Plantinga's Probability Arguments Against Evolutionary Naturalism

Branden Fitelson and Elliott Sober*

Philosophy Department
University of Wisconsin–Madison

November 20, 1997

Abstract. In Chapter 12 of *Warrant and Proper Function*, Alvin Plantinga constructs two arguments against evolutionary naturalism, which he construes as a conjunction $E\&N$. The hypothesis E says that "human cognitive faculties arose by way of the mechanisms to which contemporary evolutionary thought directs our attention (p. 220)."[1] With respect to proposition N, Plantinga (p. 270) says "it isn't easy to say precisely what naturalism is," but then adds that "crucial to metaphysical naturalism, of course, is the view that there is no such person as the God of traditional theism." Plantinga tries to cast doubt on the conjunction $E\&N$ in two ways. His "preliminary argument" aims to show that the conjunction is probably false, given the fact (R) that our psychological mechanisms for forming beliefs about the world are generally reliable. His "main argument" aims to show that the conjunction $E\&N$ is self-defeating — if you believe $E\&N$, then you should stop believing that conjunction. Plantinga further develops the main argument in his unpublished paper "Naturalism Defeated" (Plantinga 1994). We will try to show that both arguments contain serious errors.

1 The Preliminary Argument

Plantinga constructs his preliminary argument within a Bayesian framework. His goal is to establish that $\Pr(E\&N \mid R)$ — the probability of E and N, given R — is low. To do this, Plantinga uses Bayes' Theorem, which says that this

conditional probability is a function of three other quantities:

$$\Pr(E\&N \mid R) = \frac{\Pr(R \mid E\&N) \cdot \Pr(E\&N)}{\Pr(R)}.$$

Plantinga says you should assign to $\Pr(R)$ a value very close to 1 on the grounds that you believe R (p. 228). He argues that $\Pr(R \mid E\&N)$ is low. Although Plantinga doesn't provide an estimate of the prior probability $\Pr(E\&N)$, he says that it is "comparable" to the prior probability of traditional theism (TT) (p. 229), meaning, we take it, that their values aren't far apart.

This last claim should raise eyebrows, not just among evolutionary naturalists who reject the idea that their theory and traditional theism are on an equal footing before proposition R is taken into account, but also among critics of Bayesianism, who doubt that there is an objective basis for such probability assignments. Plantinga says (p. 220, footnote 7) that his probabilities can be interpreted either "epistemically" or "objectively," but that he prefers the objective interpretation. However, Bayesians have never been able to make sense of the idea that prior probabilities have an objective basis. The siren song of the Principle of Indifference has tempted many to think that hypotheses can be assigned probabilities without the need of empirical evidence, but no consistent version of this principle has ever been articulated. The alternative to which Bayesians typically retreat is to construe probabilities as indicating an agent's subjective degree of belief. The problem with this approach is that it deprives prior probabilities (and the posterior probabilities that depend on them) of probative force. If one agent assigns similar prior probabilities to evolutionary naturalism and to traditional theism, this is entirely consistent with another agent's assigning very unequal probabilities to them, if probabilities merely reflect intensities of belief.

Although Plantinga's Bayesian framework commits him to making sense of the idea that the conjunction $E\&N$ has a prior probability, his argument does not depend on assigning any particular value to this quantity. As Plantinga notes (p. 228), if $\Pr(R) \approx 1$ and $\Pr(R \mid E\&N)$ is low, then $\Pr(E\&N \mid R)$ also is low, no matter what value $\Pr(E\&N)$ happens to have.

1.1 Proposition R

For the sake of clarity, it is worth spelling out proposition R more precisely. What does it mean for our psychological mechanisms for forming beliefs to be "generally" reliable? In his unpublished manuscript, Plantinga says that R means that "the great bulk" of our beliefs are true (Plantinga 1994, p. 2). Aside from questions about how beliefs are to be counted, we don't want to challenge the truth of this summary statement. However, it drastically underspecifies the data that need to be explained. For the fact of the matter is that our cognitive mechanisms are reliable on some subjects, unreliable on others, and of unknown reliability on still others. We should divide our beliefs into categories

[1]All page references are to *Warrant and Proper Function* (*viz.*, Plantinga 1993), unless otherwise noted.

and associate a characteristic degree of reliability with each of them.[2] Perhaps certain simple perceptual beliefs are very reliable, while beliefs about other subjects are less so. Rather than trying to obtain a summary statement about all these mechanisms and the beliefs they generate, it would be better to consider a conjunction $R_1 \& R_2 \& \ldots \& R_n$, which specifies the degree of reliability that human belief formation devices have with respect to different subject matters, or in different problem situations. Plantinga (pp. 216–217, 227, 231–232, and in a personal communication) does not object to this partitioning and uses it himself to discuss the probability that $E\&N$ confers on R.

If R is true, why should one bother to spell it out in more detail? This wouldn't matter if Plantinga's argument were deductive. A sound argument stays sound when the premises are supplemented with more (true) details. However, probability arguments don't have this property. Even if $\Pr(R \mid E\&N)$ is less than $\Pr(R \mid TT)$, it remains to be seen whether $\Pr(R_1 \& \ldots \& R_n | E\&N)$ is less than $\Pr(R_1 \& \ldots \& R_n \mid TT)$.

Before we get to that comparative question, let's consider whether the conditional probability $\Pr(R_1 \& R_2 \& \ldots \& R_n \mid E\&N)$ is high or low. Suppose that evolutionary naturalism does a good job of predicting each of the conjunct R_i's, conferring on each a probability, say, of 0.99. It still could turn out that $E\&N$ confers a low probability on the conjunction. If

$$\Pr(R_1 \& R_2 \& \ldots \& R_n \mid E\&N) = \Pr(R_1 \mid E\&N) \cdot \Pr(R_2 \mid E\&N) \cdots \Pr(R_n \mid E\&N)$$

(*i.e.*, if the R_i's are probabilistically independent of each other, conditional on $E\&N$), then the left-hand term may have a low value, even though each product term on the right has a high value. Multiply 0.99 times itself sufficiently often and you get a number close to zero. This can happen to any good theory; it may confer a low probability on a massive conjunction of observations even though it confers a very high probability on each conjunct.

Once we decompose proposition R into a conjunction of claims, it is far from obvious that evolutionary theory does a worse job of predicting this conjunction than traditional theism does. Plantinga says the traditional theist "believes that God is the premier knower and has created us human beings in his image, an important part of which involves his endowing them with a reflection of his powers as a knower (p. 237)." However, an influential point of view in cognitive science asserts that human reasoning is subject to a variety of biases. It isn't just that people occasionally make mistakes, but that the human reasoning faculty seems to follow heuristics that lead to systematic error (Kahnemann, Tversky, and Slovic 1982). It would be no surprise, from an evolutionary point of view, if human beings had highly reliable devices for forming beliefs about practical issues that affect survival and reproduction, but are rather less gifted

when it comes to matters of philosophy, theology, and theoretical science. Does traditional theology also predict this result? No doubt, a theology can be specified that makes any prediction one wants. However, it is not at all clear that Plantinga's traditional theology does a good job predicting the varying levels of reliability that the human mind exhibits. Plantinga must address the same problem that Paley's design argument faces: Why would an omniscient, omnipotent, and benevolent deity produce organisms who seem to be so manifestly imperfect in the adaptations they exhibit (Sober 1993)?

1.2 Setting $\Pr(R) \approx 1$

We mentioned earlier that Plantinga sets $\Pr(R) \approx 1$ because he believes proposition R. Within the context of Bayesian confirmation theory, assigning the evidence a probability close to unity has a peculiar consequence, as we now will show.

Bayesians define confirmation in terms of probability raising; an observation O confirms a hypothesis H if and only if the posterior probability $\Pr(H \mid O)$ is greater than the prior probability $\Pr(H)$. If we rewrite Bayes' theorem as follows

$$\frac{\Pr(H \mid O)}{\Pr(H)} = \frac{\Pr(O \mid H)}{\Pr(O)},$$

it is clear that O cannot confirm H, if $\Pr(O) = 1$. With this assignment, the right-hand ratio can't be greater than unity, so the left-hand ratio can't either. Plantinga's stipulation that $\Pr(R)$ is *close* to unity doesn't quite insure that R can't confirm a hypothesis H. After all, it is possible that $\Pr(R \mid H)$ should be even closer to unity than $\Pr(R)$ is. Let us say that a hypothesis H is *quasi-deterministic* with respect to R if $\Pr(R \mid H) > \Pr(R) \approx 1$. If evolutionary naturalism isn't quasi-deterministic in this sense, then R can't confirm it, given Plantinga's assignment. Proposition R may leave the probability of $E\&N$ unchanged, or it can lower its probability; there is nowhere to go but down. Unless traditional theism is quasi-deterministic with respect to R, it too cannot be confirmed by proposition R, if $\Pr(R) \approx 1$.

Bayesians like to point out that it is a consequence of Bayes's theorem that an observation is incapable of confirming a hypothesis when the observation is completely unsurprising. However, for most predictions of any interest, a Bayesian agent isn't certain in advance that they will come true; when surprising predictions *do* come true, they provide confirmation. The wet sidewalk (W) confirms the hypothesis that it has been raining. The fact that you believe that the sidewalk is wet shouldn't lead you to assign this evidence a probability of unity. A reasonable assignment of value to $\Pr(W)$ is given by the fact that the sidewalk is rarely wet. The observation is therefore somewhat surprising, and so is capable of confirming the hypothesis. Thus, Plantinga's claim that $\Pr(R)$ is close to unity is very odd; it is crucial to his argument that $\Pr(E\&N \mid R)$ is low. Plantinga needs a better reason for this assignment than the fact that he believes R.

---

[2] An even better strategy would be to associate a characteristic degree of sensitivity with a mental faculty in a given environmental setting. Roughly speaking, sensitivity is a world-to-head relation, measured by the probability that the agent will believe p, conditional on p's being true. In contrast, reliability is a head-to-world relation, measured by the probability that p will be true, conditional on the agent's believing p. Sensitivity tends to be a more stable property of measurement devices than reliability. See Sober (1994, essays 3 and 12) for discussion.

Plantinga's preliminary argument might be replaced by a different argument, one that seeks to establish that $\Pr(TT \mid R) > \Pr(E\&N \mid R)$. The goal now is to compare two posterior probabilities, not to estimate their absolute values. This inequality is true precisely when

$$\Pr(R \mid TT) \cdot \Pr(TT) > \Pr(R \mid E\&N) \cdot \Pr(E\&N).$$

Notice that the value of $\Pr(R)$ is now irrelevant. The argument might begin with the assertion that it is more probable that our psychological mechanisms for forming beliefs are reliable if TT is true than would be the case if $E\&N$ were true. If God exists and intervenes in natural processes to guarantee that human beings end up with reliable cognitive faculties, this makes R more of a sure thing than would be the case if chancy natural processes are the only causes of the mental equipment we possess.[3] (Here we ignore the problems adumbrated in Section 1.1 concerning how R should be spelled out.) If TT and $E\&N$ are assigned the same prior probabilities, it then follows that TT has the higher posterior probability.

The present argument provides a recipe for replacing any nondeterministic theory in the natural sciences. If quantum mechanics predicts that a certain experimental outcome was merely very probable, why not accept instead the theistic hypothesis that this outcome was the inevitable outcome of God's will? Theism can be formulated in such a way that it renders what we observe as probable as you please. Those who feel the need to appeal to God's intervention in the case of human mentality should explain why they do not do so across the board.

---

[3]Here we go along with Plantinga's usage of the term "traditional theism," according to which this doctrine makes different observational predictions than evolutionary naturalism. However, there is room to argue whether this way of seeing matters is the only one that is available from various religious traditions (McMullin 1993). Plantinga understands traditional theism to be the idea, not just that God sets evolutionary processes in motion (where these are understood in terms of the best theories that science now provides), but that he occasionally intervenes in them to insure certain outcomes. The idea that God does the former, but not the latter, confers on proposition R precisely the same probability that evolutionary theory by itself confers on R. This can be seen more clearly by considering the accompanying Figure.
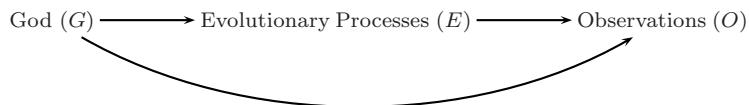

Figure: The possible relationships between G, E, and O.

If evolutionary processes (E) "screen off" God's activity (G) from what we can observe (O), then $\Pr(O \mid E\&G) = \Pr(O \mid E\&\text{not-}G)$. Plantinga thinks this equality is false; he holds that atheistic evolutionism confers on the observations (specifically, on proposition R) a probability different from the one provided by theistic evolutionism. This means that Plantinga is thinking of God as not simply acting *through* natural evolutionary processes, but as affecting the world by a separate, "miraculous" pathway.

1.3 Is $\Pr(R \mid E\&N)$ Low? Rethinking "Darwin's Doubt"

Plantinga (pp. 223–228) argues that $\Pr(R \mid E\&N)$ is low by enumerating several logically conceivable scenarios that describe how beliefs and actions might be related. For each of them, he contends that it is very unlikely that the cognitive mechanisms that evolve should be highly reliable.[4] Here are the possibilities that Plantinga considers: (i) beliefs are not causally connected with behavior; (ii) beliefs don't cause behavior, but are effects of behavior or are effects of events that also cause behavior; (iii) beliefs cause behavior, but do so in virtue of their syntax, not by virtue of their semantics; (iv) the semantic properties of belief cause behavior, but the behaviors are maladaptive; (v) beliefs cause adaptive behavior. In this last category, the adaptive behaviors may be caused by true beliefs, but they also can be caused by false ones. To illustrate this point, Plantinga describes a prehistorical hominid named Paul who manages to avoid being eaten by tigers even though he desires that they should consume him. Paul gets what's good for him by desiring what is bad; he stays out of trouble because his beliefs are false in just the right way. In each of these scenarios, Plantinga says that it is improbable that our cognitive faculties should have evolved to be highly reliable. So $\Pr(R \mid E\&N)$ is low.

In the body of his more recent unpublished manuscript, Plantinga also says that $\Pr(R \mid E\&N)$ is low under scenario (ii) (epiphenomenalism); however, he draws a different conclusion in footnote 15 (p. 8). If beliefs and actions have neural events as common causes, then Plantinga concludes that the probability is "inscrutable," meaning that he can't figure out what value it should be assigned. We take this to be Plantinga's present considered view on the matter. Although Plantinga is developing the main argument against evolutionary naturalism in

---

[4]In discussing "Darwin's doubt," Plantinga (1994, p. 4) quotes with approval a point made by Churchland (1987, p. 548) to the effect that natural selection "cares" only about how adaptive the behaviors are that a set of beliefs causes; it does not care, in addition, whether those beliefs are true. Plantinga interprets this to mean that true beliefs are no more likely to evolve than false ones, but a probabilistic representation of Churchland's point (which is about *conditional independence*) shows that this does not follow. Churchland's point is that

Pr(Belief set B evolves | B produces adaptive behaviors & B is true)
= Pr(Belief set B evolves | B produces adaptive behaviors & B is false).

However, from this it does not follow that

Pr(Belief set B evolves | B is true) = Pr(Belief set B evolves | B is false).

In just the same way, although it is true that

Pr(it will rain tomorrow | a storm is approaching & the barometer reading is low)
= Pr(it will rain tomorrow | a storm is approaching & the barometer reading is high),

it does not follow that

Pr(it will rain tomorrow | the barometer reading is low)
= Pr(it will rain tomorrow | the barometer reading is high).

this manuscript and is not talking about the preliminary argument, it is worth noting that this conclusion undercuts the preliminary argument, which depends on assigning a low value to $\Pr(R \mid E\&N)$.

Whether or not Plantinga's considered view now *is* that $\Pr(R \mid E\&N)$ is inscrutable under scenario (ii), this is what his view *should* be, given the information he considers. Assuming that beliefs don't cause actions is not the same as assuming that they are wholly unrelated. Resistance to malaria doesn't cause anemia, nor does anemia cause malaria resistance; yet, the traits are correlated in a number of human populations because they are phenotypes caused by the same gene. There is no way to tell *a priori* how probable R is under scenario (ii).

When Plantinga turns his attention to category (v) — the case in which beliefs cause adaptive action — he argues that false adaptive beliefs are just as likely to evolve as true adaptive beliefs. The reason is that the behaviors produced by a set of true beliefs also could be produced by a set of false ones. The example of Paul shows one way this could be true. Plantinga describes another in the unpublished manuscript. If "that is a tree" is a true belief, then "that is a witch-tree" is a false belief that would lead to the same behavioral consequences, and so be equally fit. Plantinga's mistake here is that he ignores the fact that the probability of a trait's evolving depends not just on its fitness, but on its *availability*. The reason zebras don't have machine guns with which to repel lion attacks is not that firing machine guns would have been less adaptive than simply running away; the trait didn't evolve because it was not available as a variation on which selection could act ancestrally (see, also, Fodor 1997).

This means that Plantinga's argument that $\Pr(R \mid E\&N)$ will be low in category (v) situations is inadequate. Plantinga might reply that witch-beliefs and other systems of adaptive false beliefs were available ancestrally. However, we don't see any reason to think that this substantive claim about the past can be successfully defended. By ignoring the question of availability, Plantinga, in effect, assumes that natural selection acts on the set of *conceivable* variants. This it does not do; it acts on the set of *actual* variants.

In general, the way to have two (logically independent) properties be well-correlated is to have one cause the other, or to have each trace back to a common cause. If belief and action failed to be causally connected in either of these two ways, then it *would* be surprising for selection on action to lead cognitive mechanisms to evolve that are highly reliable. However, if belief and action are causally connected, then it takes a more detailed argument than Plantinga provides for concluding that reliable belief formation devices are unlikely to evolve via selection on actions. Proposition R *is* improbable under scenario (i), but that's about all one can say.

1.4 The Principle of Total Evidence

Suppose Plantinga is right in saying that $\Pr(E\&N \mid R)$ is low. It does not follow that $E\&N$ is improbable relative to *all* relevant evidence. Evolutionary naturalists can happily accept the idea that the conditional probability just mentioned is low; it does not follow that they should have less confidence than they presently do in the truth of evolutionary theory and naturalism.

If you draw a card at random from a standard deck of cards, the probability is only 1/52 that you will draw the 7 of diamonds. If you do draw this card, that doesn't mean that you should conclude that the deck isn't standard or that the card wasn't drawn at random. If you have independent evidence that the deck is standard and that the draw was random, you simply accept the fact that some of the things that happen don't have high probabilities. Even if it turns out that there are features of human cognitive makeup that are improbable on the hypothesis that human beings evolved, there is lots of evidence that the human mind is a product of evolution. In this light, the sensible thing to do is to accept evolutionary theory and come to terms with the fact that evolutionary processes sometimes have improbable outcomes.[5]

As for the separate hypothesis of naturalism, by which Plantinga means atheism together with some other claims that he does not spell out, these too must be evaluated in the light of *all* the evidence, not just with respect to proposition R.

1.5 A Contradiction and Two Ways Out

We have mentioned that Plantinga thinks $\Pr(R)$ is close to 1, $\Pr(R \mid E\&N)$ is low, $\Pr(R \mid TT)$ is high, and that $\Pr(E\&N)$ and $\Pr(TT)$ are "comparable." Plantinga's preliminary argument also includes the assumption that N and TT are the only two "significant alternatives" (p. 228).

If the claim that $\Pr(E\&N)$ and $\Pr(TT)$ are "comparable" means that their values are not far apart, and if the claim that $E\&N$ and TT are the only two "significant alternatives" means that they are the only possibilities that have non-negligible prior probabilities, then this set of probability claims is contradictory. To see why, let's expand $\Pr(R)$:

$$\Pr(R) = \Pr(R \mid E\&N) \cdot \Pr(E\&N) + \Pr(R \mid TT) \cdot \Pr(TT).$$

If $E\&N$ and TT are exhaustive, then Plantinga's claim that they have "comparable" priors means that they are each close to 0.5. Substituting this and the other values that Plantinga assigns to the component expressions, we obtain

$$1 \approx (\text{low}) \cdot (0.5) + (\text{high}) \cdot (0.5).$$

This is impossible — a contradiction of the axioms of probability.

There are a couple of ways out of this difficulty. One is to retain all the probability assignments, but deny that $E\&N$ and TT exhaust the significant alternatives. If a third possibility, theory X, is countenanced, then $\Pr(R)$ expands to

$$\Pr(R) = \Pr(R \mid E\&N) \cdot \Pr(E\&N) + \Pr(R \mid TT) \cdot \Pr(TT) + \Pr(R \mid X) \cdot \Pr(X).$$

---

[5]The point we are making here accords with what Plantinga (1994) calls "the perspiration objection," which he attributes to "Wykstra, DePaul, and others." Although Plantinga discusses how the objection should be formulated, he does not, as far as we can see, provide an answer to it.

If $\Pr(R) \approx 1$, $\Pr(E\&N)$ and $\Pr(TT)$ are about the same, and $\Pr(R \,|\, E\&N)$ is low, then Plantinga must assign $\Pr(E\&N)$ and $\Pr(TT)$ negligible probabilities, so that $\Pr(X)$ is close to 1. He also must assign $\Pr(R \,|\, X)$ a value close to unity. This revision in Plantinga's argument thus requires the existence of an alternative to traditional theism that is vastly more probable *a priori*, and which entails that proposition R is very probable indeed. The effect of these assignments is to make $\Pr(E\&N \,|\, R)$ and $\Pr(TT \,|\, R)$ low and $\Pr(X \,|\, R)$ high. If a low value for $\Pr(E\&N \,|\, R)$ suffices to reject evolutionary naturalism, then one should reject traditional theism as well. This revision of his argument indicates that the most acceptable alternative is theory X.

Another way to avoid contradiction is to reinterpret what it means for $\Pr(E\&N)$ and $\Pr(TT)$ to be "comparable." Plantinga has suggested to us (personal communication) that this should be taken to mean that the agent doesn't believe that the two theories have very different probabilities. For example, suppose that the agent places $E\&N$ and TT in the same wide interval of probabilities (say, between 0.05 and 0.95) and isn't able to be more specific than this.

Plantinga's various claims can be rendered consistent by this revision. To see why, let's return to the expansion of $\Pr(R)$ and see what happens when we let $\Pr(E\&N)$ and $\Pr(TT)$ be unknown, save for the fact that each must fall between 0.05 and 0.95:

$$\Pr(R) = \Pr(R \,|\, E\&N) \cdot \Pr(E\&N) + \Pr(R \,|\, TT) \cdot \Pr(TT)$$

$$1 \approx (\text{low}) \cdot (?) + (\text{high}) \cdot (?)$$

The insertion of question marks insures that no contradiction arises. However, values for the question mark quantities are not left open; $\Pr(E\&N)$ must be very close to zero and $\Pr(TT)$ must be very close to unity. The argument is now consistent, but is entirely deprived of its probative force. Consistency (and the revised interpretation of what "comparable" means) require the assumption that traditional theism is virtually certain *a priori*, and that evolutionary naturalism is almost certainly false, again *a priori*. Those not already convinced before proposition R is considered that traditional theism is vastly more probable than evolutionary naturalism will reject the argument at the outset. In addition, this revision of Plantinga's preliminary argument undermines its original motivation. Plantinga's thought was to develop what he calls "Darwin's doubt" — that $\Pr(R \,|\, E\&N)$ is low. However, once $\Pr(R)$ is set close to 1, and $\Pr(E\&N)$ is assumed to be small, it *automatically* follows that $\Pr(E\&N \,|\, R)$ is still smaller, no matter what value $\Pr(R \,|\, E\&N)$ happens to have.

2 Plantinga's Main Argument Against $E\&N$

The argument just described is preliminary to the main event, in which Plantinga (pp. 234–235) argues that $E\&N$ is self-defeating. The main argument doesn't aim to show that the conjunction $E\&N$ is probably false (or that $E\&N$ is less probable than TT), but that people shouldn't believe $E\&N$:

1. $\Pr(R \,|\, E\&N)$ is low or its value is inscrutable.
2. Therefore, $E\&N$ is a defeater of R — if you believe $E\&N$, then you should withhold assent from R.
3. If you should withhold assent from R, then you should withhold assent from anything else you believe.
4. If you believe $E\&N$, then you should withhold assent from $E\&N$ ($E\&N$ is self-defeating).

$\therefore$ You should not believe $E\&N$.

Plantinga goes further, in both the book and the manuscript. He argues that naturalism by itself is self-defeating. In the book, he says that "if naturalism is true, then, surely so is evolution (p. 236)." In the manuscript, he says that $\Pr(E \,|\, N)$ is high (p. 11). Neither of these claims is right. Recall that proposition E adverts to the mechanisms described in *contemporary* evolutionary theory. If that theory were found wanting, it would not entail the falsehood of naturalism; naturalists could quite consistently cast about for a better scientific theory. In the manuscript Plantinga makes the point that E is "the only game in town" for a naturalist; this may or may not be true (now), but that hardly shows that naturalism on its own makes 1990's evolutionary theory probable.

2.1 Problems that the Main Argument Inherits from the Preliminary Argument

We have already discussed why we are unconvinced by Plantinga's argument that $\Pr(R \,|\, E\&N)$ is low. We also argued that even if $\Pr(E\&N \,|\, R)$ were low, that would not entail that R suffices to reject $E\&N$. The symmetrical point that pertains to the main argument is that even if $\Pr(R \,|\, E\&N)$ were low, that would not oblige people who believe $E\&N$ to withhold belief from R. After all, people who believe $E\&N$ might have other reasons for believing R. For example, they might argue that R is a basic proposition that does not need theoretical support, or that R derives its epistemic credentials from something other than the thesis of evolutionary naturalism.

The same point holds if you don't know what value to assign to $\Pr(R \,|\, E\&N)$. People who believe $E\&N$ should not regard the fact that this probability is "inscrutable" to them as a reason to reject R. We suspect that many people who are well acquainted with the theory of special relativity and who think that birds fly still don't know what value to assign to $\Pr(\text{Special relativity} \,|\, \text{birds fly})$, especially if probability has to be an objective quantity; however, that doesn't show that they should withhold belief in special relativity. The Principle of Indifference is flawed because it claims to obtain probabilities from ignorance; the start of Plantinga's main argument makes the complementary mistake of holding that ignorance of probabilities is a guide to belief.

In the light of these points, consider the following passage from *Warrant and Proper Function* (p. 229, our italics and brackets) in which Plantinga justifies the first step of the main argument:

Someone who accepts $E\&N$ and also believes that the proper attitude toward $\Pr(R\,|\,E\&N)$ is one of agnosticism [or, one of low degree of belief] clearly enough has good reason for being agnostic about [or, having a low degree of belief with respect to] R as well. *She has no other information about $R\ldots$ but the source of information she does have gives her no reason to believe R and no reason to disbelieve it.*

Notice that Plantinga assumes that evolutionary naturalists have no basis for deciding what to think about R, other than the proposition $E\&N$ itself. This crucial assumption is never defended in either *Warrant and Proper Function* or "Naturalism Defeated."

2.2 What Defeating R Means

In the second step of the main argument, Plantinga says that $E\&N$'s defeat of R means that evolutionary naturalists should withhold assent from anything else they believe — for example, from $E\&N$ itself. This goes beyond what the defeat of proposition R really entails. Proposition R says that "the great bulk" of the beliefs we have are true (Plantinga 1994, p. 2). If evolutionary naturalists should withhold assent from R, this does *not* mean that they should withhold assent from most of what they believe, much less from everything they believe. Even if $E\&N$ defeats the claim that "at least 90% of our beliefs are true," it does not follow that $E\&N$ also defeats the more modest claim that "at least 50% of our beliefs are true." Plantinga must show that $E\&N$ not only defeats R, but also defeats the claim that "at least a non-negligible minority of our beliefs are true."

2.3 Conditional Probability and Defeat

Although we, like a number of other commentators, have interpreted the main argument of *Warrant and Proper Function* as asserting that a low value for $\Pr(R\,|\,E\&N)$ suffices for $E\&N$ to defeat R, Plantinga (1994) denies that this is what he meant and tries to develop an account of defeat that clarifies how the argument is supposed to go. However, Plantinga still spends time arguing in this later manuscript that $\Pr(R\,|\,E\&N)$ is low or inscrutable, so he presumably still holds that the value of this probability is relevant to establishing that $E\&N$ defeats R.

Plantinga develops three principles that he thinks govern the defeat relation. Although he explains why he thinks these principles are correct, he never explains how they are relevant to establishing that $E\&N$ defeats R. In fact, their logical form renders them incapable of closing the gap between premises 1 and 2. The first and third principles assert *sufficient* conditions for X's *not* defeating Y. The second states a *necessary* condition for X's defeating Y.

In addition, Plantinga's "First Principle of Defeat" apparently helps establish that $E\&N$ is *not* a defeater of R. Substituting R and $E\&N$ for A and B in this principle yields

If S rationally believes that the warrant that $E\&N$ has for him is derivative from the warrant that R has for him, then $E\&N$ is not a defeater, for him, of R.

We suspect that many evolutionary naturalists rationally believe that their warrant for believing $E\&N$ depends on their being warranted in believing that their cognitive faculties are highly reliable.[6]

Not only is a low value for $\Pr(X\,|\,Y)$ not *sufficient* for Y's defeating X; it also is not *necessary*, if defeaterhood is to ground the idea of *self*-defeat. The reason is that $\Pr(Y\,|\,Y) = 1$, for all Y. And as difficult as it is to connect low probability to defeaterhood, it seems even harder to see why the inscrutability of $\Pr(X\,|\,Y)$ should help establish that Y defeats X.

In the preliminary argument, Plantinga assigns to $\Pr(R)$ a value close to unity because he believes R to be true. In the main argument as formulated in *Warrant and and Proper Function*, he gives the impression that he thinks evolutionary naturalists should withhold belief in R because $E\&N$ fails to confer a sufficiently high probability on R. These are two ways of expressing the same sentiment: *high probability is necessary for rational belief*. However, the more recent manuscript "Naturalism Defeated" repudiates the idea that there is any such simple relation between probability and acceptance.

What Plantinga is coming up against here is a close relative of the phenomenon that Kyburg's (1961) lottery paradox made vivid. Suppose there are $10,000$ tickets in a fair lottery; one ticket will win and each has the same chance of winning. Suppose you adopt the following criterion for belief — you accept a proposition if you think it has a high probability. If so, you will accept each proposition of the form "ticket i won't win." However, the conjunction of these contradicts the starting assumption that the lottery is fair. Therefore, high probability is not sufficient for rational belief. A similar counterexample can be constructed to show that high probability is also not *necessary* for rational belief. Consider any n propositions $P_1, \ldots, P_n$ such that (i) you accept each of the P_i, and (ii) each of the P_i is very highly probable. The conjunction $P_1\&\cdots\&P_n$ may turn-out to be quite *im*probable (see section 1.1 for a concrete example of this probabilistic phenomenon). Nonetheless, it apparently would be rational for you to accept the conjunction $P_1\&\cdots\&P_n$. Hence, high probability is not necessary for rational belief (see, also, Maher 1993, section 6.2.4). Philosophers of probability have extracted from these paradoxes one of two lessons — either the concepts of acceptance and rejection are suspect, or they are more subtly related to the concept of probability than the threshold criterion just described.

---

[6]Of course, one can't deduce $E\&N$ from R alone. But, evolutionary naturalists might reasonably maintain that R is one of several premises which underwrite their non-deductive inferences concerning the plausibility of $E\&N$. It is worth pointing out that Plantinga himself makes use of this kind of non-deductive "warrant derivation." On page 39 of "Naturalism Defeated," he says, in the context of discussing an objection to his argument, that the warrant that P has for you (where P is such that $\Pr[R\,|\,N\&E\&P]$ is high, but P is *logically independent* of R) is "$\ldots$ derivative from the warrant R has for you $\ldots$ it is hard to see what other source [of warrant] there could be [for P]." We see no reason why it would be irrational for evolutionary naturalists to say the same thing about the warrant that $E\&N$ has for them.

This connection with the lottery paradox suggests that the task of repairing the main argument is formidable. That argument begins with claims about probability, moves to claims about defeat, and then concludes with a claim about self-defeat. Each step along the way requires principles quite different from the ones that Plantinga has so far described. Whether plausible principles exist that forge the requisite connections we leave to the reader to conjecture.

3   A Question for Evolutionism

Although Plantinga's arguments don't work, he has raised a question that needs to be answered by people who believe evolutionary theory and who also believe that this theory says that our cognitive abilities are in various ways imperfect. Evolutionary theory does say that a device that is reliable in the environment in which it evolved may be highly unreliable when used in a novel environment. It is perfectly possible that our mental machinery should work well on simple perceptual tasks, but be much less reliable when applied to theoretical matters. We hasten to add that this is *possible*, not *inevitable*. It may be that the cognitive procedures that work well in one domain *also* work well in another; *Modus Ponens* may be useful for avoiding tigers *and* for doing quantum physics.

Anyhow, if evolutionary theory does say that our ability to theorize about the world is apt to be rather unreliable, how are evolutionists to apply this point to their own theoretical beliefs, including their belief in evolution? One lesson that should be extracted is a certain humility — an admission of fallibility. This will not be news to evolutionists who have absorbed the fact that science in general is a fallible enterprise. Evolutionary theory just provides an important part of the explanation of why our reasoning about theoretical matters is fallible.

Far from showing that evolutionary theory is self-defeating, this consideration should lead those who believe the theory to admit that the best they can do in theorizing is to do the best they can. We are stuck with the cognitive equipment that we have. We should try to be as scrupulous and circumspect about how we use this equipment as we can. When we claim that evolutionary theory is a very well confirmed theory, we are judging this theory by using the fallible cognitive resources we have at our disposal. We can do no other.

Plantinga suggests that evolutionary naturalism is self-defeating, but that traditional theism is not. However, what is true is that neither position has an answer to hyperbolic doubt. Evolutionists have no way to justify the theory they believe other than by critically assessing the evidence that has been amassed and employing rules of inference that seem on reflection to be sound. If someone challenges all the observations and rules of inference that are used in science and in everyday life, demanding that they be justified from the ground up, the challenge cannot be met. A similar problem arises for theists who think that their confidence in the reliability of their own reasoning powers is shored up by the fact that the human mind was designed by a God who is no deceiver. The theist, like the evolutionary naturalist, is unable to construct a non-question-begging argument that refutes global skepticism.

References

Churchland, P. (1987). Epistemology in the Age of Neuroscience. *Journal of Philosophy LXXXIV*, 544–553.

Fodor, J. (1997). Is Science Biologically Possible? Unpublished manuscript of the 1998 Benjamin Lecture at the University of Missouri.

Kahnemann, D., P. Slovic, und A. Tversky (Eds.) (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kyburg, H. (1961). *Probability and the Logic of Rational Belief*. Wesleyan University Press.

Maher, P. (1993). *Betting on Theories*. Cambridge University Press.

McMullin, E. (1993). Evolution and Special Creation. *Zygon 28*, 299–335.

Plantinga, A. (1993). *Warrant and Proper Function*. Oxford University Press.

Plantinga, A. (1994). Naturalism Defeated. Unpublished manuscript.

Sober, E. (1993). *Philosophy of Biology*. Westview Press.

Sober, E. (1994). *From a Biological Point of View*. Cambridge University Press.