

First Draft of "Moderate Rationalism and Bayesian Scepticism"

Brian Weatherson

May 20, 2008

1 Against Moderation

This paper is part of a larger campaign against moderation in foundational epistemology. I think the only plausible responses to a kind of Humean sceptic are, radical responses. The Humean sceptic I have in mind tells us about a sceptical scenario, *ss*, where our evidence is just as it actually is, but some purported piece of knowledge of ours is false. The sceptic names the proposition *You aren't in ss* as *s*, and calls on us to respond to the following argument.

1. You can't know a priori that *s*
2. You can't know a posteriori that *s*
3. So, you can't know that *s*

There are five interesting kinds of response to this sceptic's call, of which I think only two are ultimately adequate.

- The **sceptic** says that the conclusion is true. I include as sceptics those like Nozick and Dretske who say that we can't know we aren't in *ss*, although somehow we can know things that entail it.
- The **strict evidentialist** says that the scenario the Humean describes is incoherent. That's because of the of the sceptic's use of evidence in setting up *ss*. If the purported piece of knowledge is actually part of our evidence, and evidence is all true, this is incoherent. The contemporary paradigm here is Timothy Williamson (2000), though the core idea traces back at least to Austin.
- The **radical rationalist** says that premise 1 is false because we can have fallible a priori knowledge; for example, we know a priori that *s* even though it could turn out to be false.
- The **radical empiricist** says that premise 2 is false because we can learn a posteriori both facts about the world, and facts about how evidence justifies belief in various empirical propositions. The *radical* nature of this proposal will become a little clearer in what follows. (Obviously this bears little relation to Jamesian radical empiricism; the names are attempts to recapture some useful terms.)
- The **moderate rationalist** says that premise 2 is false, but we could have known a priori that if we had the evidence we actually had, we would have been *justified* in believing we weren't in *ss*. What we couldn't know a priori was that this justified belief is true.

As set up so far, the sceptical argument is somewhat schematic. It doesn't distinguish between religious external world scepticism, scepticism about philosophical knowledge, scepticism about other minds, and inductive scepticism.

It may well be that strict evidentialism is the right response to some of these sceptics, as (arguably) Austin thought it was the right response to the external world sceptic. There is something very natural about saying that right now my perceptual evidence includes that there is a computer on my desk. It is comparatively unnatural to say that my evidence merely includes that there *seems* to be a computer on my desk. So it is comparatively unnatural to say that I could have had just this evidence, without there being a computer on my desk. We might be forced to say this after careful investigation of the nature of perception, but I think the strict evidentialist line has prima facie plausibility.

Similarly, there are some fields where the sceptic has some plausibility. The point of Hume's *Dialogues*, after all, is to note that religious scepticism can thrive even without any principles that would generate excessive scepticism.

At risk of elevating personal preferences into philosophical insights, I think the most important of these is **inductive** scepticism. (There's a reason I said this was a **Humean** sceptic.) And here I think both of the first two options are highly implausible. There are many things I know about the future. I'm not, sadly, going to win a gold medal at the next summer Olympics. There is some possible world where I get added to some team, perhaps deep on the bench of the hockey team, and that team wins a gold medal. Perhaps there is some possible world where I get into some race, and everyone else falls down in every leg. So it's metaphysically possible. But it ain't gonna happen.

Still, it's hard to say that my evidence need be different in a world where I win a gold medal. My evidence that I'll win a gold medal is that I'm not on any of the relevant teams that could win golds, that I'm not fast enough, strong enough etc to win races without bizarre disasters striking all other competitors, that those disasters haven't ever struck in the modern Olympics, and so on. I have counterparts with exactly the same causal history as mine, and exactly the same mental and physical states that are (generally) correlated with external facts. That's to say, they have the same evidence as I do. But they will, though bizarre luck, win gold medals. So I could win a gold medal despite having the evidence I do. So the Humean sceptical scenario here is possible. So I don't think a strict evidentialist line is plausible when it comes to inductive scepticism. (Note that I haven't denied here that there might be a tight connection between knowledge and evidence. Perhaps all evidence is knowledge. I've just denied that all knowledge is evidence, particularly inductive knowledge.)

That leaves us with three options, two of them radical and one moderate. And, as I said at the top, my inclination is to favour radicalism. I think in fact that both radical options are defensible. (That is, I don't know of any knockdown arguments against either.) But I think the moderate option is inherently unstable, and ultimately indefensible.

The general style of argument against moderate rationalism is to consider some empirical proposition p that we know on the basis of our evidence e , and argue that, in order for us to know p on the basis of e , there must have been some proposition, perhaps $e \rightarrow p$, that was knowable antecedently to getting evidence e . But since e was all our evidence, then $e \rightarrow p$ must have been known, or at least knowable, a priori. And that contradicts the *moderate* part of moderate rationalism. (Note that I'm using \rightarrow to pick out material implication throughout.)

In "Scepticism, Rationalism and Externalism", I argued that this kind of argument could be shown to work on fairly weak auxiliary assumptions if the moderate rationalist in question was also assumed to be an internalist about justification. The argument there relied primarily on some moderately strong closure principles concerning knowledge and justification. But of course this argument does not tell at all against externalist moderate rationalism, and I left it as an open question at the end of that paper whether such a position was viable.

Here is the kind of view I have in mind as the target in this paper. Consider a theorist who accepts the following three principles.

1. One can't know a priori any proposition that could have turned out to be false.
2. S's belief that p is justified iff it was produced by a reliable method, and any true belief that is justified in this sense constitutes knowledge unless it is either unsafe or inferred from a false premise.
3. Principle 2 is knowable a priori (and, though this won't matter for what follows, necessary).

I'm not sure anyone accepts these three principles, especially with principle 2 stated quite so simply. But I think, or at least hope, that the simplification in principle 2 won't be too damaging to the argument to follow. Because I do think there is an interesting class of views that have this form, with a more complex clause on the right-hand-side of principle 2. And those are what I want to argue against in this paper. (Note also that the second clause of premise 2 doesn't aim to *analyse* knowledge, but merely to state some sufficient conditions for knowledge.)

Here is the argument I'm going to defend against moderate rationalism has the following structure. (In what follows I take e and p to be names for a particular body of evidence and a particular proposition respectively, and K to be a particular possible agent. I'll say more about just what those terms name below.)

1. For any A , if K can know A upon acquiring evidence e , then either (a) K must have known A before getting evidence e , or (b) K must have had reason to believe antecedently that e is evidence for A , or (c) e itself must be evidence that e is evidence for A .
2. If moderate rationalism is true, then K doesn't know $e \rightarrow p$ antecedent to getting e .
3. If K doesn't know $e \rightarrow p$ antecedent to getting e , then K doesn't have reason to antecedently believe e is evidence for $e \rightarrow p$.
4. If moderate rationalism is true, then e itself is not evidence that it is evidence for $e \rightarrow p$.
5. K can know $e \rightarrow p$ upon acquiring evidence e
6. So, moderate rationalism is false.

In section 2 I'll look at a Bayesian argument that may have sceptical implications, and argue that it supports premise 1 and 3 of the argument. In section 3 I'll say a little more about e and p , and hopefully making clear that premises 4 and 5 are true. I'll conclude by arguing in favour of premise 2, which is perhaps the most contentious of the lot.

2 Bayesian Constraints on Learning

There has been a lot of discussion in the literature recently about the following theorem.

$$Pr(A \rightarrow B | A) \leq Pr(A \rightarrow B), \text{ with equality only if } Pr(A) = 0 \text{ or } 1.$$

The proof of this is fairly straightforward, once we notice the following two ways to expand the LHS and RHS of the equation.

$$Pr(A \rightarrow B | A) = Pr(A \rightarrow B | A)Pr(A) + Pr(A \rightarrow B | A)Pr(\neg A)$$

$$Pr(A \rightarrow B) = Pr(A \rightarrow B | A)Pr(A) + Pr(A \rightarrow B | \neg A)Pr(\neg A)$$

Since $Pr(A \rightarrow B | A) \leq 1$, while $Pr(A \rightarrow B | \neg A) = 1$, the result we want follows immediately from these expansions. As David Jehle pointed out to me, when we go somewhat more slowly through the proof, it turns out to rely heavily on some distinctively classical principles. (Neither expansion is in general correct in an intuitionistic setting.) But in this paper we'll assume that our background logic is classical, so we'll take this to be a theorem. It's also easy enough to verify that there is equality only if the probability of A takes one of the extreme values. I'll assume in what follows that it does not, because that's going to be primarily the case in which we're interested in epistemologically.

The question is what we should make of the theorem. One popular Bayesian interpretation of it is that we cannot learn $A \rightarrow B$ by learning A . That's because, in general, only evidence that confirms a conclusion is the right kind of evidence on which to learn that conclusion. And only evidence that raises the probability of a conclusion (or at the very least does not lower it) confirms the conclusion. Since conditionalising on A lowers the probability of $A \rightarrow B$, it follows that we cannot learn $A \rightarrow B$ by learning A .

I think this is rather too quick. It presupposes that all learning is by conditionalisation. That is, it presupposes that what to do after getting some evidence, say A , is given by the prior conditional probability given A . That is, it presupposes that before we get evidence A , we can say what the thing to do if we get evidence A is. That's sometimes true. It's true in, for example, the games of chance for which the probability calculus was developed and in which Bayesian approaches seem undeniably correct. But it can hardly be always true, unless we want to say that we can know a priori what the right reaction is to any evidence. We'd have to say that, because A could be any given person's entire body of evidence, and what they can say prior to getting that is what they can say a priori.

In "The Bayesian and the Dogmatist" I developed this line of reasoning somewhat more extensively, and proposed a formal theory that allowed that getting evidence A could let us learn what the right thing to do with that evidence is. Here I want to note the converse point.

That possibility, the possibility that getting evidence A could let us learn what the right thing to do with that evidence is, seems to be the only way for the Bayesian argument to fail. If getting evidence A doesn't tell us anything about what to do with it, then there is, as far as I can see, no reason not to simply conditionalise on A . Perhaps there is some other evidence such that if we had it, we would have a better sense of which credences were rational given evidence A . Perhaps without that evidence the best we can do is have credences that are more imprecise than in some sense they should be. That doesn't undermine what I'm saying here, which is simply that conditionalisation on our prior credal states looks like the right thing to do unless we get some evidence that we had misunderstood the connection between a piece of evidence and the world, i.e. unless we got evidence that our prior *conditional* credal states were defective.

That's just to say that premise 1 is correct. It looks like there are three ways that K can know A after getting evidence e . She can antecedently know e , and A not defeat that knowledge. She can antecedently know that e is evidence for A , and then update when she gets e . That is, she can have antecedent reason to be highly confident in A given e , and this conditional confidence can become unconditional confidence when she gets e . Or she can learn that e is evidence for A . That's just what premise 1 says, so premise 1 is true.

Moreover, when conditionalising on e cannot raise the probability of A , the second of those options is ruled out. The second option says you weren't antecedently confident in A , but you were antecedently confident in A given e . But if the conditional probability of A given e is lower than the probability of A , that seems out of the question. And the proof showed above that when A is $e \rightarrow p$, that's exactly what happens. So of the three options left open by premise 1, only the first and third are real possibilities. That's what premise 3 says, so premise 3 is true.

3 Commitments of Moderate Rationalism

The moderate rationalist, in the sense I'm interested in here, thinks that there are substantive things we can say a priori about the nature of epistemic justification. For example, we can say that beliefs produced by reliable methods are justified. Now we can't, typically, know that a particular method is reliable a priori. But we can often learn that a method is justified. For instance, I know that believing the contents of visual representations that I form in good daylight concerning medium-sized objects is quite reliable, and all such beliefs are formed by some reliable method. (I'm not sure whether that is a method; perhaps it is too gruesome to be a method. So I don't know that it's a reliable method. But I do know that every instance of it is an instance of a reliable method.)

Since the moderate rationalist says that we can know substantive things about justification a priori, she's committed to thinking the following kind of situation is possible. There can be a knower K , a method M , a property F of methods, a piece of evidence e , and a proposition p such that:

- K knows a priori that if M is F , then beliefs produced by method M are justified;
- K knows that M is F ; and
- K knows that if she gets evidence e , using method M will produce a belief that p .

That's fairly abstract, so let's fill it out with an example. Assume the particular moderate rationalist we're interested in is a reliabilist, so F is the property of being reliable. Assume that K has learned that a particular website, truereports.com, is a highly reliable guide to the weather in London. So if truereports.com says that it is raining in London, it usually is raining in London. And if truereports.com says that it is sunny in London, it usually is sunny in London. Method M then is to believe the contents of meteorological reports from truereports.com as they pertain to London. Let e be that truereports.com says that it is raining in London. And let p be that it is raining in London. So we have the following possible case.

- K knows a priori that if believing meteorological reports from *truereports.com* is reliable, then beliefs produced by believing meteorological reports from *truereports.com* are justified;
- K knows that believing meteorological reports from *truereports.com* is reliable; and
- K knows that if she gets as evidence that *truereports.com* says that it is raining in London, using method the method *believe meteorological reports from truereports.com* will produce a belief that it is indeed raining in London.

This kind of situation doesn't seem at all unusual; we're not actually focussing on particularly odd sceptical scenarios here. But there are two points we can make about the example, if we continue it by assuming that K does in fact see that *truereports.com* says it is raining in London, and that this report is in fact correct.

First, this evidence might be quite good evidence that it is raining in London. But it is rather poor evidence for the material implication *Truereports.com says that it is raining in London* \rightarrow *It is raining in London*. That's because just learning what *truereports.com* says about the weather in London is really no evidence at all for what conclusions are justified on the basis of this report. And as argued above, that means that K should simply conditionalise on the new evidence. But conditionalising on the fact that *truereports.com* says that it is raining in London can't possibly increase the probability of that material implication. So it can't be evidence for the material implication. That's just what premise 4 says, so premise 4 is true.

Second, since this method is in fact reliable, K can come to have a justified belief that it is raining in London. This belief is not only justified, it appears to be free from any obvious defeaters. So K knows that it is raining in London. This shouldn't be too surprising. We learn what the weather is like by reading reports all the time. But once K knows that it is raining in London, she then is in a position to infer the material implication *Truereports.com says that it is raining in London* \rightarrow *It is raining in London*, i.e. $e \rightarrow p$. That's because that material implication is an obvious consequence of p itself. I'm not saying that this is the first time K can come to know $e \rightarrow p$, but whether or not she could know it before getting evidence e , she can know it now. That's what premise 5 says, so premise 5 is true.

4 Moderate Rationalism and Conditional Knowledge

The only premise I haven't yet defended is premise 2, which says that if moderate rationalism is true, then K can't know $e \rightarrow p$ antecedent to getting e . At one level it looks like this should be easy to defend. After all, the moderate rationalist says that we can't know anything that could have turned out to be false a priori. And $e \rightarrow p$ could have turned out to be false. So we can't know it a priori.

But that won't do in the context here, because I haven't assumed that e is all of K's evidence. Perhaps although K didn't know $e \rightarrow p$ a priori, she learned it on the basis of earlier acquired evidence. That seems to be more than a theoretical possibility. For all we've said so far, she might have good inductive evidence, good enough even to count for knowledge, that *truereports.com* will not misreport the state of the weather in London. If I checked the weather report on some actual sites, I would know before looking that they would be accurate. That's to say, I'd know before looking that it wouldn't say it was snowing while actually being warm and sunny. That's to say, I'd know something like $e \rightarrow p$ before getting e . I think this *isn't* the general case though.

Let's further stipulate, as seems at least possible, that K doesn't have a sufficient basis for making this kind of inductive inference. She didn't, that is, learn that *truereports.com* is reliable by seeing how reliable they had been on several past occasions. Rather, she learned this directly, perhaps by testimony from a known to be reliable source. She then received evidence *e*, i.e. saw that *truereports.com* said that *p*. And then she knew that *p*.

Still, perhaps this report that *truereports.com* was generally reliable was enough for her to form a justified belief in $e \rightarrow p$. I doubt this can be the case though. Let's consider two cases. It will make the cases easier to state if we introduce one knew name, *r*, for the proposition that *truereports.com* is generally reliable.

The first case is that it is possible for there to be no processes K could use which are, in her setting, reliable processes to infer from the general reliability of *truereports.com* to the accuracy of its next report. Perhaps in her world induction generally is not reliable. Or perhaps, more plausibly, in her world the accuracy of a website is always a *grue*-like, unprojectible feature of the world. (That sounds not entirely unlike our world.) If that's possible, we'll stipulate that it is the position K is in. So her inference from *r* to $e \rightarrow p$ is not a reliable inference, and could not be. So before getting evidence *e*, she couldn't know $e \rightarrow p$, as required.

The second case is that this is not really possible. Perhaps as a matter of necessity, inferring from the reliability of a source to the accuracy of its next report is a reliable process, or at least is an instance of some reliable method or other. If that's so, it seems like it must be true on fairly general conceptual grounds. It's hard to see how it could be a necessary a posteriori truth, like the chemical composition of water. So probably K could have figured it out before learning *r*.

Perhaps that is too strong, though it sounds plausible enough to me. But here's a weaker claim that is much harder to reject. If K knows $e \rightarrow p$ before getting *e*, she knew $r \rightarrow (e \rightarrow p)$ before getting *e*. But it is hard to see how *r*, or anything else, could be evidence for that. What ensures that her belief in this conditional is justified are some deep conceptual truths about the nature of justification, plus some very abstract facts about the necessity of the existence of certain reliable methods. Neither fact is guaranteed, or made more likely, or made clearer, by learning *r*. So she knew this conditional before learning *r*. But this argument works well enough even *r* was a piece of basic knowledge, indeed her only piece of basic knowledge. That's to say, she could know $r \rightarrow (e \rightarrow p)$ a priori. But since it could have turned out to be false, she could know something that could have turned to be false a priori. And that is inconsistent with moderate rationalism.

So in one case, she couldn't know $e \rightarrow p$ before getting evidence *e*. In the other case, moderate rationalism fails. That's to say, if moderate rationalism is true, she couldn't know $e \rightarrow p$ before getting evidence *e*. That's what we wanted to prove to establish premise 2.

That completes the refutation of moderate rationalism.

5 Where It's At

The failure of moderate rationalism leaves two reasonably different options.

One option is to simply accept that we can have a priori knowledge of things that could turn out to be false. That is, to accept fallibilism about a priori knowledge. This doesn't seem like a particularly radical step, if we also accept, as we should, fallibilism about a posteriori knowledge. But in practice it seems to be an option that many epistemologists are unwilling to accept. So I'll conclude with some words about the other option that remains open.

The above argument is no refutation of reliabilism. It is consistent with everything I've said above that reliabilism is true, but isn't knowable a priori. It's not crystal clear just what would motivate such a position; most of the arguments for reliabilism seem to be fairly a priori in nature. But it is a position that could be taken.

More generally, we could simply accept there's nothing substantive that can be said about epistemic justification a priori. There's no property of belief forming methods that we could know a priori entails that method produces justified beliefs. Perhaps that's because a kind of particularism about epistemic justification is true; there's simply no property methods can have whose possession (non-trivially) entails that beliefs produced by that method are justified. At most we can produce *ceteris paribus* generalisations about what produces justified beliefs if there is no particular reason that it doesn't work on an occasion.

The description of K's situation above relied crucially on two factors. First, the assumption that it was possible to specify such properties, so K could know simply that a method has the property that make its outputs justified. Second, the assumption that K could know not just that the method has this property, but that it is a method which produces justified beliefs. Someone who is generally sceptical that we can "factorise" claims about justification into normative facts (any method with property F produces justified beliefs) and descriptive facts (this method has property F) won't believe that. Nor will anyone who is sceptical that we can know a priori substantive normative claims such as that any method with property F produces justified beliefs. That leaves a lot of space for responses that deny premise 2 of the original sceptical argument, although it will be a space that excludes the view that we can do substantive work on the nature of justification on purely a priori grounds.