FIFTH EXCURSUS

Insulated Idealization and the Problem of Self-Doubt

Some important epistemological problems arise from cases of *self-doubt*: cases in which subjects have evidence that their own belief-forming processes are unreliable, and cases in which subjects cannot rule out the possibility that their belief-forming processes are unreliable. Cases of the former sort raise difficult questions about what subjects should believe when they have such evidence. If I have evidence that I am a poor mathematical reasoner, should I accept any mathematical conclusions? Cases of the latter sort threaten a sort of overarching metacognitive skepticism: if I cannot be certain that my belief-forming processes are reliable, then arguably I cannot be certain of the beliefs that they produce.

The problem of self-doubt raises a significant issue for scrutability theses, and it has consequences for the nature of the idealization we must invoke. Addressing these issues helps to shed some light on the problem of self-doubt more generally. I will discuss these cases first as a problem for Conditional Scrutability. I will then draw some more general conclusions and evaluate related problems for Inferential and A Priori Scrutability.

Say that John has recently been given an anti-arithmetic drug that is known to render users incompetent at doing arithmetic: any arithmetical belief of his will have (let us say) at most a 50 percent chance of being true. Let M be '57 + 65 = 122', which we can suppose has just been uttered by John. Let U be 'My arithmetic judgments are unreliable' (or more specifically, 'My judgments about M have at most a 50 percent chance of being correct'). Suppose that John comes to know that he was recently given the anti-arithmetic drug, so that he comes to know U. What should he then judge about M? There is at least a strong intuition that John should adopt a credence of at most 0.5 in M, or perhaps suspend judgment about M. To continue believing M in light of the evidence about his unreliability seems irrational.

Now suppose that John has not been given the drug, but that he cannot rule out the possibility that he has been given it. What should John judge about M, conditional on the assumption that he has been given the drug? That is, what is the ideally rational credence for him to have in M given U? As in the case above, it is plausible that this credence $cr'(M \mid U) \leq 0.5$. Even if John is in fact an ideal

reasoner, he should suspend judgment when he conditionalizes on the hypothesis that he is unreliable.

If this is right, a problem for Conditional Scrutability immediately threatens. There are subjects in the actual world who are unreliable about mathematics. For them, an analog of *U* is true. Suppose that *PQTI* is a conditional scrutability base for those subjects. Then *U* is conditionally scrutable from *PQTI*, so that $cr'(U \mid PQTI) = 1$. Given that $cr'(M \mid U) \leq 0.5$, and given that *PQTI* does not contain information beyond *U* that changes the conditional judgment about *M*, then $cr'(M \mid PQTI) \leq 0.5$. If so, *M* is not conditionally scrutable from *PQTI*. But *M* is true. So not all truths are conditionally scrutable from *PQTI*, contradicting the assumption that *PQTI* is a conditional scrutability base. If this reasoning applies to all putative scrutability bases, then the Conditional Scrutability thesis is false.

Someone might resist by holding that a scrutability base may contain relevant further information beyond *M* that changes the conditional judgment about *M*. Most obviously, the base might contain *M* itself. Still, given a paradigmatic scrutability base with only a posteriori truths, such as *PQTI*, it is hard to see how the rest of *C* will change the conditional judgment. So at the very least this problem forces us to expand the base considerably. Furthermore, one can raise an analogous problem involving an anti-scrutability drug (one that disrupts scrutability reasoning) or an anti-reasoning drug (one that disrupts all reasoning). Or one can simply note that many actual-world inhabitants are unreliable at reasoning and scrutability, so that analogs of *U* such as 'I am an unreliable reasoner' are true for them. It is then arguable (as discussed below) that even an ideal reasoner should suspend judgments that are conditional on *U*, and likewise should suspend all judgments that are conditional on bases from which *U* is scrutable. If so, there will be no base from which *U* and all other truths are conditionally scrutable.

Alternatively, one might respond that the ideally rational credences $cr'(M \mid U)$ and $cr'(M \mid PQTI)$ are 1: if John were ideally rational, then even on the supposition that *PQTI*, his ideal reasoning would allow him to know *M* with certainty. But this does not seem quite right. It is plausibly *irrational* to accept simultaneously that one's belief in *M* is unreliable and to be certain that *M*. (The statement 'It is raining and my judgment that it is raining is unreliable' seems to manifest a sort of irrationality that is reminiscent of Moore's paradoxical sentence 'It is raining and I do not believe it is raining'.) If I were to learn *U* (and to acquire no other new evidence), then rationality would require that if I do not question *U*, I should suspend judgment about *M*. So even before learning *U*, my credence $cr(M \mid U)$ should not be high. Even for an ideally rational being, $cr(M \mid U)$ will not be high. Although *U* is false in an ideally rational being's context, such a being may well have some tiny positive credence in *U*,

perhaps because they have a tiny positive credence that they have been given an anti-reasoning drug. That area of their credence space will be divided more or less evenly between M and $\sim M$. It follows that for each of us—whether or not we are ideally rational, and whether or not we have recently been given such a drug—$cr'(M \mid U)$ will not be high. The same goes for John. So the counterexample stands.

Someone might suggest that if one is reasoning ideally, one can know with certainty that one is reasoning well: perhaps by introspecting the quality of one's reasoning, or perhaps simply by introspecting one's judgments and then re-using the reasoning to determine that they are correct. This conclusion will be incompatible with U, at least if U says that one reasons poorly in every instance (as we may as well stipulate that it does). Then the objection will say that ideal reasoning can rule out U with certainty, so that $cr'(U) = 0$, and $cr(PQTI) = 0$ when U is scrutable from $PQTI$. Now, I am very doubtful that one can ever rule out U with certainty. But even if we can, this leaves open the question of what $cr(M \mid U)$ and $cr(M \mid PQTI)$ should be, especially if we allow credences conditional on hypotheses with credence zero (as I think one should). In general, a supposition (such as U) trumps any empirical evidence, including introspective evidence, that it conflicts with. So it is natural to say that under the supposition of U, any conflicting evidence deriving from introspection that tends to undermine U will be rendered irrelevant. If so, this evidence will not affect the values of $cr'(M \mid U)$ and $cr'(M \mid PQTI)$, and these credences should still be low.

This sort of problem affects much more than conditional scrutability. David Christensen (2007) has observed that a problem of this sort affects even our knowledge of logical truths. It is commonly held that if L is a logical truth, the rational credence $cr'(L)$ is 1. But it may well be rational to have a small positive credence in the thesis U that one is unreliable about logic: after all, one cannot exclude with certainty the hypothesis that one has recently been given an anti-logic drug. For the reasons above, $cr'(L \mid U)$ cannot be high. It follows that $cr'(L)$ cannot be 1.

The best way to handle this problem is to disentangle various different principles of rationality. It is clear that the principle of rationality invoked above is quite different in kind from ordinary principles of rational inference and the like. We might call it a level-crossing principle: it is a principle by which one's higher-order beliefs about one's cognitive capacity are used to restrain one's first-order beliefs about a subject matter. This principle governs how one should reason in a way quite different from standard principles of theoretical reason. In the case above, standard principles support believing M, but the level-crossing principle support suspending judgment about M, and in this case the second principle wins.

We can imagine a cognizer—call him Achilles—who is at least sometimes insensitive to this sort of level-crossing principle. On occasion, Achilles goes into the mode of *insulated cognition*. When in this mode, Achilles goes where first-order theoretical reasoning takes him, entirely insulated from higher-order beliefs about his cognitive capacity. He might acquire evidence that he is unreliable about mathematics, and thereby come to believe 'I am unreliable about arithmetic', but he will go on drawing conclusions about arithmetic all the same. We might say that in the insulated mode, his reasoning is *practically self-confident*, even if it is not *theoretically self-confident*. That is, any self-doubt manifests itself only in what Achilles believes, and not in how he goes about believing.

What if Achilles comes to believe that he has been taking a falsity pill, so that all of his mathematical beliefs are false? Then he will believe 'All my mathematical beliefs are false'. At the same time, through introspection he may figure out 'I have the mathematical belief that FLT is true' (where FLT is Fermat's Last Theorem). From these he would infer, by ordinary theoretical reasoning, that FLT is false. To avoid this result, we need to stipulate that when in the insulated mode, Achilles is also incapable of introspection.[1] In fact, to avoid indirect evidence of his beliefs through observing his behavior, we can stipulate that in the insulated mode, he is incapable of perception, too.

There is plausibly some sense in which insulated cognition is irrational, but it is a limited sort of irrationality. Suppose that Achilles is otherwise fully rational. And suppose that at a certain point of time, he might either engage in insulated reasoning or fully rational reasoning, where we stipulate that in both cases this involves armchair reasoning (without perception or introspection) that may exploit existing beliefs. Then insulated cognition will yield at least as many true beliefs as the fully rational mode, and in some cases (those in which Achilles has misleading evidence for his irrationality) it will yield more. It is only if Achilles has independent sorts of theoretical irrationality that his cross-level irrationality will be a problem, causing him to keep forming false beliefs where a fully rational creature would be restrained.

Insofar as it is reasonable to postulate ideal cognizers at all, there seems to be no bar to postulating *insulated ideal cognizers*: cognizers whose rational processes are practically insulated from higher-order beliefs, as Achilles' processes are, but

[1] What if Achilles also has a prior belief, formed before he entered insulated mode, that he will believe FLT? One could attempt to exclude such beliefs by requiring that Achilles disregard all evidence and beliefs from before he entered insulated mode, or at least that he disregard all empirical evidence and beliefs. But this would be too close to restricting Achilles to a priori reasoning, which would have the undesirable effect of making the notion of conditional scrutability depend on the notion of apriority. Instead, one can rely on the observation that if Achilles believes or supposes that he has been taking the drug, this will have the effect of undercutting his prior grounds for believing that he will believe FLT, or at least of rendering any such belief uncertain. So his insulated reasoning in support of FLT will overwhelm these defeated grounds for denying FLT.

are otherwise ideal. As we have seen, insulated ideal cognizers are in some ways more successful cognizers than fully ideal cognizers, at least where non-empirical reasoning is concerned, because their cognition is never affected by misleading self-doubts. For example, a fully ideal cognizer may have some small positive credence in its own unreliability (it cannot exclude with certainty the hypothesis that it has recently been given the drug above), so it will correspondingly never be absolutely certain of anything, even of logical truths. By contrast, there is no corresponding bar to an insulated ideal cognizer's being certain of logical truths.

Rational idealizations need not be cashed out in terms of ideal cognizers.[2] More fundamentally, they involve ideal norms or ideal warrants. One can cash out an insulated idealization in normative terms, speaking of what one ideally ought to believe (starting from one's current state) if cross-level principles are set aside. We might even define a notion of what one 'ought*' to believe that works in this way, and a corresponding notion of one's 'rational*' credence in a sentence, $cr^*(S)$. For example, where L is a logical truth, then even if one's ordinary rational credence $cr'(L)$ is less than 1 for the reasons above, it may be that the insulated rational credence $cr^*(L)$ is 1.

One can also cash out an insulated idealization in terms of warrant. An insulated warrant is a warrant that gives no role to level-crossing principles of support. It is arguable that even after John has taken the anti-arithmetic drug, there exists an insulated warrant for John to believe M, and even for John to be certain in M. A proof of M provides such a warrant, for example. It is just that John is not in a position to take advantage of that warrant. Likewise, before taking the drug, there is an insulated warrant for John to have conditional credence 1 in M given U. We can then say that $cr^*(P)$ is the credence in P for which there is an insulated warrant. It is arguable that any warrant is an insulated warrant. On this view, level-crossing principles do not play any role in ordinary warrants, so that there is an ordinary warrant for John to be certain in M in the case above (a lower credence may be rational, but it is not warranted). I will not try to adjudicate this matter here, but if this view is correct, it may be another place where warrant can play an especially basic role.

[2] Still, a bonus of the insulated idealization is that it overcomes one familiar problem in appeals to ideal cognizers. Typically, one cannot simply identify one's rational credence in P with the credence one would have in P if one were to become an ideal cognizer, as this would entail that everyone has an overly high rational credence that they are ideal cognizers, and so on. However, this problem does not arise on the insulated idealization, because of the bar on introspection. So this problem does not exclude the thesis that one's rational credence in P is the credence one would have in P if one were to become an insulated ideal cognizer (starting from one's current state). I will not rely on this thesis, in part because it is not obvious that there could be a truly ideal cognizer, as opposed to a series of more and more ideal cognizers. (A propositional version of the thesis, like other theses involving modal idealizations, also runs into trouble in cases of semantic fragility.) But the thesis may occasionally be a useful aid in thinking about insulated credences.

We can then say that S is conditionally scrutable from C if the *insulated* rational credence $cr^*(S \mid C)$ (which is identical to $cr^*(S \mid CC)$, where CC is a conjunction of the sentences in C) is high. Understood this way, the drug case poses no problem for the Conditional Scrutability thesis. In this case, although $cr'(M \mid U)$ is not high, $cr^*(M \mid U)$ is still plausibly high. The belief that one is unreliable about arithmetic has no impact on one's insulated rational credence in M, and likewise the supposition that one is unreliable has no impact on one's insulated conditional rational credence. So Conditional Scrutability is unthreatened.

I think that for many purposes involving theoretical rationality, insulated rational credence is often the most useful notion. Certainly, insulated rational credence seems to better reflect the sort of claims that theorists often make about rational credence. Where non-insulated rational credences are concerned, even tiny empirical self-doubts will infect the analysis of all sorts of otherwise well-behaved matters, in ways that are hard to regiment, and that will render many standard claims of formal epistemology false. (Logical truths will not have rational credence 1, Sleeping Beauty will not have rational credence 1/3, and so on.) The insulated idealization keeps the focus on first-order theoretical reasoning, allowing a more straightforward analysis of theoretical reason. Of course when it is relevant one can still invoke a non-insulated idealization, in order to see how first-order and higher-order reasoning interact, and to determine what it is rational (simpliciter) for a subject to believe.

The insulated idealization allows us to take seriously the thesis of Conclusive Conditional Scrutability. Here S is conclusively conditionally scrutable from C iff $cr^*(S \mid C) = 1$, and the thesis holds that $cr^*(S \mid C) = 1$ for all truths S and for the appropriate compact C. Where ordinary rational credences are concerned, a thesis as strong as this is out of the question for reasons discussed above. But for insulated rational credences, the thesis may well be correct. I will return to this matter in the next chapter.

What about Inferential Scrutability? There are presumably some domains about which I am actually unreliable: that is, there are classes U of sentences such that my beliefs about sentences in that class are only 50–50 likely to be true. Let S be the sentence: 'I am unreliable about the sentences in U'. Then S is true. S will presumably be inferentially scrutable from the right sort of scrutability base C: that is, if I came to know the sentences in C, I would come to know S. But if I knew (or even believed) S, then I should rationally suspend judgment about the sentences in U, so I could not know the sentences in U. So it appears that these sentences will not be inferentially scrutable from C. Still, one could argue that this is a Fitchian case: properly investigating S requires ideal reasoning, so that a proper investigation would render S false. This allows that if I properly came to know enough sentences in C, I would come to know the true

sentences in *U*. So it is not obvious whether self-doubt causes a problem for Inferential Scrutability. If it turns out that it does, however, we can invoke an insulated idealization as above.

What about A Priori Scrutability? If *M* is a mathematical truth, then given that *M* is a priori, any material conditional $D \rightarrow M$ will also be a priori. For similar reasons it follows that *M* is a priori scrutable from any base, even a base specifying a world where the subject has been given an anti-arithmetic drug. At worst, the Achilles worry suggests that a subject cannot come to be certain of *M* by a priori reasoning, for reasons akin to those discussed earlier in the case of logic. If one is interested in conclusive a priori knowledge, which requires certainty, one can invoke an insulated idealization in one's definition of apriority: for example, *S* is a priori if insulated ideal a priori reasoning could bring about psychological certainty in *S*. Or perhaps best, one can say that *S* is a priori when there is a conclusive (insulated) a priori warrant for believing *S*, where a conclusive warrant is one that supports certainty.

3

Adventures with a Cosmoscope

1 A Scrutability Base

To argue for scrutability theses, as I do in this chapter and the next, we first need a potential scrutability base. I will start with a reasonably generous base. In later chapters (6 and 7) I will consider whether it needs to be augmented and whether it can be narrowed down.

In the *Aufbau*, Carnap's main base was a phenomenal base, consisting just of logical expressions plus an expression for phenomenal similarity (similarity of conscious experiences). This base was rejected by Goodman on the grounds that it does not definitionally entail all truths about specific phenomenal qualities of experience, and by Quine on the grounds that it does not definitionally entail physical truths about spatiotemporal location in the external world. More generally, it is commonly believed that no set of phenomenal truths a priori entails all truths about the external world. If so, a scrutability base must involve more than phenomenal truths.

Carnap suggests in the *Aufbau* that he might instead have used a physical base. In a 1927 letter to Moritz Schlick (see Coffa 1985, p. 403), he says that he plans to publish two *Aufbau*-like books, one with a phenomenal base and one with a physical base. He says that in some ways the second book would even better deserve the title *The Logical Structure of the World*, while the first book (the actual *Aufbau*) might more accurately be called *The Logical Structure of Knowledge*. All this suggests some sensitivity to the limitations of a phenomenal base and openness to a physical base, presaging the physicalist orientation of his work in the 1930s.

A physical base would have avoided Goodman's and Quine's problems, but it would have had other problems. Just as a phenomenal base has trouble accounting for physical truths, a physical base has trouble accounting for phenomenal truths. In particular, it is arguable that phenomenal truths ('Someone is conscious',