6

PSYCHOLOGICAL RATIONALISM

In the last chapter I introduced Psychological Rationalism as a view about the extent of the Disparity. When I talk about the Disparity that is shorthand for the various respect in which homo sapiens differ from homo philosophicus. I’ve been arguing that the Disparity is extensive, at least extensive enough to put pressure on Rationalism about self-knowledge. If we are as different from homo philosophicus as I have been maintaining then it’s not usual for our propositional attitudes not to be as they ought rationally to be, and that makes it hard for us to determine what our attitudes are by determining how they ought rationally to be. In reply, Psychological Rationalists question the grounds for positing an extensive Disparity, and offer their own grounds for thinking that there can’t be an extensive Disparity. As far as Psychological Rationalism is concerned we are sufficiently similar to homo philosophicus for Rationalism about self-knowledge to be a viable account of how we know our own beliefs, desires, and so on. I will refer to this as Psychological Rationalism’s Similarity Thesis.

Psychological Rationalism has antecedents in the history of philosophy. In his book The Mind of God and the Works of Man, Edward Craig discusses how what he calls ‘The Similarity Thesis’ influenced many of the great, dead philosophers, including Berkeley, Spinoza and Leibniz. Craig’s Similarity Thesis emphasises the closeness between God and the individual human being; it says that ‘man was made in God’s image’ (1987: 14). No doubt there are countless differences between man and God but the nature of their reason there is one point where ‘the human mind and the mind of God naturally coincide’ (Craig 1987: 46). Since God is presumably a model epistemic agent whose reasoning is flawless and whose beliefs are as they ought rationally to be, the claim that the humans and God naturally coincide in respect of their reason implies that we generally reason well and that our beliefs are generally as they ought rationally to be. This suggests that the Similarity Thesis I am attributing to Psychological Rationalism is in the same ballpark, and has some of the same implications, as the thesis that Craig attributes to a bunch of 17th century rationalists.

With Craig’s discussion in mind, the model epistemic citizen I have been referring to as homo philosophicus can be viewed as a representation of some of the epistemic virtues traditionally attributed to God. By the same token, placing emphasis on the Disparity can be viewed as a way of emphasizing the discontinuities between man and God as traditionally conceived. Indeed, these discontinuities might be even greater than is implied by talk of a ‘Disparity’; for God to believe what he ought to believe it would have to be assumed that God has beliefs, and that isn’t an uncontroversial assumption (see Alston 1986). Presumably, there is also a sense in which God doesn’t really reason; he sees instantly what follows from what and so doesn’t have to move from premises to conclusions. God only reasons if genuine reasoning can be instantaneous. We can avoid such complications by reframing the Similarity Thesis in terms of the relationship between homo sapiens and homo philosophicus rather than the relationship between man and God.

Why would anyone believe the Similarity Thesis in my sense? You might think that it is an empirical question how close we are to the ideal of homo philosophicus. If there is empirical evidence that much of our reasoning isn’t critical, that we are self-ignorant and biased to believe, and that our attitudes tend to survive evidential discrediting then there is, to that extent, empirical evidence against the Similarity Thesis. By the same token, it would seem to follow that the best way to defend the Similarity Thesis and attenuate the Disparity is to produce evidence that we do by and large reason critically, that we aren’t generally self-ignorant or biased to believe, and that our attitudes do not on the whole survive evidential discrediting and aren’t recalcitrant. Such empirical evidence against the Disparity would also amount to empirical grounds for thinking that our attitudes are more or less as they ought rationally to be.

In fact, this isn’t how Psychological Rationalists argue. It’s not that they do not regard empirical arguments for the Disparity as questionable in their own terms but that they don’t regard the Similarity Thesis as one that should be defended on piecemeal empirical grounds. The alternative is to defend the thesis on “transcendental” grounds. What I mean by a transcendental defence of the Similarity Thesis is one that focuses on the question: what are the necessary conditions for a subject to have propositional attitudes at all? Suppose it turns out that for a subject to have attitudes at all the relations among his attitudes, perceptions, and actions must be by and large rational. This would be a positive argument for the Similarity Thesis, since homo philosophicus is by stipulation a being whose attitudes are as they ought rationally to be. Indeed, the result of this transcendental argument is not just that our attitudes must approximate to being as they rationally ought to be but that any being that has attitudes must in this respect be like homo philosophicus. As for the specific elements of the Disparity, there are now two ways of dealing with these other than on piecemeal grounds: one is to argue that however widespread phenomena like belief-perseverance and self-ignorance may be they do not show that relations among our attitudes, perceptions and actions aren’t by and large rational. The other would be to view the transcendental argument for the Similarity Thesis as an argument against the possibility of such phenomena being widespread. I’ll come back to this.

Before looking in detail at how a transcendental argument for the Similarity Thesis might go I’d like to say something about my use of the term “transcendental”. The notion of a transcendental argument is associated with Kant, who saw such arguments as ‘a priori’ rather than empirical. A transcendental argument tries to establish the truth of some proposition P by arguing that the truth of P is necessary for knowledge, thought, or experience. As far as Kant was concerned, such necessary conditions can only be established non-empirically, by means of a priori philosophical reflection, but there is no need to follow Kant in this respect. You could think that there are highly general necessary conditions of thought, knowledge or experience that can only be established by philosophical reflection but that what we appeal to when we engage in such reflection are high level empirical considerations. In addition, claims about what is and isn’t necessary for thought, knowledge, or experience are certainly liable to empirical refutation. Seen in this way, so-called “transcendental” arguments for the Similarity doesn’t have to be a priori though they can be; it’s one thing to say that a necessary condition for a subject to have attitudes is that his attitudes are largely rational, and another to say that this isn’t an “empirical” truth, whatever that means.

The next question is: is it true that having attitudes that are mostly rational, mostly as they ought rationally to be, is a necessary condition for one to have attitudes at all? There is a discussion of this issue in Dennett’s paper ‘Three Kinds of Intentional Psychology’. Dennett argues that we approach each other as what he calls ‘intentional systems’, that is, as entities whose behaviour can be predicted by the method of attributing beliefs, desires, and rational acumen according to the following principles:

1. ‘A system’s beliefs are those it ought to have, given its perceptual capacities, its epistemic needs, and its biography’ (1987: 49).
2. ‘A system’s desires are those it ought to have, given its biological needs and the most practicable means of satisfying them’ (ibid.).
3. ‘A system’s behaviour will consist of those acts it would be rational for an agent with those beliefs and desires to perform’ (ibid.).

On this account, the notion of a propositional attitude is fundamentally an explanatory notion. We ascribe beliefs and desires to give reason-giving explanations of actions, and beliefs and desires can themselves be given rational explanations; we make it intelligible that S believes that P by explaining why S ought to believe that P in these circumstances. This is presumably also what McDowell is getting at when he says that concepts of propositional attitudes have their ‘proper home in explanations of a special sort: explanations in which things are made intelligible by being revealed to be, or to approximate to being, as they rationally ought to be’ (1998: 328).

What Dennett means by “ought to have” in 1 and 2 is “would have if it were ideally ensconced in its ensconced in its environmental niche”. This gives us ‘the notion of an ideal epistemic and conative operator or agent’ who recognizes all the dangers and vicissitudes in its environment and desires all the benefits relative to its needs. An ‘ideal epistemic operator’ sounds like a good description of homo philosophicus but not of homo sapiens, ‘for surely we are not all that rational’ (1987: 50). Nevertheless, Dennett insists, we treat each other as if we are rational agents, and that this ‘works very well because we are pretty rational’; while we are not ideal epistemic and conative agents we ‘approximate to the ideal version of ourselves exploited to yield predictions’ (1987: 51). This is Dennett’s version of the Similarity Thesis. Folk psychological attributions of propositional attitudes predict what we will believe, desire, and do ‘by determining what we ought to believe, desire, and do’ (1987: 52).

The original question was: is it a necessary condition for having attitudes at all that the relations among our attitudes, perceptions, and actions are by and large rational? But what Dennett argues is that we have to assume that our attitudes are as they rationally ought to be when we attribute attitudes to one another. Even if his argument works it doesn’t look like an answer to the original question; just because we have to assume that are attitudes are a certain way as a condition for attributing attitudes it doesn’t follow that we couldn’t so much as have attitudes unless they are that way. Indeed, Dennett explicitly describes our rational agenthood as a ‘myth’, and the fact that this myth ‘structures and organizes our attributions of belief and desire to each other’ (1987: 52) doesn’t make it true that we are rational. The claim that we are ‘pretty rational’ can only be established on empirical grounds, but then there is no longer anything “transcendental” about this defence of the Similarity Thesis.

What this objection fails to take to take into account is Dennett’s ‘interpretationism’. This is roughly the view that what makes it true that a subject S believes or desires that P is that S can be interpreted as believing or desiring that P on the basis of what he says and does. So there isn’t a gap between what is necessary for S to have propositional attitudes and what is necessary for S to be interpreted as having propositional attitudes, that is, what is necessary for such attitudes to be attributed to S on the basis of his verbal and non-verbal behaviour. It is at this point that rationality comes into the picture; the sense in which, in attributing beliefs and desires to S, we have to assume that his attitudes are by and large as they ought to be is that the ideal of rationality has a ‘constitutive role’ in interpretation. Here is Bill Child’s lucid summary of all this:

When we interpret someone, we explain her actions in terms of her reasons. So the idea of a reason-giving explanation is central to the interpretationist conception of the mental. Internal to that form of explanation, and thus to the interpretationist conception, is the notion of rationality; the ideal of rationality has a constitutive role in propositional attitude psychology. To say that is to say (amongst other things) that if a subject has attitudes at all, the relations amongst her attitudes must by and large be rational. No actual individual is perfectly rational; so all sorts of local irrationality in thought and action are intelligible. But what is not intelligible is that a subject might have a set of attitudes that were absolutely irrational…. with an individual like that, the idea of explaining actions in terms of reasons could have no application (1994: 8).

On this account, the assumption that we are ‘pretty rational’ has a transcendental justification; our being pretty rational is a necessary condition of interpretability, and being interpretable as having beliefs, desires and other attitudes is what it is for one to have such attitudes. If we have beliefs and desires then by and large they must be as they ought to be.

I’m going to call this argument for the Similarity Thesis the ‘argument from above’ since it relies on highly abstract claims about the nature and explanatory role of propositional attitudes to show that insofar as we have beliefs and desires we must approximate to the ideal of homo philosophicus. The contrast is with what might be called an ‘argument from below’ for the Similarity Thesis, that is, one that defends this thesis on piecemeal empirical grounds. As I’ve already indicated, the argument from above needn’t be conceived of as non-empirical since the interpretationist account of the mental could in principle be regarded as a high level but nevertheless ultimately empirical theory of mind. However, there is also no denying that it is more natural to interpret the argument from above as relying on a series of ‘a priori’ claims about the nature of mental states and the constitutive ideal of rationality.

Whatever the status of the argument from above, the important question is: is it any good? Here are three reasons why you might be sceptical about this argument:

1. You might have a problem with interpretationism.
2. You might question the supposed ‘constitutive role’ of the ideal of rationality in interpretation. The aim of interpretation is to make people intelligible and predict their behaviour but doing these things does not require the supposition that their attitudes are as they rationally ought to be.
3. You might accept the argument from above, including what it says about the constitutive role of rationality, but think that the argument only looks plausible because it is vacuous; the Disparity consists of a bunch of highly specific respects in which humans typically fall short of homo philosophicus, and the very abstract guarantee of rationality supplied by the argument from above does nothing to attenuate the Disparity; for all that this argument shows the various phenomena which make up the Disparity could still be widespread among normal humans.

I’m not going to get into a discussion about (a), beyond pointing out that it is a fact that the considerations in support of the Similarity Thesis that are put forward by writers like Dennett, Davidson and McDowell are interpretationist considerations, and that interpretationism isn’t uncontroversial; it certainly wouldn’t and shouldn’t be accepted by anyone who thinks that propositional attitudes are internal states. I’m going to concentrate instead on (b) and (c), both of which raise serious questions about the argument from above.

Starting with (b), let’s go back to the case of Steve from Kahneman’s Thinking, Fast and Slow. You will recall that Steve has been described as shy and withdrawn, as invariably helpful but with little interest in people or the world of reality. This is the description you have been given, and I’m trying to decide whether you believe that he is more likely to be a farmer or a librarian. Let’s suppose also that you know that there are many more farmers than librarians. The sense in which you “know” this is that if I were to ask you to compare the number of farmers with the number of librarians in the country you would say that there are many more farmers. If you know that there are many more farmers then you ought rationally to believe that Steve is more likely to be a farmer but do I think that this is what you are more likely to believe? Not at all. Even without detailed knowledge of Kahneman’s discussion of the role of the representativeness heuristic in human thinking I might suspect that you are in fact more likely to believe that Steve is likely to be a librarian. You are more likely to believe this because Steve’s personality is that of a stereotypical librarian rather than a stereotypical farmer.

In this example the belief I attribute to you is that Steve is more likely to be a farmer, and this is the belief I attribute to you because I know enough about human psychology to know that in making judgements of probability people are prone to ignoring highly relevant statistical considerations; indeed, it’s not just that people ignore such considerations but that large sections of the population seem not to have a grasp of even the most elementary statistical principles. Unless I know that you are a particularly careful thinker with a grasp of statistics it’s a fair bet that my description of Steve will have led you to form the belief that Steve is more likely to be a librarian. So that is belief I ascribe to you even though it is not the belief you ought rationally to have. I don’t determine what you believe by determining what you ought to believe, and the basis of my ascription isn’t the ‘constitutive ideal of rationality’ but an explicit or implicit grasp of the power of the representativeness heuristic. BAT AND BALL is no different. You ought to think that the ball costs 5 cents but I don’t suppose that this is what you do think. Again, there are two points: what you believe in this case isn’t what you ought rationally to believe, and the myth of your rational agenthood isn’t what structures and organizes my thinking about what you believe. The thing that does that is my sense of how fast thinking humans are likely to approach problems like BAT AND BALL.

Other non-rational factors that need to be taken into account when attributing beliefs to others include the bias to believe, the attractions of conspiracy theories, and the prevalence of belief-perseverance and attitude recalcitrance. The evidence supports the view that the 9/11 attacks were carried out by al Qaeda but one would have to be quite optimistic to think that what people believe about the 9/11 attacks is what they ought rationally to believe. In a case like this, the socio-political context of the attribution seems far more relevant than any considerations of rationality. Or take Harman’s Karen example. Karen believes on the basis of her aptitude test scores that she has an aptitude for science and music but not for history or philosophy. Then she is told that she had been given someone else’s test results. What would Karen now think? When Harman concludes that ‘Karen would almost certainly keep her new beliefs’ (1986: 35) he isn’t appealing to the principle that what she would believe is what she ought rationally to believe. The basis on which he predicts Karen’s belief is the prevalence of belief-perseverance. It doesn’t matter what Karen ought rationally to believe because it isn’t the thought of what she ought rationally to believe that is doing the explanatory work.

In none of these cases do we have any difficulty making the believers and their beliefs intelligible other than on the basis of considerations of rationality. We don’t even try to make it intelligible that S believes that P by explaining why S ought to believe that P. And when it comes to attitudes other than beliefs it’s even clearer that intelligibility does not depend on rationality. You have no reason to fear the spider in your bathtub but it’s intelligible that you fear the spider. My grounds for judging that you fear the spider have little to do with thinking that you ought rationally to fear it. As for the suggestion that a system’s desires are the ones it ought to have given its biological needs and most practicable means of satisfying them, this might be true of homo philosophicus but humans are a different matter. A desire to smoke cigarettes is one that many humans have but ought not to have given their biological needs, while a desire for exercise is one that many of us lack even though we ought to have it. In his discussion Dennett tends to oscillate between representing the ‘ought’ in ‘S ought to desire P’ as having to do with what is rational and as a matter of what promotes survival. These aren’t the same thing (see Child 1994: 51) but either way it’s implausible that we can make sense of each other by thinking about what our attitudes ‘ought’ to be.

It might seem that the way for the interpretationist to deal with such cases is to point out that they are examples of the ‘local irrationality’ the existence of which interpretationism never sought to deny. Just because we sometimes interpret people as having attitudes other than the ones they ought rationally to have it doesn’t follow that their attitudes aren’t by and large as they rationally ought to be; it’s just unintelligible that someone who has beliefs and desires is absolutely irrational. This is what I mean when I describe the argument from above as a ‘damage limitation exercise’ or, more colourfully, as a transcendental damage limitation exercise. It allows that our beliefs sometimes persevere despite evidential discrediting, that our propositional attitudes are sometimes recalcitrant, and so on, but the suggestion is that such phenomena can’t be widespread. They may amount to respects in we are different from homo philosophicus but they don’t falsify the Psychological Rationalism’s Similarity Thesis because the limitations on the extent to which we can be different from homo philosophicus still allow us to determine what our attitudes are by determining what they ought rationally to be; the Disparity, such as it is, isn’t a problem for rationalism about self-knowledge.

There are quite a few problems with arguing this way. Here is the first problem: let’s agree that the argument from above succeeds in limiting the extent of the Disparity. But since there is no question of the argument completely eliminating the Disparity we then face the question: how much of a Disparity can rationalism about self-knowledge live with? At least on the face of it, even a small Disparity is going to be a problem for rationalism. Given that it is at least sometimes the case that your attitude toward something isn’t as it rationally ought to be, the method of determining what your attitude is by determining what it ought rationally to be will sometimes lead you astray. For example, you will sometimes conclude that you want something that you don’t actually want because you judge it is what you ought to want. But how can use of an unreliable method give you knowledge of your own attitudes?

It’s hard to assess this objection without getting into a wide-ranging epistemological discussion that is well beyond the scope of this chapter. The obvious thing to say is that just because use of a particular method for forming beliefs about our attitudes sometimes leads us astray it clearly doesn’t follow it’s an unreliable method or not reliable enough to be a source of knowledge. The reliability required for knowledge isn’t perfect reliability so you can still discover what your attitudes are by determining what they ought to be as long as using this method will generally give you the right answer. It will generally give you the right answer as long as your attitudes are generally as they ought to be, which is precisely what the argument from above claims. So the real issue is whether this argument is capable of limiting the scale of the Disparity in the way that it claims. Just how similar to homo philosophicus do we have to be if we are to be interpretable as having propositional attitudes?

The problem the argument from above faces at this point is perfectly illustrated by the passage from Bill Child quoted above. On Child’s interpretation of interpretationism, what it rules out is that ‘a subject might have a set of attitudes which were absolutely irrational, for none of which she has any reasons at all, and which it was impossible to relate intelligibly to her action’ (1994: 8). The interpretationist may be right that this would be unintelligible but ruling out the possibility of an absolutely irrational subject of propositional attitudes doesn’t do much to limit the scale of the Disparity since you can have many attitudes that aren’t as they ought rationally to be without being ‘absolutely irrational’. When Oliver thinks that the collapse of the twin towers on 9/11 was caused by a controlled demolition, or when Karen continues to insist that she has an aptitude for science and music, it’s not that they don’t have reasons for what they believe. They have their reasons, in the light of which their attitudes are intelligible. To be sure, their reasons aren’t very good reasons but that doesn’t make Karen or Oliver absolutely irrational. Their attitudes are not as they rationally ought to be, and they may open to rational criticism, but these aren’t examples of the kind of extreme irrationality that the argument from above rules out. Saying that a person’s attitudes can’t be absolutely irrational is one thing; saying that they must be by and large as they rationally ought to be is another.

To put it another way, the argument from above establishes something much weaker than Psychological Rationalism needs. Psychological Rationalism says that we approximate to ideal epistemic agents, that is, to homo philosophicus, but the argument from above only shows that if we have propositional attitudes we can’t be totally irrational. But there’s a very big difference between not being totally irrational and approximating to homo philosophicus. Ruling out extreme irrationality leaves open the possibility of an extensive Disparity between homo sapiens and homo philosophicus, certainly extensive enough to make it impossible for us to determine with any reliability what our attitudes are by determining what they ought to be. If this is right then the transcendental guarantee the argument from above provides is too weak to be of much use to Psychological Rationalism.

In any case, it’s not just a question of how much epistemic malpractice is consistent with the argument from above. This way of putting things makes it sound as though the real problem with the argument from above is quantitative: what Psychological Rationalism can tolerate is only a small Disparity but the argument from above allows the Disparity to be big. This way of putting things is fine as far as it goes but it misses a deeper point about the basis on which we ascribe propositional attitudes to each other. The deeper point concerns the role of the ideal of rationality in attitude ascriptions. The impression you get from Dennett is that when we interpret other people we make the default assumption that their attitudes are by and large as they ought to be, and that this assumption is what enables us to predict the behaviour and attitudes of other people. However, it’s only any use assuming that people generally have the attitudes they ought to have if it’s clear what attitudes they ought to have, and that’s just the problem: the notion of a person’s attitudes being as they rationally ought to be is much more opaque than the argument from above assumes.

Here is a simply illustration of the point: you believe P and you believe that if P then Q. Should you believe that Q? Not necessarily. If you have independent evidence against Q then maybe you should revise your belief that P, or your belief that if P then Q. This is a point I made in chapter 1: in practice, the question ‘Does S believe that P?’ is often much easier to answer than the question ‘Ought S to believe that P?’, just as the question ‘Does S fear that P?’ is often much easier to answer than the question ‘Should S fear that P?’. This makes it implausible that the way to answer the first of these questions is to answer the second, since we would then be answering an easier question by answering a harder question. In addition, if we can answer the question ‘Does S believe that P?’ prior to answering the question ‘Should S believe that P?’ then we must have independent means of answering the former question, as indeed we do in many cases. This suggests that what is true in cases like KAREN and BAT AND BALL is true more generally: it’s inefficient to predict a person’s attitudes on the basis of what their attitudes ought to be because we are often so unclear what their attitude ought to be, and it’s also ineffective to predict a person’s attitudes on this basis because what people believe, want, fear etc. is more often than not influenced by a wide range of non-rational psychological and contextual factors that are in danger of being ignored if the focus on is the ‘constitutive ideal of rationality’. Once again, the inescapable conclusion is that the argument from above does very little in practice to attenuate the Disparity.

This discussion leads naturally to (c). Like (b), the effect of (c) is to call into question the argument from above’s ability to limit the extent of the Disparity, but (c) does this in a different way from (b). Whereas (b) casts doubt on the role of the myth or ideal of rationality in structuring and organizing our attributions of belief and desire to each other, (c) makes the point that even if interpretation is governed by the ideal of rationality the net effect on the Disparity is negligible. How can that be? Because the various principles of rationality which interpretationists like Dennett propose are so lacking in substance as to be compatible with most of the phenomena which make up the Disparity. If this is right, then (b) and (c) are two horns of a dilemma for the argument from above: the first horn says that if the principles of rationality are substantial enough to rule out the Disparity then it’s implausible that they structure and organize our attributions of belief and desire to each other. The second horn says that if Dennett’s principles structure and organize our attributions of attitudes to each other then they can’t be substantial enough to rule out the Disparity. Either way, the argument from above provides no effective transcendental guarantee that insofar as we have beliefs and desires we must be similar to homo philosophicus.

The point that interpretationism’s principles of rationality aren’t substantial enough to attenuate the Disparity or vindicate Psychological Rationalism can be illustrated by reference to the principle that ‘a system’s beliefs are those it ought to have, given its perceptual capacities, its epistemic needs, and its biography’ (Dennett 1987: 49). Now consider Karen. Is her belief that she has an aptitude for science and music one that she ‘ought’ to have after she has been told about the mix up with the test results? I have been arguing that it’s hard to know what the subject should or shouldn’t believe in cases like KAREN but the other side of the coin is that there then isn’t a clear-cut cut case for saying that her belief is one that she oughtn’t to have given her capacities and biography. Among her capacities are her capacity to keep track of her justifications for he beliefs, but we know that this capacity is bound to be limited given the need for her to avoid too much mental clutter. Given that she might have lost track of her original justification for believing that she has an aptitude for science and music it could be argued that her belief is not in breach of Dennett’s principle even though belief-perseverance after evidential discrediting is not something that homo philosophicus would ever be guilty of. In interpreting Karen as believing that she has an aptitude for science and music but not for history or philosophy you aren’t interpreting her as believing anything other than what she ought to believe given further background assumptions about her.

Even in the variation on KAREN in which she has kept track of her justifications but still believes that she has an aptitude for science and music because she finds it hard to get rid of this belief, it’s not absolutely clear that Dennett should find this objectionable. Who is to say that a certain degree of attitude recalcitrance might not promote survival and be ‘rational’ at least to the extent that it isn’t always worth the mental effort to get rid of one’s entrenched beliefs? This doesn’t mean, of course, that Karen is no different from homo philosophicus , or that there is no Disparity if her beliefs are those she ‘ought to have’ in Dennett’s sense. Her beliefs can be those she ought to have in this sense even if the way she operates is different from the way that homo philosophicus would operate; your beliefs can be as they ought to be in Dennett’s sense even if you aren’t a model epistemic citizen.

Although it’s easy to see the point of arguing in this way, my own view is that (b) is a more effective response to the argument from above than (c). What the latter does is to try to reconcile the Disparity with the role of the ideal of rationality in interpretation but it goes too far when it implies that the myth of our rational agenthood rules very little out when it comes to the differences between us and homo philosophicus. Although it’s true that a system whose beliefs are those it ought to have can also be one whose beliefs sometimes persevere despite evidential discrediting, it’s not true that principles like Dennett’s rule nothing out. It’s hard to maintain that your beliefs are as they ought to be in BAT AND BALL, or that Oliver’s beliefs about 9/11 are in good order. It’s even clearer with other attitudes, such as self-destructive desires or irrational fears, that something is seriously amiss from the standpoint of rationality. The thing to say about these cases is not that they are ones in which the subject’s attitudes are as they ought to be, either from the standpoint of rationality or the standpoint of survival. The thing to say is that the fact that something is seriously amiss with such attitudes doesn’t make it impossible or especially difficult to interpret ordinary humans as having them. However, this takes us back to (b), and to the suggestion that the argument from above doesn’t limit the extent of the Disparity because many of our attitudes do not conform to demanding principles of rationality.

So much for the argument from above: it tries to vindicate Psychological Rationalism by means of a transcendental argument against the possibility of a substantial Disparity but the proposed argument is no good. What it shows is that there are limits to how irrational a being with propositional attitudes can be, but these limits do not in any sense vindicate the Similarity Thesis. That leaves the argument from below. When I introduced this argument I described it as an argument for the Similarity Thesis, but this is misleading. It’s not so much a positive argument for this thesis as an attempt to deflect arguments against Psychological Rationalism. The idea is really very simple: the main line of attack on the Similarity Thesis has been to point out the various ways in which humans fail to approximate to the ideal of homo philosophicus. This is an empirical argument against the Similarity Thesis and is only as good as the empirical evidence for a substantial Disparity. This evidence takes the form of experiments that reveal all manner of human epistemic malpractice, but the issue is whether the various failings that emerge in contrived experimental situations tell us anything about ordinary human thinking and reasoning. The argument from below challenges the relevance of the supposed experimental evidence for the Disparity; it says that this evidence does not generalize, and doesn’t show that that there is a substantial Disparity in real life.

I mentioned an argument along these lines chapter 5. Just as behavioural economists argue that laboratory findings of economic irrationality fail to generalize to real markets, so Dennett objects on roughly similar grounds to empirical evidence from psychology against the Similarity Thesis:

How rational are we? Research in social and cognitive psychology (e.g. , Tversky and Kahneman 1974; Nisbett and Ross 1978) suggests we are only minimally rational, appallingly ready to leap to conclusions or be swayed by logically irrelevant features of situations, but this jaundiced view is an illusion engendered by the fact that these psychologists are deliberately trying to produce situations that provoke irrational responses – inducing pathology in a system by putting strain on it- and succeeding, being good psychologists…. A more optimistic impression of our rationality is engendered by a review of the difficulties encountered in artificial intelligence research. Even the most sophisticated AI programs stumble blindly into misinterpretations and misunderstandings that even small children reliably evade without a second thought (…). From this vantage point we seem marvellously rational (1987: 52).

This, in essence, is the argument from below: not a grandiose transcendental argument from on high for the Similarity Thesis but a more prosaic empirical argument against the Disparity or, more accurately, an empirical argument against empirical arguments for the Disparity.

Is the argument from below any good? Talk of ‘laboratory finds’ and of psychologists trying to engineer irrational responses is well wide of the mark. Consider BAT AND BALL again. This problem figures in Shane Frederick’s Cognitive Reflection Test of one type of cognitive ability. Frederick reports that CRT was administered to 3,428 people in 35 separate studies. Most respondents were undergraduates who were paid $8 to complete a 45-minute questionnaire. They were told only: “Below are several problems that vary in difficulty. Try to answer as many as you can” (2005: 28). It’s true that BAT AND BALL is specifically designed to produce an answer that is intuitive and wrong, and it makes a difference that the example is BAT AND BALL rather than, say, BAGEL AND BANANA: a banana and bagel cost 37 cents. The banana costs 13 cents more than the bagel. How much does the bagel cost? As Frederick points out, respondents miss the “bat and ball” problem far more often than they miss the “bagel and banana” problem. There is, to this extent, an element of experimental manipulation but this doesn’t vindicate the argument from below; since the aim is to test our susceptibility to certain kinds of cognitive illusion, it’s obvious that the chosen example needs to be one that is potentially illusion-generating. BAT AND BALL doesn’t so much provoke as evoke a flawed response, and the fact that it does that is the point of the example.

Dennett’s accusation is if anything even less pertinent in relation to by Ross, Lepper and Hubbard’s early work on belief-perseverance. Subjects were presented with the task of distinguishing between authentic and genuine suicide notes. Initially subjects were provided with false feedback indicating varying levels of success in the task, only to be told later that the initial feedback had been false. Subjects were then asked to fill out questionnaires asking them to estimate their actual performance at the task. The results showed a remarkable degree of postdebriefing perseverance, with subjects who had initially been given the false feedback that they were good at distinguishing authentic and fake suicide notes continuing to rate their abilities far more favourably than those who had initially been given negative feedback. As with BAT AND BALL there is little evidence here of illicit experimental manipulation or any attempt to ‘induce pathology’ by ‘putting strain’ on subject’s belief systems. The tests shows that test subjects’ responses to undermining evidence are in some sense less than ideal; it doesn’t induce them to respond ‘irrationally’, if that means that it gets them to respond in a way that is unrepresentative of how they and other humans would respond to such scenarios in real life. Anyway, as I’ve already argued above, it’s not clear that belief-perseverance is irrational.

This leads on to another, related point about the argument from below: the argument represents the evidence for the Disparity as experimental evidence, and then raises questions about the extent to which the evidence generalizes. However, it’s arguable that the idea of a significant Disparity is also part of folk psychology, and that there is plenty of evidence from everyday life that humans don’t approximate to homo philosophicus. The psychological data only confirm what many people who haven’t been influenced (or should that be corrupted?) by philosophy believe anyway; isn’t exactly news that our attitudes are in many cases not as they ought rationally to be, that we often reason carelessly, that most of us are no good at statistics, and that self-ignorance is a pervasive feature of human life. We don’t really need experimental psychologists to tell us these things, though for those who believe there is a significant Disparity it’s certainly reassuring that the scientific evidence supports their view.

The case of Oliver is particularly telling in this regard. We might be dismayed by the fact that there are so many real world Olivers with bizarre views about 9/11 and little sense of why they think the weird things they think about such events. But are we surprised? Hardly. For those of us who don’t start out with the touchingly optimistic and naïve vision of man as a model; epistemic citizen made in the image of God, the natural reaction to OLIVER and even KAREN is: of course that’s how it is. Why would you think otherwise? Dennett’s attempt to discredit the empirical evidence for the Disparity therefore misrepresents the source of the evidence as well as its credentials: it doesn’t all come from artificial laboratory experiments, and the simple reason it generalizes to the real world is that a lot of it is drawn from non-specialist observation of the real world.

No doubt there is much more to be said about all this but I hope I have said enough to justify concluding that the argument from below is no good. Its heart is in the right place since it’s certainly a good idea to assess the Disparity by looking at the actual evidence for and against rather than by armchair reflection. If you are sceptical about the prospects for a transcendental deduction of the Similarity Thesis then the obvious alternative is to challenge the evidence against this thesis. The problem is that this evidence is really rather strong: it turns out that there are solid empirical grounds for positing a Disparity that is large enough to create problems for the Similarity Thesis.

Where do we go from here? When I introduced Psychological Rationalism at the start of this chapter I characterized it as the view that humans are sufficiently similar to homo philosophicus for rationalism about self-knowledge, with its reliance on TM, to be a viable account of how we know our own beliefs and other attitudes. Most of this chapter has been about how far we are from being ideal epistemic agents, but a substantial Disparity might be something a Psychological Rationalist could live with at least to this extent: suppose it turns out that rationalism about self-knowledge is compatible with a substantial Disparity. I called this position ‘compatibilism’ in chapter 1, and if this turns out to be defensible then there is no need for Psychological Rationalism to try to show that we approximate to ideal epistemic agents. All it needs to show is that the Disparity, however substantial, still allows us to know our own attitudes by employing the Transparency Method.

I will come back to compatibilism in chapter 9. Before then, there are a couple of other questions to consider. One is whether it’s right to conclude from the Disparity that we are irrational or even, as has recently been claimed by the behavioural economist Dan Ariely, ‘predictably irrational’. This question, which I will discuss in the chapter after next, brings into focus an issue which I have mentioned a few times but not yet discussed in any detail: what is meant by the term ‘irrational’, and does the suggestion that we aren’t much like homo philosophicus commit one to endorsing the view that we are irrational? The other question is this: I have represented Psychological Rationalism as a descriptive thesis, as a view about what humans are actually like. This makes Psychological Rationalism something of a hostage to empirical fortune, and that is the point I have been exploiting in this chapter. However, it’s worth noting that there is a quite different way of thinking about Rationalism. The alternative is to represent it as a view about what is normal for humans, or with what we should be like, rather than with what we humans are actually like. So, for example, it could be normal for our beliefs not to survive evidential discrediting even if belief-perseverance isn’t uncommon in practice.

The key to this response to the Disparity is the thought that what is normal for us isn’t a statistical matter. This allows Rationalism to live with the Disparity because as far as the normative issue is concerned the Disparity is neither here or there. One question is: how do humans actually reason and revise their attitudes when faced with undermining evidence? A quite different question is: what is the normal or correct way for humans to reason and revise their attitudes in the face of undermining evidence. These questions can, of course, have quite different answers, and there is a form of Rationalism that concentrates on the second question and deems the Disparity to be irrelevant for the purposes of answering this question. This form of Rationalism points out that what is normal for us may not be what is common. I’m going call this view Normative Rationalism. Is Normative Rationalism any good and, does it have anything useful to say about the sources, character, and value of our self-knowledge? These are among the questions I want to address in the next chapter.