It's No Illusion!

Alvin C. Plantinga

Section A of chapter 7, "Evolutionary Naturalism: Epistemically Unseated or Illusorily Defeated?" — Plantinga's reply to William J. Talbott's "More on the Illusion of Defeat" (ch. 6).

7

EVOLUTIONARY NATURALISM:
EPISTEMICALLY UNSEATED OR ILLUSORILY DEFEATED?

ALVIN C. PLANTINGA AND WILLIAM J. TALBOTT

A. It's No Illusion!

Alvin C. Plantinga

First, thanks to Bill Talbott for his characteristically acute and insightful contribution. I learned something significant from his original comments on the EAAN ("The Illusion of Defeat"); now he generously offers "More on the Illusion of Defeat" (hereafter "MID"). I say it's no illusion.

Talbott begins by conceding, if only for purposes of argument, that

(1) It is rational for the naturalist to believe P(R/N&E) is low.

What I argued, of course, is not merely this, but that in fact P(R/N&E) *is* low—though naturally I do also believe that it is rational for the naturalist to believe this. What Talbott denies is the second step of the EAAN, namely that

(2) The naturalist who sees (believes) that P(R/N&E) is low has a defeater for R in her own case.

I argued for (2) by analogy, giving several relevantly similar cases where, as I saw it, the analogue of the naturalist does indeed get a defeater for R. Talbott categorizes these (together with some others he presents in his original, 2002 article) in a useful way and then adds another he thinks maximally similar in relevant ways to the EAAN situation. This crucial case is an extension of my emendation of his "first example of the tardy revelation" (I know, I know, this is starting to sound like the sort of thing that gives analytic philosophy a bad name).

My emendation of that first example of the tardy revelation went as follows. At *t*, I come to believe I've taken XX,[1] a drug that (globally) destroys cognitive reliability in 95 percent of those who take it, compromising memory, perception, reasoning, and the rest. At *t* I also believe 5

percent of the population has a blocking gene that renders XX harmless. Presumably, at *t* I have a *defeater* for R (a purely alethic defeater—see MID, 154). At *t+n* it seems to me that I receive a call from my physician, telling me that I am one of the lucky 5 percent that have that blocking gene. And the question is this: what would the proper function of my cognitive faculties and processes—my *truth-aimed* cognitive faculties and processes—require in this situation? I thought they would continue to require that I refrain from believing R. That this is correct, it seems to me, is easily seen if we reflect on what we should think in the analogous third-person case. We learn that Sam has taken XX; we also learn that after he took it, it seemed to him that he received such a call from his physician; surely proper function requires continuing to withhold the belief that Sam's faculties are reliable.

Talbott thinks I am wrong in this evaluation, but he doesn't explain why. What he does instead is to elaborate on the example (the "augmented XX example"). For the details of the elaboration, see MID. Fundamentally, the idea is this: Each of twenty scientists takes XX, a drug they believe destroys cognitive reliability; each believes that the other nineteen have also taken it; each also believes that one of the twenty has the blocking gene, but doesn't know which; and each also believes that the experiences and memories of all twenty will be coherent after taking the drug.

Under these circumstances, says Talbott, each of the scientists has an easy positive test for finding out that she has the blocking drug: noting that (a) it seems to her that the other scientists start behaving as if they are perceiving pink elephants, and as if they have lost all knowledge of chemical theory, but (b) that she herself has no experiences as of pink elephants, and also believes that she is a scientist investigating the effects of XX; it also seems to her that she has the relevant chemical knowledge; and she wonders whether she has the blocking gene. In a word (or two) it seems to her that her memories and current beliefs and experience are coherent. As Talbott puts it, "the issue is precisely whether or not the continuity and coherence of Helen's memories and experience, including the experience of her doctor calling to tell her that she has the drug blocking gene, does provide her with a positive test that determines that she does have the gene" (MID, 158). "That" he says, "is the issue that needs to be discussed, not stipulated away."

Right; that is certainly an issue that needs to be discussed, and far be it from me to stipulate it away. I'd like to discuss this issue first from a relatively abstract point of view, and then in terms of the specific example Talbott proposes. Abstractly, then, Talbott's idea seems to be this: Suppose I come to believe something (something relevant) with respect to which it is unlikely that my cognitive faculties are reliable. Perhaps I believe that I've taken XX and that 95 percent of those who do suffer massive cognitive distortions; or perhaps I believe N&E and then come to see that P(R/N&E) is low. This may initially give me a defeater for R. But I can still acquire a defeater-defeater for that defeater: I can note that my experience and memories are coherent. If I do note

this, then I no longer have a defeater for R. More exactly, perhaps the idea is that if I believe this all along, so to speak, I don't get a defeater for R in the first place; my belief that my memories and experience are coherent is a "defeater-deflector."[2]

This seems to me mistaken. First, the notion of coherence; here there are several questions. (1) Talbott speaks repeatedly of coherence of my *experience* with my memories (and presumably other beliefs). Coherence is not an easy notion,[3] and much about it is unclear; it is particularly unclear, however, that or how *experience* could be coherent with *beliefs.* I take it experience is here being thought of, roughly, as or as similar to *ways of being appeared to:* but how could my being appeared to in such and such a way be coherent (or incoherent) with a belief or other proposition? Of course *the proposition* that I am being appeared to in such and such a way can be coherent with other propositions; so let's suppose this is how to understand Talbott here. Coherence requires that there be no conflict between my belief that I am being appeared to in such and such a way and the rest of my beliefs.

There is more to coherence of experience with belief. What more? Well, (2) experience often *inclines us to believe* something or other; on the basis of experience, it often seems that things are a certain way. Thus, on the basis of my present sensory experience it seems to me that I'm seated in front of a computer, that birds are singing outside my door, that the trees have fully leafed out, etc.; and this inclines me to believe those things. Sometimes experience and belief can be at odds, as when confronted with what one knows is an illusion of some sort. I know those two lines in the Muller-Lyre illusion are the same length; but one *seems* longer than the other. I cross my index and middle finger and put a marble between them; it seems as if (feels as if) there are two marbles; but I believe there is only one. Each of the propositions involved in the deduction of *non-self-exemplification both does and does not exemplify itself* seems true; but I believe at least one of them must be false. Accordingly, there can be incoherence, not just between a belief and the proposition that I am appeared to thus and so, but also between a belief and the proposition experience inclines me to believe. So coherence would also require, not that there be no conflict at all between belief and how things seem (the propositions experience inclines me to accept), but that there be minimal conflict—no doubt a weasely thing to say here, but perhaps the best that we can do.

(3) Coherence also requires an appropriate relationship among the propositions I believe. What relation? That's not an easy question. Like most of us, I am such that there is no possible world in which all of my beliefs are true, and quite properly so. That is because I believe that I believe at least one false proposition. So let B be the set of my first-order beliefs—beliefs that are not about my beliefs: I believe that B contains at least one false belief. But then there is no possible world in which all of my beliefs—second-order as well as first-order—are true. What's required here isn't

logical consistency but something else extremely hard to state—but including, presumably, believing no explicit contradictions (propositions of the form P&-P), and perhaps believing no proposition P and also believing its denial, -P. Perhaps it's also required that I believe no obviously inconsistent triad of propositions, and maybe the same for obviously inconsistent quartets—beyond that, it's hard to say.

Clearly, this isn't sufficient for coherence in the relevant sense. Also required would be (4) a certain hard to characterize regularity in my experience, and a certain consonance between experience and belief. Perhaps this is required just for the coherence of my experience in itself; but at any rate it is surely required for the coherence of my experience with the beliefs I have about how the world ordinarily goes. This regularity would be violated if, e.g., at one moment it seems for all the world as if there is a house twenty-five feet in front of me, and then at the next moment it seems that I haven't moved, but now there is instead a two-acre lake, and then at the next moment a small mountain, and then a flat jungle, and then an opera house, etc. Or if at one moment it seems that my left arm is about the usual two and a half feet long, at the next moment it seems that it's fifty feet long, at the next an inch and a half, etc. Or if at one moment Helen seems to be five years old, but at the next sixty-five years old, or if I believe she is a scientist working on XX but she acts like she thinks she's an elephant trainer, or if my son Harry suddenly seems to turn into a small horse. In these cases there would be no logical inconsistency, but still a lack of coherence in some obvious sense.

So it's monumentally difficult to say just what coherence is, in the relevant sense, the sense in which Talbott no doubt intends it. And the next question: why think my believing that my experience and beliefs are coherent in this sense (call it "T-coherence") is a defeater-defeater or -deflector? Note first that it is obviously possible to be mistaken about whether one's experience and beliefs are T-coherent. Perhaps (and perhaps not) it is impossible to be mistaken about whether one is being appeared to redly; beliefs of that sort are conceivably incorrigible.[4] But the same certainly does not go for coherence, as Frege (as well as most any philosopher you pick) learned to his sorrow. I can easily believe, falsely, that some set of my beliefs is coherent. Furthermore, T-coherence involves a kind of continuity in experience; therefore, judgments of coherence involve reliance upon memory. I have to rely on memory to know that, e.g., it's not the case that at one moment it seems there is a house before me, at the next moment a lake, at the next a snowcapped volcano, etc.; and of course memories can be mistaken. So suppose I believe I've taken XX (and that 95 percent of those who do suffer from massive cognitive unreliability); then I can't sensibly take it for granted that my beliefs and experiences are T-coherent, just because it seems to me that they are. Under these conditions I can't take it for granted that there is a horse in front of me, just because it seems to me that I see a horse there; the same goes for my experience's being T-

coherent.[5] The best I can do, along these lines, is to take it for granted that it *seems to me* that my beliefs and experience are T-coherent.

Hence the basic question is this: why think that if my experience and beliefs seem T-coherent, then (probably) my faculties are functioning reliably? What is P(R/XX & STC) (where "XX" denotes the proposition that I've taken XX and nineteen out of twenty people who do so are rendered unreliable, and "STC" denotes the proposition that my beliefs and experience seem T-coherent)? Is this probability high, as it must be if STC can serve as a defeater-defeater or defeater-deflector? I think not. We can see this most clearly by considering a third-person case. Suppose we initially assume R_{Sam}: that Sam's faculties function reliably; we then come to believe XX_{Sam}: that he has ingested some XX and that 95 percent of those who take it (all but that lucky 5 percent who have the blocking gene) suffer from massive cognitive dysfunction. In believing XX_{Sam} we have a defeater for R_{Sam}. Next, suppose we learn STC_{Sam}: that it seems to Sam that his experience and beliefs are T-coherent. Do we then have a defeater-defeater? If we had believed STC_{Sam} all along, would we have had a defeater-deflector? One has only to ask the question to see the answer: clearly not. Ninety-five percent of those who take XX become massively unreliable; Sam has taken XX; $P(R_{Sam})/(STC_{Sam})$ is .05; adding STC_{Sam} to the condition doesn't appreciably change that probability. If so, that proposition is neither a defeater-defeater nor a defeater-deflector.

Now surely the same goes in my own case. P(R/XX) is low, around .05; adding STC to the condition doesn't appreciably raise that probability; hence that belief is neither a defeater-defeater nor a defeater-deflector in my own case any more than in Sam's.

I said we should consider this question both from the more abstract and from the more concrete point of view: so let's turn to Talbott's example. There's no need to repeat the details; what's relevant here is (1) Helen believes the experimental set-up is as Talbott says it is, (2) she has no experiences as of pink elephants, (3) it seems to her that the others are behaving as if they have lost their knowledge of chemistry and believe that they are training pink elephants and (4) it seems to her that she has not lost her knowledge of chemistry. In a word, her experiences are T-coherent. Our question is: does the fact that her beliefs and experiences seem T-coherent give her a defeater-defeater or -deflector with respect to R?

Consider again the third-person perspective. We initially assume R with respect to Helen; we then learn that she has taken XX, and that 95 percent of those who do become massively unreliable. Thus we acquire a defeater for our initial assumption of R with respect to her. We then learn that it seems to Helen that her beliefs and experience are T-coherent. Does this give us a defeater-defeater—or, if we suppose we learned this at the very beginning of the story—a

defeater-deflector? Surely not. True, she seems to remember the story about the blocking gene, how those without it would act as if they were elephant trainers, etc., while the person with it would have no elephant experiences, etc. But of course she's taken XX; so how can we credit these apparent memories of hers? $P(R_{Helen}/XX_{Helen}) = .05$; we add to the condition XX_{Helen} the proposition that Helen believes the story, and also believes that her present beliefs and experiences are T-coherent. Clearly $P(R_{Helen}/XX_{Helen} \ \& \ TC_{Helen})$ is not significantly greater than $P(R_{Helen}/TC_{Helen})$. We therefore don't have a defeater-deflector or defeater-defeater in the belief that Helen believes the story, and that her present beliefs and experiences are coherent with that story.

Surely the same goes for Helen herself. She believes that she's taken XX and that nineteen out of twenty people who take it develop massive unreliability. This gives her a defeater for R in her own case—that is, this gives her a purely alethic rationality defeater. No doubt proper function would call for her to continue to believe or anyway assume that her faculties are functioning reliably. What else could she do? But, so I say, if only the truth-aimed cognitive faculties were working in her, she would have a proper-function defeater and cease believing or assuming R. And the fact that it seems to her that her beliefs and experience were T-coherent would make no difference.

So here Talbott and I seem to be at loggerheads: I say Helen would have a purely alethic defeater; Talbott says she would not. Is this as far as we can go?

Maybe not. First, note that the third-person perspective gives us a bit of insight into the difference between a proper-function defeater here and a purely alethic defeater. Truth-aimed processes may be compromised or overridden in one's own case. There is this enormously powerful inclination to assume R in one's own case, and of course it is easy to see the utility of this inclination; failure to believe or assume R can make a shambles of one's entire noetic structure. Of course, one doesn't have nearly as strong an inclination to believe R in the case of someone else. So a sensible way to proceed here is to consider the analogous third-person situation, as I did above. It is clear, I think, that in the third-person analogues to the case in question, one doesn't get a defeater-defeater or defeater-deflector for the defeater for R.

A second consideration: Talbott complains that "one curious feature of Plantinga's position on purely alethic rationality defeaters is that he says almost nothing about how best to design cognitive faculties that best realize the goal of truth or reliability" (158); he goes on to say that "The question is whether Helen's reasoning would be recommended as good alethic design" (158); and he adds later that "In any case, my suggestion is that, of the alternatives under consideration, the best alethic design is one that endorses reasoning that one's cognitive faculties are reliable in the XX examples and does not endorse maintaining beliefs about the external world and reasoning

that the cognitive faculties producing those beliefs are reliable in the malicious demon and brain in the vat examples" (161). He goes on to make interesting observations about good cognitive design.

The question," he says, "is whether Helen's reasoning would be recommended as good alethic design": but why is *that* the question? Our question is whether the devotee of N&E who knows or believes that P(R/N&E) is low has an alethic defeater for R; but *that* question has to do with what *our* truth-aimed faculties or belief-producing processes require. It is certainly possible that there be better cognitive designs than ours (perhaps enjoyed by angels); and no doubt God could design a cognitive system that always and automatically came up with only true beliefs across a wide variety of topics; still, our question is really about the truth-aimed portion of *our own* cognitive design. Talbott's thought is that a really good cognitive design plan would be such that in XX type situations, one's experience and beliefs seeming coherent would furnish a defeater-defeater or defeater-deflector. I'm inclined to doubt that; in any event our design plan doesn't enjoy that feature. We can easily see this as follows. Suppose I believe that nineteen out of twenty who take XX suffer massive cognitive distortion and are also such that they believe their beliefs and experience are T-coherent. I then come to believe I've taken XX; I am therefore in an XX situation. But now add that my experience and beliefs seem to me to be T-coherent. Don't I nevertheless still have an alethic defeater for R? As I would in the corresponding third-person case?

Be that as it may (and no doubt it will), Talbott then produces still another example, one he thinks most like the actual case involving the EAAN: the ZZ example. Here BT is told (and presumably believes) a story about ZZ, a substance producing global unreliability, and a cure. According to the story, he was exposed to ZZ but also (along with all the other children in the world) given the cure. Talbott suggests that BT doesn't, under those conditions, have a purely alethic defeater for R. I'm inclined to agree. Certainly, in that circumstance, I would continue to believe that my cognitive faculties are reliable, and I can't see that I get a defeater in being told (and believing) the story. It is as if I learn in one breath both that there was once a very high probability that my faculties would become massively unreliable, and that this eventuality was forestalled by my receiving the cure. Under those conditions I don't get a defeater for R. In learning these two things at the same time, perhaps I have a defeater-deflector, or at any rate a limiting case of a defeater-deflector—"limiting," in that in the usual cases, the defeater-deflector is a belief I have before I acquire the otherwise defeating belief.

But the ZZ case is not analogous to the EAAN situation. The reason is simple: in the EAAN situation, there is nothing analogous to the bacterium-induced cure in the ZZ story. The believer in N&E doesn't learn of any bacterial cure, or any cure of any other sort; and it doesn't seem likely or perhaps even possible that such an element could sensibly be added to the story. He who accepts

N&E and sees that P(R/N&E) is low has no knowledge of any cure for the looming unreliability. He has nothing but the grim realization that the probability—given what is crucially important here, namely, the origin and provenance of his cognitive faculties—is low that his cognitive faculties are reliable.

One final consideration: perhaps Talbott thinks that what plays the role of the bacterial cure, for the evolutionary naturalist, is just the fact that it seems to her that her faculties are T-coherent; perhaps he thinks that this belief is a defeater-deflector or a defeater-defeater. Perhaps the idea is that while P(R/N&E) is low, P(R/N&E&C) (where C is the proposition that her faculties seem to her to be T-coherent) is high; and perhaps the idea is that one can simply tell, whether reliable or not, whether one's memories and beliefs seem to one to be reliable.

He also says:

If, in addition to accepting the Low Probability Thesis, the evolutionary naturalist believed that his memories and experience would be equally continuous and coherent, regardless of whether his cognitive faculties were globally reliable or whether they were globally unreliable, that would be enough to give him a rationality defeater of both kinds for all (or almost all) of his beliefs, including the belief that his cognitive faculties were reliable. Hume is one of the few naturalists I can think of who would have accepted such an extreme claim (163).

The idea, I think, is that P(C/R) > P(C/-R), which means that C is evidence for R.

First, we must remember that the evolutionary naturalist can't sensibly propose, as a defeater-deflector, that his beliefs and experiences are in fact coherent. That would be just to assume that the faculties involved in the production of the belief that they are coherent are in fact reliable; and that is precisely part of what is in question. But we conceded, at least for purposes of argument, that he can't make a mistake about whether they *seem* to him to be coherent; and presumably they do. So the question is really this: what is P(SC/R) and P(SC/-R)? What is the probability of his faculties seeming coherent, given, respectively, R and -R? Talbott apparently thinks there is a big difference here: the first probability is high, while the second is low. I can't see the slightest reason for thinking this. It's a truism that seriously deluded people often have coherent sets of beliefs; it is even more likely that their beliefs and experience will *seem* to them to be coherent. Just as there is evolutionary advantage in believing R, whether or not it is true, so there may well be evolutionary advantage in being such that one's experience and beliefs seem coherent, whether or not R is true. We'd expect that if Sam were reliable, his experiences and beliefs would seem coherent; we don't suppose that if he were not reliable, they wouldn't seem that way. Maybe they would and maybe they wouldn't; it's hard to say what that probability is; either it's inscrutable or it isn't far from .5. Neither of these will be of use to Talbott.

B. The

Notes

IT’S NO ILLUSION! (Alvin C. Plantinga)

1. Not to be confused with the Mexican beer of the same name.

2. See my “Reply to Beilby’s Cohorts,” in James Beilby, ed., *Naturalism Defeated?* Ithaca, NY: Cornell University Press, 2002, 244.

3. See Laurence BonJour, *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press, 1985, 93ff.

4. A better candidate for incorrigibility might be “I’m being appeared to like *that*.”

5. In dreams it usually seems that one’s experience and beliefs are T-coherent; waking sometimes reveals that they were not.