Some Questions about *the Problem of the Problem of Induction.*

Fabrizio Cariani
Northwestern University

White advances two central claims. The first is that a version of the problem of induction involves the reliability principle.

> **reliability principle**: If S considers whether the methods or rules she followed in concluding P are reliable, and she is not justified in believing that they are reliable, then she is not justified in believing P.

The reliability principle threatens the status of beliefs that are formed by inductive methods because it seems difficult to justify the reliability of induction. Despite the threat, White identifies the beginnings of an anti-skeptic rejoinder to this argument. The second claim is that [J1] we are justified in believing that induction is reliable and [J2] in fact, we might be *a-priori* justified. White does not claim to have established either point: it might be enough for his purposes to cast sufficient doubt on their negation (so that an opponent can't simply assume them to be false). My comments focus entirely on this second half of White's paper—and in particular on the considerations that support [J1].

1. The argument for J1

How can we be justified in believing that induction is reliable? Make the task concrete by imagining a particular reasoner, call him Theo. Theo goes through a sequence of observations—say, for simplicity, binary observations (e.g. rain/no rain; sun/no sun, etc.). Let *E* denote the entire sequence of Theo's observations. At any given time, Theo observes an initial segment of *E* and can form beliefs about the next element of the sequence. He does so by applying an inductive method to that initial segment. White does not say exactly what it means to follow an inductive method (this won't matter until the end of this commentary).

He also does not explicitly say which of many possible notions of reliability is at work in the reliability principle. There are a few candidates. According to one notion, a method is reliable just in case it can be expected to result in more true beliefs than false ones. According to another, a method is reliable just in case it results in more true beliefs than false ones given the actual sequence of observations. Many of White's arguments—and all of the ones I discuss here, presuppose the latter notion of reliability.[1]

Another aspect of White's setup will play a role in my discussion: there must be a gap between *reliable* methods (relative to a sequence *E*) and *unreliable* ones. That is, there must be methods that (relative to some sequence) are neither reliable nor unreliable. To see why, note that after considering a rather irregular sequence of observational inputs, White remarks:

1 Specifically, White supposes it is legitimate to determine the reliability of a method relative to a single observational history. This is only legitimate if we are talking about actual reliability.

> while this may not be the best sequence on which to apply inductive rules, it is not one that will tend to lead inductive reasoners into *error*. For induction is barely *applicable* to this sequence.

Though this passage does not mention reliability, the conclusion to be drawn from it is that one can't claim that induction is unreliable on the basis of highly irregular sequences. There must then be a three-way classification of methods relative to sequences. *M* is *reliable* relative to *E* iff (a) *M* recommends formation of a large enough number of beliefs and (b) most of these beliefs are true. *M* is *unreliable* iff (a) holds but (b) fails. *M* is *not evaluable for reliability* iff (a) fails.

Back to Theo. We do not suppose that he has access to the whole of *E*. Nonetheless, we could imagine looking at the whole of *E* 'from outside', so to speak. We could, then, classify sequences like *E* in three ways. *E* is *regular* iff a majority of steps are predictable by a simple regularity. It is *unexcitingly irregular* iff there is no regularity whatsoever. Finally, *E* is *excitingly irregular* iff there is enough regularity that Theo's inductive methods recommend forming several beliefs but enough irregularity that they are often wrong.[2]

We can now state an argument for [J1]:[3]

> (W1) If *E* is regular, induction is reliable and hence it is not unreliable.
> (W2) If *E* is unexcitingly irregular, induction does not recommend forming many beliefs and hence is not unreliable.
> (W3) It is very likely that *E* is either regular or unexcitingly irregular.
>
> ---
>
> (C0) It is very likely that induction is not unreliable.

Let us grant that the argument from (W1)-(W3) to (C0) is valid.[4] Note that (C0) does not rule out the possibility that induction is unreliable. If Theo's world is excitingly irregular, after all, induction may systematically produce false beliefs. (C0) is a bit short of the conclusion we want to reach. For one thing, given how I set things up, "not unreliable" does not entail "reliable", because condition (a) might fail. Nonetheless, we might also add the assumption that, in the sorts of cases we care about, induction does recommend forming many beliefs, so condition (a) is satisfied. If so, it will follow that:

> (C1) It is very likely that induction is reliable.

An additional step is necessary to reach (J1), which recall is our target. (C1) must imply:

---

2 Note that what counts as excitingly irregular depends on what sorts of methods Theo uses.

3 This is my reconstruction: I do not know if White would accept it. As I mentioned earlier, it's not clear that White expects to have an argument for (J1) It might be enough for him to cast sufficient doubt on the negation of (J1). I set this up as an argument because it facilitates my making of certain points.

4 I believe it is valid. But proving that it is requires making assumptions about the semantics of conditionals and probability operators that are not necessary here.

(J1) We are justified in believing that induction is reliable.

Depending on how one understands justification, one may reasonably push back against this transition. Not I, however: I am happy to stipulate the conditional *if (C1), then (J1)*, thus making the argument to (J1) valid.

Should we accept the premises of this argument? It is reasonable to concede (W1) and (W2). The philosophical core of the argument is (W3). White's reasoning for (W3) invokes something like the principle of indifference.

> When we consider the vast array of ways that stars can be arranged, it would be unreasonably arbitrary to put a lot of confidence in some possible arrangements over others. Star arrangements that compose a portrait of me make up a minuscule proportion of the total possibilities. So we should put very little confidence in that obtaining, and hence be quite sure that it won't.
>
> So if a situation in which induction was genuinely unreliable required a *very specific* arrangement of things, then perhaps we can be a priori justified in believing that this is not the case. Of course we can't entirely rule it out. But it should seem far less disturbing.

In other words, if the world is not conspiring against Theo, it is extremely unlikely that his experience will be excitingly irregular. After all, excitingly irregular sequences are few and far between. What if the world does conspire against Theo? Might there not might be a demon whose task is to systematically deceive him? But that hypothesis too, White notes, requires the conjunction of several individually improbable assumptions: there must be a demon; Theo must pay attention when the demon sets him up to have a certain experience (as opposed to looking elsewhere); the demon must be physically capable of producing the right observation (most demons can't stop the sun from rising) and so forth. Again, then, we might be justified on the basis of indifference considerations in believing that such formidable coincidences are improbable.

2. Indifference and the reliability principle.

I worry about the legitimacy of White's appeal to the indifference considerations in this dialectical context. As normally understood, the principle of indifference constrains distributions of credences in situations of evidential symmetry.[5] The anti-skeptic's argument needs something slightly different from this principle, though obviously related to it. The argument must invoke a method for forming justified qualitative beliefs. Let's say then that the indifference method is the following belief-forming method.

5 There are notorious problems with formulating the principle of indifference, but none of these are the basis of my worry. I am convinced by White's contention that there ought to be some cleaned up (or appropriately restricted) version of the principle that rationalizes the sorts of intuitive judgments that underlie the principle of indifference (e.g. that a randomly chosen string of letters is unlikely to be the complete text of Moby Dick).

indifference method: If nothing in our evidence favors any member of a very large partition over any others, then for each member *P* of the partition, one is justified to believe that *P* is false.

The indifference method implies, for instance, that I am justified in believing that my descendants won't all have the same birthday.

So, does the indifference method contribute to a justification of the reliability of induction? I am not sure. The reliability principle—the higher-order principle we started with—applies to the indifference method just as it applies to the induction method. Applying the principle to a belief in P, formed by the indifference method, we get:

> if S considers the indifference method, and she is not justified in believing that it is reliable, then she is not justified in believing P.

The worry is that the reliability principle challenges the indifference method just as much as it challenges inductive methods. What is missing is some non-circular reason to think that the indifference method is reliable. Moreover, any such justification must appeal to some feature of the indifference method that is relevantly different from parallel features of induction methods. We cannot simply point to the excellent track record of beliefs formed on the basis of the indifference method. After all, induction methods too have an excellent track record. Unless we can identify some way of justifying the reliability of the indifference method, the rejoinder to the skeptic's argument is incomplete.[6]

3. Inductive Methods.

Before concluding, I want to develop one more point: it is difficult to evaluate White's argument without a more precise sense of what counts as an induction method. In support of this point, I will show that if certain belief-forming methods are allowed to count as inductive, then we can run the argument for (J1) without appealing to the indifference method. We might then replace (W3) with this stronger claim and run the argument for (J1) as a straightforward instance of constructive dilemma.

Consider White's discussion of sequences of binary experiences. The discussion aims to show that sequences on which inductive methods are unreliable are *rare*. To support this claim, White considers a handful of such sequences. The most threatening sequence ends up being:

> (*) 1111100001111100001111100001111100000…

However, as White himself points out, this is arguably not an induction-defeating sequence. After a couple mistakes, a sophisticated inductive reasoner will latch on the fact that she was projecting the wrong regularity, and will start predicting the right regularity.

---

6 An interesting option here would be to explore the possibility that we might have default justification for the outputs of the indifference methods even if, following White's argument, we do not have default justification for the consequences of inductive methods.

So, the question is: if we had a realistic induction method, would it be always possible to construct sequences that show the method to be unreliable? To answer this question, we must be more precise about counts as an induction method. In the case of sequences of 1's and 0's, a minimal requirement would be that an induction method be a function that inputs a sequence up to a particular point and outputs a prediction (1, 0 or N, for "no prediction") for the next element of the sequence. Here is a very rough induction method:

> M_1: consider the last 5 elements of the sequence (for the first 5 elements output N). If the proportion of 1's exceeds some threshold (say, 90%), predict 1 (similarly for 0's); otherwise predict nothing.

This algorithm is unreliable when applied to sequence (*): in this sequence M_1 only makes a prediction after five 1's, and every such prediction is wrong. After five 1's comes a 0. However, for the reasons we noted, M_1 is hardly a plausible induction algorithm. A better one would encourage our reasoner, Theo, to look for arbitrary discernible patterns to project. If no pattern is found, he should not form a belief (one problem with doing this is that it's difficult to say exactly what counts as a "discernible pattern": set this issue aside).

Here is the bad news: whatever fancy induction method Theo may adopt, there must be sequences on which induction is not reliable.[7] A well-trained demon will be able to take a method *M* and produce a sequence of observations *s* that makes *M* not reliable. One can even define *s* as a function of *M*. Let *s(x)* be the *x*-th coordinate of *s*. Let *s|x* be the subsequence of *s* consisting of its first *x* entries (*s|0* is the empty sequence). Given any finite sequence *s|x,* let *M(s|x)* denote *M*'s prediction about *x+1*. To create a M-defeating sequence, simply say that $s(x)=|1-M(s|x-1)|$, *if* $M(s|x-1)=0$ or 1; 1 otherwise. In plain English, if *M* makes a prediction about the *x+1*-th step of the sequence, flip it; otherwise predict 1. If Theo has to commit to a single induction method *M* at the beginning of his observational history, then *s* defined as above will make him wrong every time he makes a prediction. If the demon plays her hand wisely and arranges Theo's experience as described in constructing *s*, *M* cannot be reliable.

But here is the good news: it does not yet follow that for every induction method, there is a sequence on which that method is *un*reliable. In fact, I think there are methods that are guaranteed not to be unreliable. Here is one such method, call it M_2: make a handful of predictions at the beginning of your observational history (you can make these predictions on the basis of whatever first-order evidence you have available). If you're ahead by 2 or more, keep on making predictions as long the difference between true and false beliefs is +2. You can keep on making predictions even if you are not ahead by 2 as long as one more mistake would not make you unreliable. If you are not ahead by 2 and you have made enough predictions that another mistaken prediction would make you unreliable, stop guessing. Informally, the method comes down to the following: if you were lucky enough to get ahead at the beginning of your observational history by following your first-order evidence, keep on doing that. Otherwise, stop making predictions.

---

[7] At the Episteme conference, I expressed some hope that this might be false, but the argument in this paragraph and a conversation with Selim Berker convinced me otherwise.

A reasoner who follows this method can reach one of two outcomes. Either she accrues enough advantage that she can keep on forming inductive beliefs. In that case, the method will be reliable. Or she will stop forming beliefs. This reasoner cannot be unreliable, although sometimes she won't be reliable either. As it turns out, in many of the inductive tasks we perform (e.g. predicting sun-risings), we have accrued enough advantage that they can keep on inferring. When we use induction, we are more often than not in the first category. If all of this is correct, then for users of M_2 we *can* strengthen premise (P3) to the claim that that it is certain that *E* will be either regular or unexcitingly irregular and without appealing to the indifference method.

Of course, I am not suggesting that such a method defeats skepticism about induction—or even the specific skeptical challenge that White entertains. Nonetheless, I think that a disjunctive conclusion is warranted. Either (i) there is, after all, a strengthening of White's argument that does not have to go through an appeal to the indifference method or (ii) there is a principled reason why M_2 does not count as an induction method or (iii) M_2 does count as an induction method but is so far from our actual inductive practices that we should not give it serious consideration (or (iv) some other element of the setup I have ascribed to White is problematic). Any choice at this point requires an argument.

[**Acknowledgments:** I am grateful to Mike Caie, Roger White and the participants to the 10th Episteme Conference for discussion. I also thank Jennifer Lackey for organizing the conference]