. There I drew a distinction between ‘substantial’ and ‘trivial’ self-knowledge and pointed out that knowledge of your own attitudes – say your belief that all races are equal, or that you want another child – can be substantial. Interestingly, the intentional self-knowledge which rationalism does by far the best job of accounting for is the trivial variety; the more substantial your intentional self-knowledge, the less convincingly TM accounts for it. So the more strongly you believe that the philosophy of self-knowledge should focus on substantial self-knowledge, the less satisfied you should be with rationalism about self-knowledge. Even when it comes to trivial self-knowledge, rationalism runs into a variety of problems, not all of which are connected to the Disparity. In the end I think that the right thing to say about TM is that it is a minor player in the true story about intentional self-knowledge, but that it’s not an objection to TM that it only delivers evidence-based self-knowledge: pretty much any self-knowledge that is worth having is going to be evidence-based.

In getting to this conclusion I want to begin by reminding you of the main tenets of rationalism, and then introducing the distinction between two versions of Rationalism as a response to some of the problems I mentioned in chapter 1. The way I introduced rationalism in that chapter was by quoting Gareth Evans as saying that he gets himself in a position to answer the question whether he believes that P by putting into operation whatever procedure he has for answering the question whether P. In making a self-ascription of belief, Evans says, ‘one’s eyes are, so to speak, or occasionally literally, directed outward- upon the world’ (1982: 225). This idea is taken up by Richard Moran in his book Authority and Estrangement. In Moran’s terminology, the question “Do I believe that P?” is an inward-directed question, whereas the question “Is P true?” is outward-directed. These questions have different subject-matters, but Moran argues that the inward-directed question is as he puts it ‘transparent’ to the corresponding outward-directed question. Furthermore:

[I]f the person were entitled to assume, or in some way even obligated to assume, that his considerations for and against believing P (the outward-directed question) actually determined in this case what his belief concerning P actually is (the inward-directed question), then he would be entitled to answer the question concerning his believing P or not by consideration of the reasons in favour of P (2004: 457).

In brief:

(a) To say that the question whether you believe that P is transparent to the question whether P is true is to say that you can answer the former question by answering the latter question.

(b) What makes it possible for you to answer the inward-directed question by answering the corresponding outward-directed question is your assumption that your belief concerning P is determined by your reasons for and against believing P.

What makes this a rationalist account of self-knowledge is that it takes your belief concerning P to be determined by your reasons and so to be knowable by reflecting on your reasons.

Both Evans and Moran reckon that the transparency account can explain the epistemic privileges of self-knowledge. Evans talks about his procedure not allowing ‘even the most determined sceptic’ to ‘insert his knife’, which suggests that he thinks that his version of the Transparency Method delivers a kind of infallible self-knowledge. For Moran, the key issue isn’t infallibility but immediacy. Sometimes what he means by immediate self-knowledge is knowledge not based on behavioural evidence. At other times he means knowledge not based on any evidence. Then there is the idea that immediate self-knowledge is non-inferential, which may or may not be equivalent to saying that is not based on evidence. There’s also the point that non-inferential knowledge can be either psychologically or epistemically immediate, and it’s not clear what kind of non-inferential self-knowledge TM is supposed to deliver. I’ll come back to this.

One question about TM which came up in chapter 1 was: how can it account of self-knowledge of attitudes other than belief? Even if the question “Do I believe that P?” is transparent to the question “Is it true that P?”, the same can’t be said for “Do I desire that P?” or “Do I fear that P?”. I’m going to call this the Generality Problem for rationalism: the problem is that the version of TM you get from Evans and Moran can only account for a sub-class of intentional self-knowledge. It isn’t just knowledge of your own sensations and lots of substantial self-knowledge that it can’t account for; it can’t even account for your knowledge of your own desires and fears.

Finkelstein’s solution to the Generality Problem on behalf of rationalism is to read TM as proposing that the question whether you believe that P is transparent to the question whether you ought rationally to believe that P; you can answer the first of these questions by answering the second, and the same method can be used to account for knowledge of your fears and desires: you can determine whether you desire that P by determining whether you ought rationally to desire that P, just as you can determine whether you fear that P by determining whether you ought rationally to fear that P. Determining whether you ought rationally to believe (or desire or fear) that P is a matter of asking yourself whether what Finkelstein calls ‘the reasons’ require you to believe (or desire or fear) that P. Suppose that you ask yourself this question and you judge that the reasons do require you to have the attitude in question. How can judging that you ought rationally to believe that P enable you to know that you do believe that P? Because, and only insofar as, you assume that the attitude you have is one that, by your lights, the reasons call for you to have. Finkelstein calls this the Rationality Assumption, and it’s tempting to see this assumption as mediating the transition from a claim about what your attitude ought to be to a claim about what your attitude is.

This way of putting things is a little problematic if you want your knowledge that you believe that P to come out as immediate. If your knowledge is mediated by the Rationality Assumption doesn’t that make it, by definition, mediate rather than immediate knowledge? In response it might be argued that the role of the Rationality Assumption isn’t to mediate but to enable TM-based self-knowledge, and that in any case the fact that a piece of knowledge is ‘mediated’ by highly general assumptions such as the Rationality Assumption doesn’t mean that it isn’t, in the relevant sense, ‘immediate’. But this is all a bit mysterious. For a start, it’s not at all clear how to distinguish between the idea that the Rationality Assumption mediates self-knowledge and the notion that it merely enables it. Also, why does the fact that an assumption is ‘highly general’ not threaten the immediacy of any knowledge that is ‘mediated’ by it? Maybe we shouldn’t be bothered by the fact that self-knowledge acquired by using TM isn’t immediate, but insisting that it is immediate is an entirely different matter.

The next problem with TM is that it represents us as substituting what is often a more difficult question (“Do the reasons require me to believe that P?”) for an easier question (“Do I believe that P?”). This is the Substitution Problem for rationalism. The idea of substituting one question for another is borrowed from Daniel Kahneman, but substitution in Kahneman’s sense is the exact opposite of the substitution involved in applying TM. Substitution in Kahneman’s sense happens when you answer a difficult question by answering an easier one. For example, you might find yourself answering the question “How happy are you with your life these days?” by answering “What is your mood right now?”. Or faced with “How popular will the President be six months from now?” the question you answer is, “How popular is the President now?”. The motto is: if you can’t answer a hard question, find an easier one you can answer. But when it comes to TM the motto seems to be: even if there is an easier question you can answer, find a harder question you can’t easily answer.

Why think that the question “Do the reasons require me to believe/ want/ fear that P?” is any harder than the question “Do I believe/ desire/ fear that P?”. The intuition here is that there are many occasions when it’s more obvious to you that you have a given attitude than that you ought rationally to have it, or that ‘the reasons’ require you to have it. As I sit down for dinner it’s perfectly obvious to me that I want to start with a vodka martini but I would be flummoxed if someone asked “Do the reasons require you to want a vodka martini?”. I have no idea what I ought rationally to want to drink, and if that is the case then why would I think that figuring out whether I ought to want to have a vodka martini is a good way of figuring out whether a vodka martini is what I want? I know without reflecting on my reasons that I want a vodka martini, just as I know without reflecting on my reasons that I’m scared of the spider in my bathtub.

Another way of making the point I am making here is to pick up on something that Jonathan Way says in an Analysis paper on the limits of transparency. Way’s specific target is the idea that we can know whether we have a given attitude M by asking ourselves whether to have M. Way points out that there are cases in which what is required is only that you have attitude M or attitude N, and that in such cases transparency doesn’t work because there is no such thing as ‘the attitude to have’. If I’m driving and there are two equally good routes to where I’m going, I can know which one I want or intend to take even though there is no sense in which I ought rationally to take that route rather than the equally good alternative. Equally, my evidence might be good enough to permit belief in some proposition P, without being so good as to require belief in P. As Way puts it, ‘the claim that there is always a uniquely correct attitude to take towards P, when one is considering whether P, remains a strikingly strong claim’ (2007: 228). The claim that there is always a uniquely correct attitude to take towards P is not just strikingly strong but downright implausible, and this generates a version of the Substitution Problem: the answer to “Do the reasons require me to believe that P?” might be no even though the answer to “Do I believe that P?” is plainly yes. Here it’s not just that the outward-directed question is harder but that it’s the wrong question given that there isn’t a uniquely correct answer to it, whereas there is (let’s suppose) a uniquely correct answer to the inward-directed question.

An especially interesting attitude to think about in relation to the Substitution Problem is the attitude of hoping that P. Suppose you have been single for a long time and that all your previous relationships ended badly. Still, you live in hope. You hope the next person you date will end up as your significant other even though your close friends think that your hopes will almost certainly be dashed. Their sage advice is: don’t hope for too much and you won’t be disappointed. So what do the reasons require you to hope? As you head for your next date would it be right to think that you ought rationally to hope that things will work out? The problem with these questions is not just that they are hard to answer but that it’s difficult to know how to even go about answering them. Presumably it’s permissible for you to hope for the best, but what would it be for hoping for the best in this case to be a rational requirement? Even Kant only asks “What may I hope?”, not “What must I hope?”. Figuring out what the reasons require you to hope for looks like a distinctly unpromising way of figuring out what you do hope for.

Obviously, all this talk of questions like “Do I believe that P?” being easier to answer in many cases than questions like “Do I believe that P?” begs an obvious question: how is the inward-directed question to be answered if not by answering the corresponding outward-directed question? What’s needed to make any progress with this is a positive theory of self-knowledge, and I will outline my own theory chapter 11. However, to avoid any confusion, it would be worth noting the following now: when I say that the answer to an inward-directed question is easier to come up with than an answer to an outward-directed question I’m not saying that it doesn’t require cognitive effort to answer an inward-directed question or that it is never appropriate to use TM. A more sensible view is that there are different pathways to knowledge of one’s own attitudes, and that it’s much easier in some cases than in others to know what your attitude is. Equally, it’s much easier in some cases than in others to know what your attitude ought rationally to be. The fact remains, however, that figuring out what our attitudes ought to be is often harder than knowing what they are, and this suggests that there has to be something wrong with the idea that using TM is a fundamental way for us to acquire knowledge of our own attitudes; the fact is that we often know what we want or fear or even believe without having to reflect on our reasons, and reflecting on our reasons might actually make it harder to know our own attitudes because our reasons are often opaque.

The next problem for TM is what I’m going to call the Matching Problem. Here is a crude statement of this problem: what you believe isn’t always what you ought rationally to believe, what you want may not be what you ought rationally to want, and what you fear may not be what you ought rationally to fear. In each of these cases, there is the possibility of a mismatch between you actual attitude and what your attitude ought to be (assuming now that it’s clear what your attitude ought to be). But if there is always the possibility of this kind of mismatch how can you determine what your attitude is by determining what it ought to be? If you judge that you ought to have a certain attitude M that doesn’t guarantee that you have M, so how can you know that you have M on the basis that you ought to have it?

A popular way of illustrating the Matching Problem is to start talking about belief perseverance. Brie Gertler characterizes belief perseverance as a ‘common psychological phenomenon that threatens the method of transparency’ (2011: 138). She says that it ‘occurs whenever a dispositional belief that P endures despite the discovery that one has no evidence for P, or that one’s evidence favours not-P’ (2011: 137). For example, people hold fast to the conviction that ‘their spouses are faithful, and their children innocent, despite abundant evidence to the contrary’. These are all cases in which people believe things that, as rational beings, they shouldn’t believe, and in which they will therefore go wrong if they try to figure out what they believe by reflecting on what they should believe.

One response to this which Gertler considers says that belief-perseverance is after all a relatively rare phenomenon, that one’s evidence will generally match one’s beliefs, and that ‘the method of transparency may achieve a degree of reliability that is high enough to qualify its results as knowledge’ (2011: 138). Gertler isn’t convinced because she thinks that belief perseverance isn’t all that rare. However, there is a question about how Gertler characterizes belief perserverance. As we saw in the case of KAREN, after the total discrediting of the sole evidence for her belief that she is especially good at science and music she has no evidence for this belief but she doesn’t realize that her sole evidence has been discredited. That is why her belief persists. Although in this case there is ‘the discovery’ that she has no evidence for her belief, this is not something that Karen has discovered. The whole point of the example is that the subject herself doesn’t know that she has no evidence for P, or that her evidence favours not-P.

The phenomenon Gertler has in mind is what I’ve been calling recalcitrance rather than belief perseverance. This is significant because while Gertler might be right that belief perseverance is relatively commonplace, recalcitrance is a different matter; how is it even possible, let alone commonplace, for someone to continue to believe that P despite knowing that her evidence for P has been totally discredited? Presumably, the wife who continues to believe her husband is faithful despite overwhelming evidence to the contrary doesn’t accept that there is overwhelming evidence to the contrary and so isn’t irrational in the sense that her attitudes fail to conform to her own judgements about what she is warranted in believing. So there is nothing to stop her from answering the question whether she believes her husband is faithful by answering the question whether this is something she ought rationally to believe.

In chapter 2 I tried to make sense of the possibility of belief recalcitrance by drawing on Harman’s idea that once a belief has become established a great deal of effort might be needed to get rid of it, even if the believer comes to see that he or see ought to get rid of it because all the evidence for the belief has been undermined. For example, I envisaged Karen as realizing that her evidence for the belief that she has an aptitude for science and music has been undermined but as still being disposed to think that she has an aptitude for science and music when the question arises and as using that thought that she has an aptitude for science and music as a premise in reasoning about what to do. She therefore still believes something she knows she oughtn’t to believe, which makes this a case of recalcitrance rather than belief-perseverance.

One reaction to this attempt to make sense of belief recalcitrance might be to suggest that after the discovery of the mix-up over the test results Karen no longer believes that she has an aptitude for science and music but alieves it. In Tamar Gendler’s terminology, an alief is a ‘habitual propensity to respond to an apparent stimulus is a particular way’ (2008: 557). What Harman sees as Karen’s ‘habits of thought’ are in fact belief-discordant aliefs. Unlike beliefs, aliefs are arational, and the fact that Karen realizes that she no longer has grounds for believing that she has an aptitude for science and music doesn’t mean that she doesn’t alieve that she has an aptitude for science and music, though it does mean that this is something she no longer believes.

The belief/ alief distinction helps TM deal with the Matching Problem. The worry for TM is that what you yourself think you ought rationally to believe might not match what you do believe, and that this stops you from determining what you believe by determining what you ought rationally to believe. This assumes that your beliefs can be irrational, but Gendler is suggesting that supposed examples of irrational belief are in reality examples of arational alief. Unlike your aliefs, your beliefs line up with what you take yourself to have reason to believe, and this allows you to use TM to discover your own beliefs. What we have here is a combination of rationalism about belief with rationalism about self-knowledge; the idea is that Karen is a problem for rationalism about self-knowledge if it’s an example of evidence-discordant belief but not if it’s an example of belief-discordant alief. Of course, what you believe can be out of sync with what there is reason to believe, or with what the evidence actually shows, but not with what you think there is reason to believe or what you take the evidence to show.

The idea that Gendler’s alief/ belief distinction can be used to deal with the Matching Problem is, though superficially appealing, no good at all. Obviously quite a lot depends on what you think belief is, but I’m inclined to agree with Scanlon that it’s sufficient for you to believe that P if you are disposed to think that P when the question arises and use the thought that P as a premise in practical reasoning. Since Karen satisfies these conditions with respect to the thought that she has an aptitude for science and music it follows that she believes (and doesn’t merely alieve) that she has an aptitude for science and music. It’s also worth pointing out that when it comes to attitudes other than belief, a Gendler-type response to the Mismatch Problem isn’t even superficially plausible. Suppose you have a liver problem and have been warned of the disastrous consequences for your health of continuing to drink vodka martinis. Should you still want a vodka martini before dinner this evening? No. Do you want a vodka martini before dinner this evening? Yes. Your desire for a vodka martini is recalcitrant, and it would be absurd to suggest that you don’t really desire a vodka martini, you only arationally ‘asire’ one. You can’t introduce new categories of mental state by linguistic fiat in order to keep your desires rational, and the same goes for your beliefs: what look like irrational beliefs that can’t be uncovered by using TM really are irrational beliefs. They aren’t mere ‘aliefs’ the positing of which allows your beliefs to be discovered by using TM.

Another superficially appealing but equally unconvincing way of dealing with attitude recalcitrance is to argue that recalcitrant attitudes are ones with respect to which we are in some way alienated, and that they aren’t a problem for rationalism because rationalism is only interested in accounting for unalienated self-knowledge, that is, knowledge of those of ones attitudes that are responsive to how they ought rationally to be by one’s own lights. It’s not clear where this restriction came from. There was certainly no sign of it in the initial set up of rationalism, and saying that rationalism doesn’t deal with alienated self-knowledge is a significant concession given the range of attitudes that are, to varying degrees, recalcitrant. Anyway, the idea that an attitude is alienated just in virtue of being recalcitrant has very little going for it. Alienated attitudes are ones that one can’t identify with, but in reality it is the attitudes that a person identifies with most wholeheartedly that are most likely to be recalcitrant. If a particular belief is fundamental to your self-conception or weltanschauung it’s hardly surprising that you find it very hard to give up despite realizing that it has been undermined.

So rationalism is stuck with the Matching Problem, and can’t get it itself off the hook by talking about aliefs or alienation. There remains the option of arguing that the kind of mismatch between what your attitudes are and what they ought to be that is a problem for TM is, though not impossible, nevertheless rare. As I’ve emphasized, the mismatches that are a problem for TM are mismatches between what you attitudes actually are and how they ought rationally to be by your own lights. This is the sort of mismatch that casts doubt on the possibility of determining what your attitudes are by reflecting on your reasons. But is there any reason for thinking that such mismatches are rare? And if mismatches are rare, does that get TM off the hook, or is there still a Matching Problem for TM? For example, you could think that no more than the possibility of recalcitrance is required to put pressure on the idea that TM can be a pathway to epistemically privileged self-knowledge.

When it comes attitudes other than belief it’s not clear how to defend the suggestion that attitude recalcitrance is rare or unusual. After all, there is nothing particularly unusual about fearing a spider you know you have no reason to fear, or wanting a martini you know you have strong and decisive reasons not to want. In such cases, what you take yourself to have reason to want or fear isn’t a sufficiently reliable guide to what you want or fear to give you knowledge, let alone epistemically privileged knowledge, of your actual wants or fears. In contrast, it is surely much more unusual for you to believe that P despite realizing that the evidence shows that not-P. This suggests a hybrid view according to which you can acquire knowledge of your own beliefs by using TM but not knowledge of your desires, fears, hopes, or other attitudes. You can know your own beliefs by using TM because what you believe is reliably related to what you take yourself to have reason to believe.

One problem with the hybrid view is that it severely limits the scope of rationalism and reintroduces the Generality Problem. I’ve said that there are two key questions for rationalism: does it explain how we can know our own beliefs, desires, and other attitudes, and does it vindicate the idea that our intentional self-knowledge is epistemically privileged? If we go for the hybrid approach then what we are saying is that rationalism can explain our knowledge of our own beliefs but not self-knowledge of other attitudes. That leaves rationalism with a large explanatory hole that will need to be filled in in some other way. In fact, rationalism is in one way even more limited in scope than the hybrid view implies. Even when it comes to explaining self-knowledge of beliefs TM has its limitations. Suppose you believe that it’s raining, then look out of the window and see that it’s not raining. You now no longer regard yourself as having a warrant to belief that it’s raining, and it’s hard to think of circumstances in which you would continue to believe that it’s raining. But in the case of beliefs which tangle with your self-conception, or beliefs you have been wedded to for years, it’s not hard to imagine how you might hang on to them despite now knowing or believing that they are unwarranted. You judge that they are unwarranted but fail to take your judgement to heart. These are cases in which knowledge of what you believe would be a form of substantial self-knowledge, and the issue is whether TM is sufficiently reliable to be a source of such self-knowledge given the genuine, and not just theoretical, possibility of a mismatch. The answer to this question depends in part on how reliable is “sufficiently reliable” but it seems a bit optimistic to suppose that you can always use TM to know what you believe about a given topic regardless of the topic.

This leads to another question: suppose that recalcitrance is sufficiently rare for TM to be capable of delivering knowledge of what you believe, at least in cases in which the subject matter is trivial. It’s then a further question whether it can deliver infallible or immediate knowledge of what you believe. The answer to this further question is almost certainly ‘no’. Whatever the subject matter, concluding that you ought rationally to believe that P doesn’t guarantee that you believe that P, so your beliefs about your own beliefs formed by using TM are still fallible even if they are unlikely to be wrong. Your self-knowledge in such cases isn’t immediate either. I’ve already commented on the way in which, when you come to know what you believe by using TM, your knowledge is mediated by the Rationality Assumption. Another way of putting pressure on the idea that TM can deliver immediate self-knowledge is this: suppose you judge you ought rationally to believe that P and know that this is your judgement (another piece of self-knowledge that needs accounting for). Judging that you ought rationally to believe that P doesn’t entail that you believe that P but it indicates that you believe that P; to put it another way, the fact that you judge that you ought to believe that P looks like pretty good evidence that you do believe that P, at least on the assumption that what you believe in trivial matters generally in line with what you think you ought rationally to believe. But your knowledge on the basis of ‘pretty good evidence’ that you believe that P can’t be regarded as immediate if immediate knowledge is knowledge that isn’t based on any evidence; in effect, you infer that you believe that P from your judgement about what you ought rationally to believe, and inferential knowledge clearly isn’t immediate knowledge.

Perhaps it’s worth adding that immediacy can be understood either epistemically or psychologically, and that rationalism is in trouble either way. If you believe that you ought to believe that P and infer that you do believe that P, then your knowledge that you believe that P (assuming that you do know this) isn’t psychologically immediate because it is the product of conscious reasoning or inference. And it isn’t epistemically immediate either because your justification for believing that you believe that P comes in part from your having justification to believe other supporting propositions, such as the proposition that you ought rationally to believe that P, and that what you believe about P is determined by what you think you ought rationally to believe.

None of this is good news for rationalism. It suggests that it fails to explain quite a lot of what it sets out to explain, and that it lacks a wholly convincing response to the Generality and Matching Problems, as well as the Substitution Problem. However, we are not quite in a position to declare ‘game over’ for rationalism. Indeed, many rationalists would argue that the real game hasn’t even started because I have so far neglected to mention a key dimension of their view, namely, the fact that we are active rather passive in relation to our own beliefs and other attitudes. For rationalism, intentional self-knowledge is a form of what might be called ‘agent’s knowledge’, and this is the key to understanding rationalism. So what I want to do next is to explain this aspect of rationalism and argue that it doesn’t get rationalism off the hook: the agential version of rationalism – what I call Activism- runs into versions of the same problems I have just been discussing.

So far I’ve been taking it that believing that P is one thing, while judging that you ought rationally to believe that P is a different matter. I’ve taken it that if you judge that you ought rationally to believe that P, that is evidence that you believe that P, and that that is how it enables you to know (though not infallibly or immediately) that you believe that P. Now consider a different view of the matter: suppose that the question arises whether you believe that P. The first thing you do in response is to deliberate, that is, consider whether there is reason to believe that P. You conclude that there is, and your conclusion is expressed in the form of a judgement: you judge that you ought rationally to believe that P. Put more simply, your deliberation leads you to judge that P. Now comes the crucial rationalist move: the relationship between judging that P and believing that P isn’t evidential; rather, your judging that P constitutes your believing that P. To judge that P, or that you ought rationally to believe that P, is to take P to be true, and that can only mean you believe that P. By deliberating about whether you ought rationally to believe that P you make up your mind about P. By acknowledging that you ought rationally to believe that P you make it the case that you believe that P, and thereby know that you believe that P.

The key to this version of rationalism is the idea that we have ‘an ability to know our minds by actively shaping their contents’ (Boyle 2009: 134). We aren’t just passive observers of our own attitudes, and it’s because we actively shape them that we have a special insight into what they are. That is why I call this version of rationalism Activism. Activism makes self-knowledge a species of what is sometimes called maker’s knowledge, the knowledge you have of what you yourself make. Idealists like Kant think that we know what the world is like because it is the mind’s construction, and now it turns out according to rationalism that we know our own attitudes because they are also the mind’s own construction. We can see what the rationalist is getting at by noting an ambiguity in the word ‘determine’. I’ve said that according to TM you determine what your attitudes are by determining what they ought rationally to be. Here, both occurrences of ‘determine’ are epistemic: in the case of belief, the view is that you come to know that you believe that P by coming to know that you ought to believe that P. However, there is also a constitutive sense of determine, according to which what you do when you determine that you believe that P is that you make it the case that you believe that P. The epistemic sense of determine is ‘determinee’ and the constitutive sense is ‘determinec’. In these terms, the Activist’s s proposal is this: by determininge that you ought rationally to believe that P you determinec that you believe that P, and can thereby determinee that you believe that P.

Does Activism give a plausible account of how we relate to our own attitudes, and does it solve the Generality, Substitution and Matching Problems in a way that allows TM to count as a source of epistemically privileged self-knowledge? On the one hand, there is something right about the idea that we are sometimes active rather than passive in relation to our attitudes. There is such a thing as reasoning yourself into believing or wanting something, and it certainly isn’t correct in these cases to say that you are a mere passive observer of your attitudes. On the other hand, as Moran concedes, ‘we’d end up with many fewer beliefs for coping with the world than we actually have if we could only acquire them through explicit reasoning or deliberation’ (2004: 458). Perceptual beliefs are a case in point; I see that there is a computer screen in front of me and believe on this basis that there is a computer screen in front of me. I know I believe there is a computer screen in front of me but I don’t know this because I have reasoned my way to this belief. Neither the belief itself nor my knowledge of it is the product of explicit deliberation. Indeed, even in the case of beliefs originally acquired by deliberation I don’t need to keep deliberating in order to know that I have them. If I have the stored belief that P, and the question arises whether I believe that P, what I need to do is not to form the belief (I already have it) but retrieve it from storage.

The question whether Activism can account for our knowledge of our stored beliefs comes to the fore in an exchange between Moran and Nishi Shah and David Velleman. Shah and Velleman argue that the question “Do I believe that P?” can either mean “Do I already believe that P?” or “Do I now believe that P?”. If the question is whether I already believe that P, the way to answer it is to pose the question whether P and see what one says, or is spontaneously inclined to answer. However, this procedure ‘requires one to refrain from any reasoning as to whether P, since that reasoning might alter the state of mind one is trying to assay’ (2005: 506). In reply, Moran objects that my stored beliefs only count as beliefs insofar as I take them to be true, and that if I relate to a stored belief as something I take to be true ‘it will be hard to see how I can see my relation to it, however spontaneous, as insulated from the engagement of my rational capacities for determining what is true or false’ (2012: 221). I am, in this sense, ‘active’ in relation to my stored beliefs. Something similar can be said about one’s perceptual beliefs; the fact that they are passively acquired doesn’t mean that they are insulated from the engagement of one’s rational capacities. Even passively acquired perceptual beliefs must be sensitive to our grasp of how they fit in to the rest of our network of beliefs. Perceptual beliefs are, to this extent, “active” but the relevant sense of “activity” is ‘the ordinary adjustment of belief to the total evidence’ (Moran 2004: 460).

What is right about this is that nothing that is recognizable as a belief can be totally insulated from the engagement of one’s rational capacities, but the question is whether the sense in which we are “active” in relation to our stored or perceptual beliefs casts any light on how we know them. Let’s suppose that I have the stored or perceptual belief that P, and that I stand prepared to revise this belief if I encounter good grounds for revising it. For example, P might be the perceptual belief that there is a computer screen in front of me, and the grounds for revising it might include the discovery that I am the subject of an experiment in which what I seem to see will be deceptive. But how is the fact that I stand prepared to revise my belief, or that I would revise it in certain circumstances, supposed to explain my knowledge that I believe that P? There are two scenarios to consider: in the first scenario, my belief that P is automatically extinguished by evidence against P regardless of whether I realize that I have until now believed or that P needs to be revised; we can imagine that all the necessary adjustments are made sub-personally rather than by me, the subject of the belief. In this case, talk of the ordinary adjustment of belief to the total evidence neither explains nor presupposes the knowledge that I believe that P. In the second scenario, I revise my belief that P because I realize that it needs to be revised, but how can I realize that a given belief of mine needs to be revised if I’m not even aware that I have it? In this scenario, my knowledge that I believe that P is presupposed but not explained. It’s not an explanation of my knowledge that I believe that P to point out that my belief is subject to revision. The explanans and explanandum just don’t connect in the right way.

The underlying problem here is that Activism operates with two very different senses of activity. In one sense, we are active in relation to our attitudes to the extent that they are formed through acts of deliberation. In another sense we are active in relation to our attitudes insofar as that they are sensitive to one’s reasoning or deliberation. Activism’s master thought is that we are able to know our own attitudes by forming or determiningc them through acts of deliberation, and the challenge is to explain how this casts any light on our self-knowledge of attitudes that aren’t formed by deliberation. Activism responds to this challenge by bringing in the idea of a readiness to deliberate and revise in the light of deliberation but it’s just not clear how this helps. Deliberating is something you actually do, and it’s one thing to say that when you form the belief that P by deliberating you can thereby determinee that you believe that P. The ‘thereby’ in this formulation needs explaining but isn’t totally opaque. What does seem opaque is the suggestion that being prepared to deliberate and revise you belief that P is what puts you in a position to know that you believe that P.

Let’s agree, then, that Activism has trouble accounting for self-knowledge of stored beliefs or beliefs not formed by deliberating. This is a version of the Generality Problem for Activism, and attitudes other than belief bring this issue even more sharply into focus; if your desire for exercise is the result of deliberating then maybe the reasoning that is responsible for the formation of the desire also enables you to know that you have it. But many desires aren’t like that; they come over you without any activity on your part but this doesn’t mean that you don’t know that you have them. Again, your knowledge in these cases can’t be accounted for along Activist lines. How can you have agential knowledge of desires that have nothing to do with your agency? Saying that desires that come over you are still responsive to reasoning or deliberation doesn’t cast any light on your ability to determinee what you desire; if you know that you want a martini but can be reasoned out of wanting one, saying that your desire is this sense responsive to reason that doesn’t explain how you know you have it.

Even when it comes to attitudes that are formed through explicit deliberation there are questions about the epistemology of Activism. Suppose that by deliberating you make it the case that you believe that P. How does making it the case that you believe that P enable you to know that you believe that P? To put it another way, when Activism says that you make it the case that you believe that P and thereby know that you believe that P, what is the force of the ‘thereby’? After all, it’s not a necessary truth that if you make it the case that P then you know that P; Boyle has the nice example of someone making it the case that his hair is on fire by standing too close to the fire but not realizing that his hair is on fire. This raises a more general question about Activism (cf. O’Brien): it sees intentional self-knowledge as a product of our rational agency with respect to our attitudes, but how is rational agency supposed to give us self-knowledge?

Here is what an Activist might say in reply: suppose I am considering the reasons in favour of thinking that it is raining, and that these reasons are convincing enough to lead me to judge that I ought rationally to believe that it is raining. To judge that I ought rationally to believe that it is raining is, in effect, to judge that it is raining. This latter judgement is the conclusion of my reflection of the reasons in favour of rain. The next question is: how do I get from judging that it is raining to knowing that I believe that it is raining? Moran writes:

I would have a right to assume that my reflection on the reasons in favour of rain provided an answer to the question of what my belief about rain is, if I could assume that what my belief here is was something determined by the conclusion of my reflection on those reasons’ (2003: 405).

The ‘conclusion of my reflection on those reasons’ is the judgement that it is raining. If I understand what it is to judge and what it is to believe then I will understand that judging that it is raining makes it the case that I believe that it is raining. So if I know that I judge that it is raining, then I ‘thereby’ know that I believe it is raining.

As we have seen, this approach doesn’t get round the Generality Problem; there are many attitudes our knowledge of which can’t be explained in this way because judging that you ought to have such attitudes doesn’t, even in the normal case, constitute your judging or believing that you have them. The Substitution Problem is also still an issue for Activism; it’s still the case that when you answer the question whether you have a given attitude by asking whether you ought rationally to have that attitude you are answering what is typically a much easier question by answering a much harder one. Intuitively, the cognitive effort required to reflect on the reasons in favour of P is often much greater than the cognitive effort needed to determine whether you believe that P; your reasons may well be opaque even if your beliefs are not. That leaves the Matching Problem, and I want to end this chapter by seeing whether Activism has a response to this problem.

 Here is a version of the Matching Problem for Activism: suppose that reflection on the reasons in favour of some proposition P leads you to judge that P. However, judging that P is not the same as believing that P. Even if you know that you judge that P this leaves it open that you still don’t believe that P, and so don’t know that you believe that P. Here is how what you judge and what you believe can come apart:

Someone can make a judgement, and for good reasons, but it not have the effects that judgements normally do – in particular, it may not result in a stored belief which has the proper influence on other judgements and on action. Someone may judge that undergraduate degrees from countries other than her own are of an equal standard to her own, and excellent reasons may be operative in her assertions to that effect. All the same, it may be quite clear, in decisions she makes on hiring, or in making recommendations, that she does not really have this belief at all (Peacocke 1998: 90).

Obviously, if judging that P normally results in the stored belief that P you can still infer from the fact that you judge that P that you believe that P but this clearly isn’t what Activists have in mind when they talk about a person’s reasoned judgements constituting his beliefs, and thereby providing direct access to them.

Not everyone finds Peacocke’s example convincing or even coherent. One thought I have already mentioned is that to judge that P is to take P to be true, and that you can’t take P to be true without believing that P. So there can be no gap between believing that P and judging that P. But then what is going on in Peacocke’s example? Here is one possibility: you recognize that you have excellent reasons to judge that undergraduate degrees from other countries are as good as undergraduate degrees from your own country, and that this is what you ought rationally to judge, but you fail to make the judgement. You might say the words “undergraduate degrees from other countries are of an equal standard” but you don’t actually take this to be true, and therefore neither judge nor believe that undergraduate degrees from other countries are of an equal standard; the very considerations which show that you don’t really believe that undergraduate degrees from other countries are of an equal standard also show that you don’t judge that undergraduate degrees from other countries are of an equal standard.

This doesn’t really help because it closes one gap while opening up another. It closes the gap between judging that P and believing that P while opening a gap between judging that you ought rationally to judge or believe that P and actually judging or believing that P. This is the mismatch I’ve been talking about for most of this chapter, and Activism doesn’t show that there can’t be this kind of mismatch. Even if judging that P amounts to believing that P, at least at the moment that you judge that P, Activism still owes us an account of how you know your own conclusions or judgements. The point is that just because you take yourself to have good reason to judge that P it does not follow that you actually judge that P. As O’Brien points out, ‘a subject can have warrant for thinking that he has judged that P when she has in fact only concluded that she has good reason to judge that P and then drawn back in a fit of risk aversion’ (2003: 379). Obviously, there is still the option of arguing that what you take yourself to have reason to judge is a good enough guide to what you actually judge (and believe) to give you mediate knowledge of what you actually judge (and believe), but this is no different from the position that Simple Rationalism ended up in. There is certainly no indication that Activism has intentional self-knowledge come out as infallible or immediate.

It might be that what is really going on here is that Activism’s conception of belief is different from mine, and that we are to some extent talking at cross purposes. My view of belief is broadly dispositional. Whether you actually believe that P depends on whether you are disposed to think that P when the question arises, act as if P is true, and use P as a premise in reasoning. On this conception of belief, it’s easy to think of the relationship between what you judge, or think you ought to judge, and what you believe as evidential: judging that P is not guaranteed to result in the stored dispositional belief that P, even if this is what normally happens. Judging, unlike believing, is a mental action, and the effects of this action may vary. But Activism seems to think of at least some beliefs as occurrences, the idea being that when you judge that P you occurrently believe that P. Occurrently believing that P might not result in your believing that P in the dispositional sense – maybe this is what happens in Peacocke’s example- but if you know that you judge that P then at least you thereby know immediately that at that moment you occurrently believe that P.

Setting aside any doubts one might have the notion of an occurrent belief, this move is still of limited help to Activism. It still leaves Activism with the problem of accounting for a person’s knowledge that he judges that P, and doesn’t close the gap between taking yourself to be warranted in judging that P and actually judging or occurrently believing that P. There is also the worry that, at best, Activism is only equipped to account for immediate knowledge of our own occurrent attitudes. When it comes to dispositional beliefs and other attitudes, the Activist is in the same boat as the Simple Rationalist: if A is a recalcitrant (dispositional) attitude, then your judgement as to whether you ought or ought not to have A will cast little light on whether you have A. In other cases, your judgements about whether you ought to have A might tell you whether you have A, but your knowledge will still be mediated by the Rationality Assumption and your knowledge of what you ought rationally to judge. Even for Activism, unless you are homo philosophicus judging that you ought rationally to believe that P can only be an indication that you judge or believe that P.

Does this matter? That depends on why rationalism is so keen on the notion that we have epistemically privileged access to our own attitudes. There isn’t much to be said for the idea that self-knowledge of attitudes is infallible: you can think you believe that the present government will be re-elected but your lack of surprise when the election goes the other way suggests otherwise. When the waiter asks you what you would like to drink what you think you want is a gin martini but when you take your first sip you realize that what you wanted all along was a vodka martini. These are all straightforward cases in which you are fallible about your attitudes, and it’s not a requirement on a theory of self-knowledge that it has our knowledge of our own attitudes come out as infallible. An account of self-knowledge that has this consequence has got to be wrong. Maybe in the cases I have described there is some kind of (defeasible) presumption that you aren’t wrong about your attitudes, but whether there is such a presumption depends on the nature and content of the attitude.

For many rationalists it isn’t the infallibility but rather the immediacy of self-knowledge that is the starting-point, but is there actually a presumption that self-knowledge is immediate? On one reading, immediate self-knowledge is not based on behavioural evidence. On another reading, it is knowledge that is not based on any evidence. If judging that you ought rationally to believe that P doesn’t count as a form of “behaviour”, and you know on the basis of this judgement that you believe that P, then your knowledge in this case isn’t based on behavioural evidence but it isn’t true that it’s based on no evidence. If judging is a mental act then your “evidence” is psychological rather than behavioural. To the extent that you have “privileged access” to your attitude it is because the evidence at your disposal is excellent and readily available.

In other cases, you might have to rely on behavioural evidence. For example, you might need to rely on behavioural evidence broadly construed to know whether you really believe in God or that your spouse is faithful. These are among your “deepest” attitudes, and it’s not counter-intuitive to suppose that you need evidence, including behavioural evidence, to know such attitudes. Indeed, when it comes to attitudes other than belief, even relatively superficial self-knowledge might need to draw on behavioural evidence; if you order a vodka martini and the question arises how you know you want a vodka martini it wouldn’t be beside the point to point out that you’ve just ordered one. You might know independently of placing your order that you want a vodka martini, but there again, you might not. Sometimes the best guide to what you want is what you chose; it would be nice to think that you chose a vodka martini because you already knew that that is what you wanted but it isn’t always like that.

In a weird way these observations are actually helpful to rationalism. Rationalism sets itself the target of explaining how it’s possible to know our own attitudes without relying on behavioural or other evidence. It succeeds in explaining how, by using TM, you are able in at least some cases to know your own attitudes without relying on behavioural evidence but not without relying on any evidence. What I’m now questioning is whether it was ever sensible to assume that intentional self-knowledge doesn’t require any evidence. Once we give up on this idea, and are prepared to think of knowledge of our own (dispositional) attitudes as evidence-based knowledge, then we will need to tell a story and the kinds of evidence that are relevant. One kind of relevant evidence is, as rationalism suggests, judgemental. If you are lucky (or unlucky) enough to be homo philosophicus, and judge that you ought rationally to believe that P, that is excellent evidence that you do believe that P. If you are homo sapiens, and you judge that you ought rationally to believe that P, that is less than excellent, but still potentially good evidence that you believe that P, depending on the nature of the belief. If you are homo sapiens and judge that you ought rationally to want that P that is not very good evidence that you want that P. The evidence that you want that P might have to come from a range of other sources, including what you say and do.

We now have at least the makings of an answer to what I referred to in chapter 1 as the sources question: what are the sources of self-knowledge for humans? The first thing to say about this is that it depends on the kind of self-knowledge that is at issue. When it comes to knowledge of our own attitudes it’s tempting to look for a magic bullet, and single source that will account for all our intentional self-knowledge. Unfortunately, there just is no magic bullet; there are multiple sources of intentional self-knowledge, depending on the type and content of the attitude known. Among the sources of some intentional self-knowledge is TM, but TM only accounts for a narrow range of intentional self-knowledge, and even in the best case the knowledge it accounts for doesn’t come out as based on no evidence. So the natural next step would be to look for other sources of intentional self-knowledge which help account for the varieties of intentional self-knowledge that rationalism doesn’t handle well. I will come back to this ‘multiple pathways’ approach to self-knowledge in chapters 11 and 12. But before then, there is another piece of business that needs to be taken care of.

One thing I didn’t say, and probably should have said, when I introduced rationalism via Evans’ idea of gaining self-knowledge by ‘looking outwards’ was that this model of self-knowledge didn’t just come out of thin air. It arose as a response to the historically influential idea that the way to explain self-knowledge is to appeal to the idea of looking inwards. Locke, Kant, and many others operated with the idea that humans have a quasi-perceptual faculty of ‘inner’ or ‘internal’ sense, and that we know our own states of mind by exercising this faculty. Evans characterizes this as the idea that knowledge of our own mental properties ‘always involves an inward glance at the states and doings of something to which only the person himself has access’ (1982: 225). This is the model to which the ‘looking outwards’ approach is opposed, so before going any further it would be worth pausing to ask whether the ‘inward glance’ model has anything going for it. This is the question for the next chapter.