

# Against "Sensible" Naturalism (2007)

by Alvin Plantinga

---

*In the present work, Alvin Plantinga responds to the worry that  $P(R/N\&E)$ , or the probability that our belief-forming mechanism is reliable given the suppositions of naturalism and evolution, would not be low or inscrutable, but rather very high; for, a species that has adaptively evolved to be very successful at procreating and surviving very likely would have mostly true beliefs.*

[Paul Draper] proposes that my argument can be boiled down to one key premise:

**(1)  $P(R/N\&E)$  is low or inscrutable.**

and two key inferences: from (1) to

**(2) Informed naturalists cannot rationally believe that R is true.**

and from (2) to

**(3) Informed naturalists cannot rationally hold any beliefs at all, including their belief in naturalism.**

Here 'informed' naturalists are naturalists who see (believe) that (1) is true. Draper's fundamental criticism is that (2) doesn't follow from (1), at least if the 'N' in (1) is what he calls "sensible" naturalism. He agrees that (2) does follow from

**(1\*)  $P(R/N\&E)$  is low,**

but claims that (1\*) is false.<sup>[1]</sup> Now Draper agrees that a naturalist who believes (1\*) has a defeater for R (and hence for naturalism); so Draper seems to agree that if (1\*) is true, the argument against naturalism is cogent. Contra Draper, I believe (2) follows from (1); at present, however, I propose to argue in reply that in fact (1\*) is true. ...

P is an NP [i.e., neurophysiological] property that causes C, the property of having a certain proposition Q as content.

We may assume that P is adaptive in that it is a part cause of adaptive behavior.<sup>[4]</sup> But (given no more than sensible naturalism), we have no reason at all to suppose that this content, the proposition Q such that C is the property *having Q as content*, is true. We know that P, the NP property that causes S to have Q as content, is adaptive: but that provides not the slightest reason to think Q is true. (We do not, for example, have any reason to think having P causes S to have Q as content *because* Q is true.) Q might be true, but it might equally well be false; it doesn't matter to the adaptiveness of P.

Possibly some true proposition is that first bit of content; equally possibly, some false proposition is. Further, given just sensible naturalism and E, it is as likely that Q, that first bit of content, be false as that it be true. P is indeed adaptive; it is adaptive by virtue of the fact that it causes adaptive behavior. But (given just E and sensible naturalism) there is no more reason to suppose that content true than to suppose it false. Sensible naturalism doesn't give us any connection between the truth value of Q, the content of that belief-structure S, and the adaptiveness of the behavior caused by S. This property P is selected for, not because it causes the content it does, but because it causes adaptive behavior. S causes adaptive behavior by virtue of its content, all right; but it doesn't cause adaptive behavior by virtue of having the property of having *true* content. There would have to be something special about the situation--something beyond sensible naturalism--if P's being adaptive made it more likely than not that Q is true. Natural selection will ordinarily select for adaptive properties, properties that cause adaptive behavior; but that gives us no reason at all to think Q is in fact true.

What holds for that first bit of content will hold for subsequent bits as well. Take any subsequent belief-structure S\* and the property P\* it has such that having P\* causes S\* to have some proposition Q\* as content: P\* will have been selected for, not because Q\* is *true*, but because P\* causes adaptive behavior in the relevant circumstances. And P\* can cause adaptive behavior whether or not Q\* is true. But then it is not likely that natural selection, in modifying the structures that cause beliefs in the direction of greater adaptiveness, will also modify them in the direction of greater reliability--in the direction, that is, of producing a greater proportion of true beliefs.

What holds for *C. elegans*, naturally enough, will hold for other species as well, including that hypothetical species we've been considering. We can assume that the NP properties P displayed by the beliefs enjoyed by members of that species are adaptive; in accordance with sensible naturalism, we can suppose that these properties cause content properties, properties of the form *has Q as content*. But (given sensible naturalism) it doesn't follow that these content propositions are likely to be true. We are supposing that the relevant NP properties cause content properties: a neural structure's having that NP property causes that neural structure to have a certain content. We are therefore supposing there is something like a causal law linking the possession of NP properties of that sort to the possession of content: all neural structures that have that NP property P also have the property of having such and such a proposition as content. Here sensible naturalism differs from 'sensible theism' (the conjunction of theism with Draper's S); according to sensible theism, God has created us human beings in his image, part of which involves giving us the capacity for knowledge. If so, however, he would have instituted causal laws linking NP properties with content properties in such a way that the beliefs in question would be (given appropriate qualifications) mostly true. Not so for sensible naturalism; it doesn't even give us reason to think that content in any way represents environmental circumstances of the

creature in question. That NP property Q is adaptive; sure enough. No doubt it is adaptive by virtue of causing behavior (in a wide sense of the term) that is adaptive in that creature's environmental circumstances, whether short term or long. That same NP property, furthermore, causes content. But why think that content would be *true*? Indeed, why think it would be in any way connected with the circumstances of the creature in question? The content of these beliefs could be anything at all. Perhaps it's like the way we think things go in our dreams. I dream that I am climbing a steep rock face in Yosemite; I believe that I am climbing that rock face. No doubt it's by virtue of the instantiation of a certain NP property P that I have a belief with that content; and no doubt my having P is adaptive. But it doesn't follow that the belief in question is probably true, or even in any way about my current environmental circumstances.

Natural selection, in modifying content properties in the direction of greater adaptiveness, is therefore not likely to be modifying belief-producing processes in the direction of greater reliability. So consider a belief-structure B with its content Q and content-causing property P; what, given that having that belief is adaptive (and given sensible naturalism), is the probability that Q is a true proposition? Well, since we have no reason to think the adaptivity of P makes the truth of Q likely (given sensible naturalism), Q could be true, but is equally likely to be false. We'd have to estimate the probability that it is true as about the same as the probability that it is false. But then if the creature in question has 1000 probabilistically independent beliefs, the probability that, say,  $\frac{3}{4}$  of them are true (and this would be a modest requirement for reliability) will be very low--less than  $10^{-58}$ . And even if the beliefs in question are maximally dependent, probabilistically speaking,  $P(R/N\&E)$  could not be greater than  $\frac{1}{2}$ --low enough to provide a defeater for R. So on sensible naturalism (and E), the probability of R appears to be very low:  $P(R/N\&E)$  (N being sensible naturalism) specified to these creatures, is low.

This is my argument for thinking that  $P(R/N\&E)$  is low, specified to that hypothetical population, and taking N to be sensible naturalism; of course the same goes for us. Draper, on the other hand, thinks the fact that we have evolved and survived provides strong evidence for R. "More generally," he says, "the long term survival of our species is much more to be expected if our cognitive faculties are reliable than if they are unreliable, and that entails that the long term survival of our species is strong evidence for R." What Draper presumably means is that the probability of the long term survival of our species is much more likely on N&E&R than on N&E&-R. So let's suppose that hypothetical species we've been thinking about has in fact survived for a very long time. Does that give us good reason to think its members have reliable cognitive faculties? That depends on how broadly we conceive 'cognitive faculty.' We might limit the term to belief-producing processes; then if our cognitive faculties are reliable, most of our beliefs will be true. On the other hand, we might use the term more broadly, as indeed is often done, in such a way that, for example, the frog who tracks and captures flies has *cognitive faculties*, whether or not it has *beliefs*. What the frog clearly does have are "indicators," neural structures that receive input from the frog's sense

organs, are correlated with the path of the insect as it flies past, and are connected with the frog's muscles in such a way that it is able to flick out its tongue and capture that unfortunate fly.

But of course indication of this sort does not require belief. In particular, it does not require belief in the obtaining of the state of affairs indicated; indeed it is entirely compatible with belief *inconsistent* with that state of affairs. Fleeing predators, finding food and mates--these things require cognitive devices that in some way track crucial features of the environment, and are appropriately connected with muscles; but they do not require true belief, or even belief at all. The long term survival of organisms of a certain species certainly makes it likely that its members enjoy cognitive devices that are successful in tracking those features of the environment--indicators, as I've been calling them. Indicators, however, need not be or involve beliefs. In the human body there are indicators for blood pressure, saline content, temperature, insulin level, and much else; in these cases neither the blood, nor its owner, nor anyone else in the neighborhood ordinarily holds beliefs on the topic.

The fact that a population of animals has survived is evidence for its having *indicators* of this sort, cognitive features that vary with the environment and enable the creatures in question to respond appropriately to their environment. It doesn't follow, as I say, that these creatures have mostly true beliefs, or even beliefs at all. But suppose we are thinking about that hypothetical population of creatures like us; of course they do have beliefs. Given that they have beliefs, does their survival make it likely (relative to N&E) that these beliefs are mostly true? Does their survival make it likely that their belief-producing processes are reliable? Draper argues that false belief would lead to maladaptive behavior. Why does he think that? Consider Draper in the bathtub with that alligator--or rather, consider some member *m* of that hypothetical population in a bathtub with an alligator. Suppose *m* holds false beliefs, believing at the time in question that the alligator is a mermaid, or even that he's sitting under a tree eating mangoes. Will that adversely affect his fitness? Not just by itself. Not if *m* has indicators and other neural structures that send the right messages to his muscles, messages that cause his muscles to contract in such a way as to bring it about that he hops out of that tub. It's having the right neurophysiology and the right muscular activity that counts. We are supposing that belief content supervenes on neurophysiology; as I argued above, however, we have no reason to think that if the neurophysiology is adaptive, the belief content will consist in true propositions. If belief content supervenes on neurophysiology, there will be causal laws connecting NP properties with belief content; but why suppose these laws are such that if the NP properties are adaptive, the belief content, those propositions, will be *true*? It doesn't matter whether the propositions believed, the content of the belief, are true or false; it doesn't matter whether the causal laws that connect neurophysiology with belief content and behavior associate true content with adaptive action, or false content with such action.

If so, however, false belief doesn't make maladaptive behavior likely, even if the beliefs cause the behavior, and do so by virtue of their content. So think

again about  $m$ , that Draper counterpart in the tub with an alligator. Suppose  $m$  has a certain belief  $B$ .  $B$  has NP properties that cause him (it) to leap out of the tub, thus frustrating the alligator.  $B$  also has NP properties on which its content supervenes.  $B$  causes the behavior it does by virtue of that content: if it hadn't had that content, it would not have caused that behavior. But the content needn't be true; and indeed there is no reason to think it would be true. If it is false content that gets associated by the causal laws with those NP properties, then false content will cause the adaptive behavior; and there is no more reason to think the causal laws will associate true content with those properties, than false content. Hence the probability of maladaptive behavior, given false content, will be no greater than the probability of adaptive behavior. That means, contra Draper, that the long term survival of this hypothetical species is *not* much more probable on their having reliable belief-producing processes than on their having processes that produce mostly false belief.

Why does Draper think or assume that those causal laws would be such as to associate *mostly* true content with adaptive NP properties? Given *theism*, of course, that is what we would expect: according to theism God has created human beings in his image, an important part of which involves our being able to have knowledge. But given *naturalism*, it seems just as likely that the causal laws in question would associate false content with adaptive action. Still more likely, perhaps: truth or falsehood is just irrelevant; sometimes true content gets associated, but just as often false content does.

So why does Draper believe or assume that those causal laws would be such as to associate mostly true content with adaptive NP properties and behavior? Why does he assume that if N&E were true, the relevant causal laws would associate *true* belief with adaptive neurophysiology and behavior? So that if a population has survived, it is likely that it displays adaptive neurophysiology and behavior, and hence also likely that its beliefs, if it has some, are mostly true and its belief-producing processes reliable? If the cognitive faculties of these creatures were in fact reliable, this would be a sensible assumption. But of course in the present context ... we can't sensibly assume that our cognitive faculties *are* reliable. To do so would be to argue, not that  $P(R/N\&E)$  is high, but that  $P(R/N\&E\&R)$  is high. Indeed it is, but it has no bearing whatever on the question whether (1\*) is true.

I therefore conclude that Draper has failed to show any problem ...

## Notes

[1] If, as he says, he thinks  $P(R/N\&E)$  is inscrutable, he shouldn't also claim that (1\*) is false; what he should say, perhaps, is that there is no reason at all to believe it.

[4] The property itself, naturally enough, doesn't cause anything; the relevant cause will be the structure that *has* the property. Following current practice I will ignore this distinction in what follows.