

# Alvin Plantinga's Evolutionary Argument

against

**Naturalism**

**By: Arash Naraghi**

## Introduction

After years defending the rationality of theism in general and Christian beliefs in particular, Alvin Plantinga has taken the offensive. He proposes an argument against naturalism, and claims the falsity, or at least, irrationality of naturalism. He made three major claims against naturalism:

*First claim:* The probability of our cognitive faculties being reliable given the conjunction of what he calls "metaphysical or philosophical naturalism" and some version of the theory of evolution is fairly low or inscrutable.( The Probability Thesis- hereafter (PT) )<sup>1</sup>

*Second claim:* Naturalism is false.

*Third claim:* Naturalism is self-defeating, and therefore, it cannot be rationally held.

Here by "naturalism" or "metaphysical/philosophical naturalism", Plantinga means the belief that there are not any supernatural beings- no such person as God, for example, but also no other supernatural beings, and nothing at all like God.<sup>2</sup>

He proposed three major arguments to establish these three claims.

This essay is divided into two parts. In the first part, Plantinga's arguments for these three major claims will be presented. And then in the second part, some important objections to these arguments will be discussed.

Among Plantinga's claims, the first one in itself is not against naturalism, but as we will see, it is a crucial premise for any of Plantinga's arguments for the second and third claims.

The two other claims are directly against naturalism, but recently Plantinga himself admitted

that his attempts to establish the second claim were not very successful. So the third claim seems to be his main objection to naturalism. Therefore, here I will focus mainly on the first, and especially the third claims.

## **Part One: Arguments**

### **I.**

#### **The Probability Thesis**

Plantinga's argument begins with an articulation of doubts about the *reliability* of our cognitive faculties (namely our memory, perception, and reason), given metaphysical naturalism and the typical mechanisms of evolution (such as natural selection, genetic drift, etc.). According to Plantinga a cognitive faculty is reliable if "the great bulk of its deliverances are true."<sup>3</sup>

Plantinga's first main claim can be expressed as follows:

(PT):  $\Pr(R/N\&E)$  is low or inscrutable.

where 'R' is the proposition that our cognitive faculties are reliable, 'N' the proposition that naturalism is true, and 'E' the proposition that we have evolved according to the suggestions of contemporary evolutionary theory.

But why should one believe PT?

Plantinga follows evolutionists<sup>4</sup> who believe that natural selection is directly interested only in behavior, not belief, and that it is interested in belief, if at all, only indirectly, by virtue of the relation between behavior and belief. If adaptive behavior guarantees or makes probable reliable faculties, then perhaps  $\Pr(R/N\&E)$  will be fairly high. On the other hand, if having reliable faculties *isn't* guaranteed by or even particularly probable with respect to adaptive behavior, then presumably  $\Pr(R/N\&E)$  will be rather low.

Therefore, according to this view, the nature of the relationship between belief and behavior is integral to determining  $\Pr(R/N\&E)$ .

Plantinga produces four "mutually exclusive and jointly exhaustive" accounts of the relationship between belief and behavior.<sup>5</sup>

But before that, he suggests that we think, not about ourselves, and our behavior, but about a hypothetical population a lot like us. These creatures are *rational*, that is, they form beliefs, reason, change beliefs, and the like; and also their cognitive faculties evolved by way of the mechanisms to which contemporary evolutionary theory directs our attention. Now we estimate  $\Pr(R/N\&E)$ , specified to them, not to us.

There are four possible relationships between belief and behavior, according to Plantinga:

(A<sub>1</sub>) Beliefs do not causally influence behavior.(Epiphenomenalism)

(A<sub>2</sub>) Beliefs causally influence behavior syntactically, but not semantically. (And here by "semantic" he means the *content* of belief, and by "syntax" he means some neurophysiological or electrochemical properties that a belief as a , for example, long-term pattern of neural activity might have.) Plantinga calls this account "*semantic epiphenomenalism*".

(A<sub>3</sub>) Beliefs causally influence behavior semantically as well as syntactically, but are maladaptive.

(A<sub>4</sub>) Beliefs causally influence behavior semantically as well as syntactically, and are adaptive.

Now  $\Pr(R/N\&E)$  will be the weighted average of  $\Pr(R/N\&E\&A_i)$  for each of these four possibilities  $A_i$ - weighted by the probabilities , on  $N\&E$ , of those possibilities. It can be calculated according to the following formula:

$$(F_1)\Pr(R/N\&E)=[\Pr(R/N\&E\&A_1)\times\Pr(A_1/N\&E)]+[\Pr(R/N\&E\&A_2)\times\Pr(A_2/N\&E)]+ \\ [\Pr(R/N\&E\&A_3)\times\Pr(A_3/N\&E)]+[\Pr(R/N\&E\&A_4)\times\Pr(A_4/N\&E)]$$

To simplify this formula, note that :

(1) Among these four alternative options,  $A_3$  seems very unlikely, and so it can be ignored safely.

(2)  $A_1$  and  $A_2$  unite in declaring or implying that the *content* of belief lacks causal efficacy with respect to behavior. So these two possibilities can be reduced to one: the possibility that the content of belief has no causal efficacy. Let's call it '-C'. Now we have two major possibilities  $A_4$  (for the sake of uniformity, let's call  $A_4$ , hereafter 'C') and '-C'.

Now we can reduce  $(F_1)$  to the following form:

$$(F_2) \Pr(R/N\&E) = \Pr(R/N\&E\&C) \times \Pr(C/N\&E) + \Pr(R/N\&E\&-C) \times \Pr(-C/N\&E)$$

Here Plantinga claims that:

(a)  $\Pr(R/N\&E\&C)$  *prima facie* has a fairly high probability, but in fact it is not as high as one might think. For, he believes that "beliefs do not causally produce behavior *by themselves*; it is beliefs, desires, and other things that do so together."<sup>6</sup> Assuming this fact, Plantinga offers two scenarios in which the basic cognitive systems could be adaptive while generating falsehood.

*First scenario:* According to Plantinga, natural selection can generate suitable desires and goals that, when conjoined with appropriate false belief, give rise to fitness-enhancing behavior. For example, consider Paul, a prehistoric hominid who *likes* the idea of being eaten by a tiger, but when he sees a tiger, always runs off looking for a better prospect, because he thinks it unlikely that the tiger he sees will eat him. This will get his body parts in the right place so far as survival is concerned, without involving much by way of true belief.<sup>7</sup>

But here one might say no one denies that there could exist false beliefs that, when coupled with certain desires, would produce adaptive behavior in a certain way. But the critical question at this point is whether or not a process of natural selection is likely to generate cognitive mechanisms that are *systematically* unreliable but somehow prove adaptive. It is easy to see, for just *one* of Paul's actions, that there are many different belief-

desire combinations that yield it; but it is less easy to see how it could be that most of all his beliefs could be false but nonetheless adaptive or fitness-enhancing. In other words, could it be the case that one holds a mostly false belief system, and nonetheless survives over time?

Here Plantinga offers his second scenario.

*Second scenario:* In the second scenario, Plantinga supposes that the organism suffers from some sort of profound misconception that undermines the veracity of its beliefs without undermining the belief's usefulness. He says: "Perhaps Paul is a sort of early Leibnizian and thinks everything is conscious (and suppose that is false); furthermore, his ways of referring to things all involve definite descriptions that entail consciousness, so that all of his beliefs are of the form *That so-and-so conscious being is such-and-such*. Perhaps he thinks all the plants and animals in his vicinity are witches, and his ways of referring to them all involve definite descriptions entailing witchhood. But this would be entirely compatible with his belief's being adaptive; so it is clear, I think, that there would be many ways in which Paul's beliefs could be for the most part false, but adaptive nonetheless."<sup>8</sup>

Even though this scenario is quite conceivable, but there is no reason to think it was an actual or available option on which natural selection can act. Here Plantinga seems to ignore 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", but it is because that option was not available as a variation on which selection could act ancestrally.<sup>9</sup>

But can we find some actual cases in which (at least from the perspective of a naturalist) we have a *system* of mainly false but adaptive beliefs? Plantinga offers a third scenario to show an *actual* instance of such system (specified to the naturalist's way of looking at the world):

*Third scenario:* Suppose naturalism is true and in fact there is no such person as God. But some naturalists believe that belief in God, while false, is nonetheless adaptive. Now suppose a tribe of cognitively gifted creatures believe that everything (except God Himself) has been created by God. Suppose therefore that their only way of referring to the various things in their environment is by way of such definite descriptions as 'the tree creature before me' or 'the tiger creature approaching me'. Suppose still further that all their beliefs are properly expressed by singular sentences whose subjects are definite descriptions expressing properties that entails the property of creaturehood- such sentences as 'The tiger creature approaching me is dangerous' or 'The tree creature before me is full of apple creatures'. Suppose, finally, that their definite descriptions work the way Bertrand Russell thought definite descriptions work. 'The tallest man in Boston is wise', for example, abbreviates 'There is exactly one tallest man in Boston, and it is wise'. Then from the naturalist perspective all their beliefs are false. Yet these can still be adaptive: all they have to do is ascribe the right properties to the right 'creatures'.<sup>10</sup>

To sum up Plantinga's main point on  $\Pr(R/N\&E\&C)$ , we can say, since there are indefinitely many belief-desire systems that would yield adaptive behavior, but are unreliable, one does not quite know what to say about  $\Pr(R/N\&E\&C)$ . But if one thinks this probability is high, then perhaps it would be reasonable to estimate it as somewhat more than 1/2.

(b) Plantinga also argues that  $\Pr(-C/N\&E)$  is high. His argument for this claim can be stated as follows:

- (1) One who accepts metaphysical naturalism will be a materialist with respect to human beings.
- (2) From the perspective of a materialist  $-C$  is hard to avoid.
- (3) Therefore, the probability that  $-C$  is the case, given  $N\&E$ , is high.

The reason Plantinga thinks (2) is true, is that if there *are* such things as beliefs, then, according to materialism, they will have to be material processes or events- maybe something like long-standing neural events of some kind in the nervous system. But a belief has to be the belief that P for some proposition P. In other words, P is the content of that belief. Now he believes that "it is far from clear how a neural structure can acquire a content....Granted that such a structure can acquire content, however, it still seems extremely difficult to see how it can enter the causal chain leading to behavior *by virtue of its content*."<sup>11</sup>

(C) Now Plantinga claims that  $\Pr(R/N\&E\&-C)$  is low.

According to the evolutionary theory natural selection can eliminate certain structures (those that come at the cost of reproductive fitness) and encourage others (those that enhance reproductive fitness). Now if false belief caused maladaptive action, therefore, natural selection could presumably modify belief-producing structures in the direction of greater reliability i.e., a greater proportion of true as opposed to false belief. But if content does not enter the causal chain that leads to behavior, then of course it will not be the case that a belief causes maladaptive behavior by virtue of its being false, and it will not be the case that a true belief causes the behavior it does by virtue of being true. And then it is hard to see how natural selection can promote or enhance or reward true belief and penalize false belief. As Plantinga puts it, in this situation the hypothetical population's beliefs would be "*invisible to evolution*".<sup>12</sup>

Now we are in position to estimate  $\Pr(R/N\&E)$  according to (F2):

$\Pr(-C/N\&E)$  is high. Suppose it is 0.8. Then  $\Pr(C/N\&E)$  will be 0.2.  $\Pr(R/N\&E\&-C)$  is low. Suppose it is around 0.2. And  $\Pr(R/N\&E\&C)$  is high, but not as high as one might think. Suppose it is 0.7. Then  $\Pr(R/N\&E)$  will be less than 1/2 (under these assignments, 0.3).

So far, what Plantinga shows (if any) is that  $\Pr(R/N\&E)$ , *specified to the hypothetical population*, is low. Now at this stage, Plantinga quickly derives the same conclusion about

"us" i.e., human beings. He says: "The reasoning that applies to these hypothetical creatures, of course applies to *us*; so if we think the probability of R with respect to *them* is relatively low on N&E[...], we should think the same thing about the probability of R with respect to *us*."<sup>13</sup> This is because he believes we are relevantly like them in that *our* cognitive faculties have the same kind of origin and provenance as *theirs* are hypothesized to have.<sup>14</sup>

Of course, Plantinga confesses that we cannot assign specific real numbers to the probabilities involved in estimating  $\Pr(R/N\&E)$ ; the best we can do are vague estimates. Therefore, his estimates of the various probabilities involved in (F2) is "imprecise and also poorly grounded" and so "the right course here is simple agnosticism: one just does not know what that probability is. ... Then this probability is *inscrutable* for you."<sup>15</sup> Therefore,  $\Pr(R \text{ (specified to us)}/N\&E)$  is either low or inscrutable.

## II. The Preliminary Argument

Plantinga's second argument (which he called the "preliminary argument") was intended to be an argument for the *falsehood* of naturalism. The first version of the argument was based on two major premises<sup>16</sup>:

- (1) The probability of  $\Pr(R/N\&E)$  is fairly low;
- (2) Our cognitive faculties are reliable, i.e.,  $\Pr(R)\approx 1$ .

Now by using Bayes's Theorem, we seem to have a straightforward argument against naturalism. Because according to Bayes's Theorem, we have:

$$\Pr(N\&E/R) = \Pr(N\&E) \times \Pr(R/N\&E) / \Pr(R)$$

Plantinga already argued that  $\Pr(R/N\&E)$  is not more than 1/2. And according to the second premise, since we believe R, we assigned it a probability of 1 or nearly 1. It means  $\Pr(N\&E/R)$  will be no greater than 1/2 times  $\Pr(N\&E)$ . Now even if we assign a high probability to N&E independent of the consideration of R ( a value less than one any way),  $\Pr(N\&E/R)$  will be less than 1/2. Therefore, if we think R is true, then we have evidence

against evolutionary naturalism. We do think R is true; therefore, we do have evidence against evolutionary naturalism (N&E).<sup>17</sup>

As Plantinga confessed this argument is "straightforwardly incorrect"<sup>18</sup> The second premise is false. The fact that we believe R, does not make its probability 1. Of course if we conditionalize R on our background knowledge (B), then  $\Pr(R/B)$  will be high, or even close to unity. But then the relevant application of Bayes's Theorem would be as follows:

$$\Pr(N/R\&B) = \Pr(N/B) \times \Pr(R/N\&B) / \Pr(R/B)^{19} \text{ ( Here for simplicity I just dropped E)}$$

But what Plantinga argued for is the claim that  $\Pr(R/N)$  is low, and what he needs here is the claim that the probability of R given N&B is low. He did not provide any reason for the latter claim. On the other hand one might say here what Plantinga means by  $\Pr(R)$ , is the *absolute* probability of R (conditioned only on necessary truths). If so, then there is no reason to assume that  $\Pr(R)$  is one, or nearly one. The fact that R is true in the *actual* world is no reason to believe that in a large proportion of all possible worlds, R is also true.<sup>20</sup> Therefore, the first version of the preliminary argument fails.

In the second version, in order to avoid those difficulties, Plantinga discharged  $\Pr(R)$ . But how? Since after all Plantinga's argument is supposed to provide a ground for preferring theism (T) to naturalism (N), he compared these two alternative "comparable" positions as follows:

$$\Pr(N/R) / \Pr(T/R) = \Pr(N) \times \Pr(R/N) / \Pr(T) \times \Pr(R/T)^{21}$$

$\Pr(R/N)$ , according to Plantinga, is low. On the other hand, Plantinga claims that  $\Pr(R/T)$  is high. According to traditional theism, we human beings have been created in the image of God. This means, among other things, God created us with the capacity for achieving *knowledge*. Therefore, given T, R is just what we would expect. So  $\Pr(R/T)$  is not low, or at least we have no reason for thinking so. Now given that we do not assign N a considerably higher absolute probability than T( since initially we seem not to have any

reason to prefer one to the other), we can conclude that  $\Pr(R/N)/\Pr(R/T)$  is less than 1, and therefore,  $\Pr(N/R)/\Pr(T/R) < 1$ . It means  $\Pr(N/R)$  is lower than  $\Pr(T/R)$ . If R is the only available evidence, then we have a reason to prefer T to N.<sup>22</sup>

### III.

#### The main argument

Finally, Plantinga's major argument is an argument against *rationality* of naturalism.

He calls it "the main argument". The main argument is as follows:

- (1):  $\Pr(R/N\&E)$  (specified to us) is low or inscrutable. (The probability thesis (PT))
- (2): For a devotee of naturalism, PT is a defeater for R. In other words, N&E&PT (Let's call it 'Q') defeats R.
- (3): For any naturalist (S), if S has a defeater for R, then S has an undercutting defeater for every belief B she holds, including N&E itself.
- (4): Therefore, evolutionary naturalism (N&E) is self-defeating.
- (5): Therefore, S cannot hold N&E rationally.

In this argument there are two sets of rationality defeaters and defeatees at work. In (2), Q functions as an undercutting defeater for R, and in (3), the denial of R functions as an undercutting defeater for N. So it seems clear that the notion of "rationality defeater" plays an important role in Plantinga's argument. What does he mean by "rationality defeater"?

Plantinga tried initially to explain the notion of rationality defeater in his argument as follows:

(D) D is a defeater of B for S at *t* if and only if (1) S's noetic structure N(i.e., S's beliefs and experiences and salient relations among them) at *t* includes B, and S comes to believe D at *t*, and (2) any person (a) whose cognitive faculties are functioning properly in the relevant respects, (b) whose noetic structure is N and includes B, and (c) who comes at *t* to believe D but nothing else independent of or stronger than D would withhold B(or believe it less strongly).<sup>23</sup>

Defeaters, including defeaters for defeaters, can be divided into *rebutters* and *undercutters*. If a defeater gives us evidence against belief B (but not strong enough to require the belief that not-B), we may call it a "rebutting defeater". An "undercutting

defeater" is a belief that indicates not that the defeatee is false but that one's grounds for it are inadequate or neutralized.

How does Plantinga support the main claims in this argument?

1. To move from (1) to (2) is a very crucial step in this argument. The crucial claim(CC) at this step seems as follows:

(CC): If  $\Pr(R/N\&E)$  is low or inscrutable, then one who believes  $N\&E$  has a defeater for  $R$ .

But what is the reason for (CC)?

There might be a general principle behind this claim as follows:

Principle (1): If  $\Pr(B/A)$  is low or inscrutable, then one who believes  $A$  has a defeater for  $B$ .

But this principle is obviously false. The fact that  $\Pr$  (Plantinga's evolutionary argument against naturalism is sound/UCSB has a philosophy department) is low or inscrutable, does not seem a defeater to Plantinga's argument for someone who believes that UCSB has a philosophy department.

Plantinga believes even though this principle does not work for *all* cases, but it might be sufficient at least for *some* cases.<sup>24</sup> And in fact he claims in the case of naturalist there is such connection between  $PT$  and  $R$ . But how is he going to support this claim?

Plantinga's argument for (CC), in principle, is an argument from analogy. He does not appeal to any general principles to establish (CC), instead he suggests cases where low or inscrutable probability seems to lead to a defeater, and then suggests that the evolutionary naturalist's position is, regarding  $R$ , relevantly analogous to those cases.

*First analogous case:* Suppose I believe that I have been created by a Cartesian demon who takes delight in fashioning creatures who have mainly false beliefs (but think of themselves as paradigms of cognitive excellence). Or the contemporary version of this scenario: suppose I come to believe that I have been captured by Alpha-Centaurian superscientists who have made me the subject of a cognitive experiment in which I have been

given mostly false beliefs. In both cases, I have a defeater for my natural belief that my faculties are reliable (i.e., R).

*Second analogous case:* Suppose I come to believe that I am a brain in a vat, and also believe that the probability of my cognitive faculties are reliable, given the belief that I am a brain in a vat, is low or inscrutable. Again I have a defeater for R.

*Third analogous case:* Suppose you believe that XX is a drug that causes widespread cognitive disorder: an hour after XX is ingested, the subject's cognitive function is no longer reliable. And this drug also prevents its victim from detecting her unreliability. Now suppose someone believes that she has taken this drug, and suppose she also believes that 90 percent of those who take the drug are no longer reliable, although of course unable to detect their unreliability. The person in this condition will continue to form perceptual beliefs, memory beliefs, and the like; and she will also continue, in some sense, to assume that her cognitive faculties are reliable. Again here Plantinga thinks that if the person believes that  $\Pr(R/I \text{ ingested } XX)$  is low, and also she believes that she ingested XX, then she has a defeater for R.

In the UCSB case, the soundness of Plantinga's evolutionary argument against naturalism is conditionalized on a fact which is *irrelevant* to it. But in Plantinga's cases, R is conditionalized on the facts which are directly relevant to R. But is there any general principle to capture this *relevancy*? It seems in all these cases, including the naturalist case, the crucial point is that, for example, N&E, or creation by Cartesian demon, involves a claim about the *origin* of our cognitive faculties. And because of that, if there is a low, or inscrutable, probability, given such an origin, that those faculties are reliable, then this does provide a defeater for R. Plantinga believes that it is hard to find or establish any principles along this line. But if a principle is wanted, he suggests a very limited one, something about beliefs specifying the origin and provenance of cognitive faculties. The antecedent of such

principle would capture what is in common to the situation in which (a) I believe that I have taken XX, or am a brain in vat, or a victim of a Cartesian evil genius, and (b) I also believe that the probability of R on the condition in question is low or inscrutable; the consequent would say that in those situation I have a rationality defeater for R.<sup>25</sup>

As we see, in all cases Plantinga tries to show that the naturalist's situation is relevantly like the situation of the person who is created by Cartesian demon, or of the brain in a vat, or of the person who digested XX. Since they have a rationality defeater for R, therefore the reflective naturalist (who is aware of PT and Plantinga's argument) also has a defeater for R. And of course in all these cases, the defeater of R is an undercutting defeater (as we will see).

2. (3), in the main argument, at least *prima facie* does not seem problematic. In general, if we have considered the question whether a given source of information or belief is reliable, and have a defeater for the belief that it is, then we have a defeater for any belief such that we think it originates(solely) from that source. Here we have a defeater for R, and it means we get a defeater for all of our beliefs.

But this conclusion and also the final conclusion of the main argument will follow only if some other conditions are met:

C<sub>1</sub>: D is a defeater of B only if D is a defeater of B *for someone*; D is a defeater for someone S only if S believes both D and B. (C<sub>1</sub> simply follows from the definition of "rationally defeater")

C<sub>2</sub>: The defeater must be itself undefeated. If we get a defeater for the defeater of R, then we can hold R (and therefore, all deliverances of our cognitive faculties, including N&E) rationally.

C<sub>3</sub>: There must be no sufficient independent evidence for R. Because the naturalist

(who accepts N&E&PT) might have such strong support for R as to outweigh the negative impact of N&E.<sup>26</sup>

In the case of the reflective naturalist, C<sub>1</sub> is obviously met. Plantinga believes that in the case of the reflective naturalist, both C<sub>2</sub> and C<sub>3</sub>, are also met, i.e., the devotee of N&E has no defeater for this defeater, and no independent evidence for R. He argues that:

(a) One cannot have non-question-begging support for R. For example, suppose one argues inductively for R: she might consider any of her beliefs. She believes that A<sub>1</sub>; that is she believes that A<sub>1</sub> is true; A<sub>2</sub> is also one of her beliefs and A<sub>2</sub> is true, and so on. By induction, she argues that all or nearly all of her beliefs are true, and therefore, her faculties are probably reliable. Here, when the person claims that, for example, A<sub>1</sub> is true, and when she relies on induction to conclude the reliability of her faculties, she has to rely on one or other of her cognitive faculties (i.e., perception, memory, and so on), and she cannot rely on them unless she assumes they are reliable. In other words, in so doing she assumes R. So any argument for R is in this context delicately circular or question-begging. Of course it is not *formally* circular, i.e., its conclusion does not appear among its premises. It purports to give a reason for trusting our cognitive faculties, but is itself trustworthy only if those faculties (at least the ones involved in its production) are indeed trustworthy. Once we come to doubt the reliability of our cognitive faculties, we cannot properly try to allay that doubt by producing an *argument*; for in so doing we rely on the very same faculties we are doubting.<sup>27</sup>

(b) The same goes for the defeater of defeater of R. The defeater of R cannot be defeated because any defeater would arise from the very faculties or belief-producing processes in question. For example, the defeater might take the form of an *argument*, perhaps for the conclusion that those belief-producing processes are reliable after all. But then we would have the same defeater for each of the premises of this argument, as well as for our belief that if the premises are true, then so is the conclusion.<sup>28</sup>

3. Now suppose R has an undefeated undercutting defeater. It mean the reflective naturalist has a good reason to believe  $\neg R$ . If  $\neg R$  is the case (i.e., if reflective naturalist's cognitive faculties are not reliable), then she has an undercutting reason for doubting every belief she has, including N&E, unless she has other evidence for it; but any purported other evidence will be subject to the same defeater as N&E. Therefore, she is not rational in holding N&E any more.

But there seems to be a problem here: As soon as N&Eer comes to doubt R, she should also come to doubt her *defeater* for R; for that defeater, after all, depends upon her beliefs, which are a product of her cognitive faculties. So her defeater for R is also a defeater for that defeater, that is, for *itself*. But then when she notes *that*, and *doubts* her defeater for R, she no longer *has* a defeater (undefeated or otherwise) for N&E; and so in spite of Plantinga, N&Eer seems to be rational to hold her belief in N&E.

To meet this objection, Plantinga offers the following scheme: Imagine a person who comes to believe  $\neg R$ , that is, that her faculties are not reliable. Then she can see that there is an infinite series of propositions related in an interesting way. At the first level, there is  $\neg R$ , which she believes. But there is a connection between  $\neg R$  and any belief she has, including  $\neg R$  itself: if  $\neg R$  is true, then any belief she has, including  $\neg R$  itself, will be unreliably formed. But *that* belief, i.e., the belief that  $\neg R$  is unreliably formed, gives him a potential defeater for  $\neg R$ . Suppose we let ' $\neg R(p)$ ' express the proposition that p is unreliably formed. Then at the second level we have  $\neg R$  and  $\neg R(\neg R)$ , which is a defeater of  $\neg R$ . But then at the third level we have  $\neg R$  and  $\neg R(\neg R(\neg R))$ , which is a potential defeater for  $\neg R(\neg R)$ , the defeater of  $\neg R$ ; and so on. We can schematically represent the structure as follows:

Level 0:  $\neg R$

Level 1:  $\neg R$  and  $\neg R(0)$  (i.e.,  $\neg R(\neg R)$ ) which is a potential defeater for level 0, i.e.,  $\neg R$ )

Level 2: -R and -R (1) (which is a potential defeater for level 1 and a potential defeater-defeater for level 0)

Level 3: -R and -R (2) (which is a potential defeater for level 2 and a potential (defeater-defeater)-defeater for level 0)

.....  
 Level n: -R and -R (n-1) (which is a potential defeater for level n-1)

.....  
 There is an infinite series of propositions here, but not an infinite series of defeaters and defeatees. Because, according to  $C_1$  (see, p.14), D is a defeater of B for S only if S *believes* both D and B. But no one could believe all the propositions in this infinite series. So even though each member is a *potential* defeater for the preceding member, nevertheless at some level, it is possible for S not to have any *actual* defeater for her belief B. In this scheme, the original defeatee (-R) shows up at every subsequent level. When that happens, the defeater-defeater does not nullify the defeater. The defeater gets defeated, but the defeatee remains defeated too. Accordingly, any time at which the person believes -R, she has a defeater for -R, even if she also has, at the time, a defeater for that defeater. This position, then, really is self-defeating, even if it is also the case that the person has a defeater for his defeater. In the case of N&Eer, Q ( i.e., PT &N&E) is a defeater for R, and the reflective naturalist (i.e., the naturalist who believes Q) has a defeater for N&E at any time which she accepts Q. She has such a defeater for N&E even if at that time she also has a defeater for that defeater. Her problem, after all , is that Q gives her a defeater for *everything* she believes, including themselves<sup>29</sup> Therefore, N&E is a self-defeating position, and it is not rational to be held by the reflective naturalist.

4. In all Plantinga's analogous cases, the probability in question *is low*. Now suppose we are unable to determine the value of  $\Pr(R/N\&E)$ . In other words, it is not low, but inscrutable. Is that sufficient, along with other relevant features of the situation, to make N&E a defeater of R? Plantinga's response is again, yes. And his strategy to support this answer is again to argue from analogy: imaging a person who comes into a factory, sees an

assembly line carrying apparently red widgets. Suppose she is agnostic about the probability of a widget's being red, given that it looks red. Now the shop superintendent tells her that those widgets are being irradiated by red light; but then a vice-president comes along and tells her that the shop superintendent suffers from a highly resilient but fortunately specific hallucination, so that he is reliable on other topics even if totally unreliable on red lights and widgets. Still, the vice president *himself* does not look wholly reliable. Then she does not know *what* to believe about those alleged red lights. She will presumably be agnostic about the probability of a widget's being red, given that it looks red; she won't know what that probability might be; for all she knows it could be very low, but also, for all she knows, it could be high. Therefore, Plantinga believes that the rational course for her is to be agnostic about the deliverances of her visual perception (so far as color detection is concerned) in this situation.<sup>30</sup>

## **Part Two: Objections**

### **I.**

#### **About the Preliminary Argument**

The conclusion of the preliminary argument (hereafter PA), is a conditional statement: If R is the only available evidence for us, and we believe PT, then we have a reason to prefer theism to naturalism.

(1) But it seems clear that R is *not* the only relevant evidence available to us.

Therefore, if we consider *all* relevant evidence (O), we may get a different result. In other words, it seems quite possible to find some additional relevant information (O) in such a way that  $\Pr(T/R\&O)$  comes out less than  $\Pr(N/R\&O)$ . The argument does not rule out such possibility.

But there seems to be a problem here: It might be difficult to say what O could be.

For according to one version of Bayes' Theorem, we have:

$$\Pr(T/R\&O)/\Pr(N/R\&O)=\Pr(R/T\&O)\Pr(O/T)\Pr(T)/\Pr(R/N\&O)\Pr(O/N)\Pr(N)^{31}$$

Suppose in this case,  $\Pr(R/N\&O)$  is greater than  $\Pr(R/N)$ , and suppose it also makes  $\Pr(R/N\&O)$  greater than  $\Pr(R/T\&O)$  (even though the latter clearly needs some reason to be acceptable), but still we cannot say that  $\Pr(N/R\&O)$  comes out greater than  $\Pr(T/N\&O)$ . This is because we have to look at  $\Pr(O/N)$  and  $\Pr(O/T)$ , and see how likely that information  $O$  is, given naturalism and given theism. Even if  $O$  raises the probability of  $R$  given naturalism, it may be improbable given  $N$ . Therefore, whatever  $O$  is, the probability of  $O$  given  $N$  must be higher than  $\Pr(O/T)$ .<sup>32</sup> And now, whatever  $O$  is, one may argue for the claim that  $\Pr(O/T)$  is high, in the same way that Plantinga argued for the claim that  $\Pr(R/T)$  is high. (see, p.9)

(2) Suppose we admit that it is not easy to find some additional information such that  $\Pr(O/N)$  is greater than  $\Pr(O/T)$ . But still Plantinga's way of reasoning in PA provides a recipe for replacing any nondeterministic theory in the natural sciences. For example, suppose quantum mechanics (QM) predicts that a certain experimental outcome ( $E$ ) was merely very probable. But theist may claim that this outcome was inevitable outcome of God's will ( $T$ ). Presumably,  $\Pr(E/T)$  is much greater than  $\Pr(E/QM)$ , and if we compare  $\Pr(T/E)$  and  $\Pr(QM/E)$  based on Bayes' Theorem, we will see that  $\Pr(T/E)$  is greater than  $\Pr(QM/E)$ , and so we have a reason to prefer  $T$  to  $QM$ .<sup>33</sup> This example makes it clear that the pattern of Plantinga's argument does not seem good at all.

(3) Suppose atheist accepts the conditional conclusion of PA. But then he can provide the same type of argument to prefer atheism (which is implied by naturalism) to theism.

For example, the atheist might appeal to the existence of evil in the world (or some particular cases of evil, or the amount of evil in the world), and provide a parallel argument in favor of atheism as follows:

Suppose  $O$  is the only available evidence we have:

(O): There is a great amount of evil in this world.

And suppose we have only two alternative hypotheses to compare with respect to O:  
 (T): The nature and the condition of sentient beings on earth is the result of benevolence actions performed by God.

(N) The nature and the condition of sentient beings on earth is not the result of the action of theistic God as an omnipotent omniscient, and perfectly good being.

If O is the only available evidence, it is reasonable to say that  $\Pr(O/N)$  is greater than  $\Pr(O/T)$ . According to Bayes' Theorem we will have:

$$\Pr(N/O)/\Pr(T/O) = \Pr(N) \times \Pr(O/N) / \Pr(T) \times \Pr(O/T)$$

Now given that we do not assign N a considerably higher absolute probability than T, and  $\Pr(O/N)$  is greater than  $\Pr(O/T)$ ,  $\Pr(N/O)$  comes out greater than  $\Pr(T/O)$ , and so we have a reason to prefer N to T.<sup>34</sup> And again it shows that the pattern of Plantinga's argument is not good.

(4) And finally, as we saw (p. 9), the claim that  $\Pr(T/R)$  is greater than  $\Pr(N/R)$  depends upon the claim that  $\Pr(R/T)$  is greater than  $\Pr(R/N)$ . But if  $\Pr(R/N)$  is *inscrutable*, then no one can figure out what value it should be assigned: it could be very high or very low, or somewhere in between. And so we are in no position to make any judgment about the relation between  $\Pr(R/T)$  and  $\Pr(R/N)$ . And it means if  $\Pr(R/N)$  is inscrutable, the preliminary argument will fail.

Therefore, either PA fails, or the conditional conclusion of it does not seem a great achievement: because at best it only provides *some* reason to prefer T to N.

## II.

### About the Probability Thesis

There are two important types of objection to the Probability Thesis(PT): One type is about PT *per se* – regardless its role in Plantinga's main argument; and the second type is concerned with the role that PT plays in the main argument as a defeater for R (as a part of Q i.e., N&E&PT).

## Objections to the Probability Thesis *per se*:

### 1. The Belief-Behavior Problem:

As we saw, Plantinga's crucial point to estimate  $\Pr(R/N\&E)$  was that  $\Pr(R/N\&E\&C)$  is not as high as it might seem (where 'C' is the claim that beliefs causally influence semantically as well as syntactically, and are adaptive( i.e.,  $A_4$  –see, p.3 ).) It is because, as he claimed, false adaptive beliefs are just as likely to evolve as true adaptive beliefs.

But there are some points regarding this claim:

(1) According to the commonsense view, though some of our actions on false beliefs have succeeded from time to time, nevertheless in general, our actions on false beliefs have *failed*; and they failed precisely *because* they were actions on false beliefs. So we have lots of inductive evidence that the holder of a mainly false belief-system will get into trouble sooner or later, even though the system might work in the short run.<sup>35</sup>

(2) False beliefs contain some truths.<sup>36</sup> For example, the claim that *that appletree witch is blooming* is false, but this belief contains some truths, like, *that is blooming; that's tree; something is blooming; something is there; appletree witch is not tiger witch....* And these true beliefs can explain that person's behavioral successes.

(3) As we saw, Plantinga's most promising reason for a mainly false but adaptive belief-system is the scenario of Paul - a prehistoric hominid, who suffers from some sort of "deep misconception" i.e., a person who has false beliefs about essential or definitional properties, as opposed to merely contingent features. Paul is a pan-psychist or an animist and, given that this position is false, most of Paul's beliefs are false. It seems Plantinga has an implicit argument here:

(a) Paul suffers from a deep misconception of some subject;

Therefore,

(b) All (or most) of Paul's beliefs about that subject are false.

It is not obvious how we can go from (a) to (b). Plantinga's strategy to defend this move (in the *third scenario*- p.5) explicitly *assumes* that the only way of referring to the various things is by way of definite descriptions, and these descriptions work the way Russell thought they do. Suppose Paul is a prehistoric hominid who believes every thing is a kind of witch. Therefore, when Paul says: "The tiger before me is dangerous", he means something like: "There is exactly one tiger-witch before me, and it is dangerous". This belief is false. But suppose we follow the work of Kripke, and say that in such an utterance the reference of the term "tiger" is due in large measure to causal connections that exist between the speaker and actual instance of tiger. In that case, it seems that the speaker can refer to a tiger, and can predicate truths about the tiger, even if he suffers from profound ignorance and/or false beliefs about the nature of tigers themselves. It seems plausible that a person can truthfully say some tiger is dangerous- and really be referring to a tiger- even though the person mistakenly thinks that that tiger is, say, essentially a creature of a certain witch kind.<sup>37</sup>

Now suppose all these critiques are true. This line of reasoning, at best, proves that  $\Pr(R/N\&E\&C)$  is high. But it should not be taken as a serious threat to PT. Because, even if  $\Pr(R/N\&E\&C)$  comes out high,  $\Pr(R/N\&E)$  still has a great chance to be low (as long as we accept Plantinga's argument for the claim that  $\Pr(-C/N\&E)$  is high). It is because, as we saw,  $\Pr(R/N\&E)$  can be calculated as follows:

$$(F_2): \Pr(R/N\&E) = \Pr(R/N\&E\&C) \times \Pr(C/N\&E) + \Pr(R/N\&E\&-C) \times \Pr(-C/N\&E)$$

It shows that  $\Pr(R/N\&E)$  is not entirely dependent on  $\Pr(R/N\&E\&C)$ . Now following Plantinga, suppose that  $\Pr(-C/N\&E)$  is high, and  $\Pr(R/N\&E\&-C)$  is low (the latter is the case if we assume that  $\Pr(R/N\&E\&C)$  is high). Let's say the first one is 0.8, and the second one is 0.2. Then  $\Pr(C/N\&E)$  will be 0.2; and then even if we assign  $\Pr(R/N\&E\&C)$  a value of 1,  $\Pr(R/N\&E)$  will be less than 1/2(0.36, to be exact).

## **2. The Inscrutability Problems:**

There are some features in Plantinga's argument which make us to say that  $\Pr(R/N\&E)$  is *inscrutable* rather than low. Some of the reasons are as follows:

(4) Belief and action might be related in a way which is not among Plantinga's four possibilities: Beliefs and actions might have neural event as common cause. In this scenario even though beliefs do not cause action, but they are, in an important and relevant way, related. (This scenario, in fact, appeared in the first version of Plantinga's argument, but in the recent versions, he dropped it.) Plantinga believes that the probability of R given this common-cause scenario is "inscrutable".<sup>38</sup> Since  $\Pr(R/N\&E\& \text{common-cause scenario})$  is inscrutable, and that  $\Pr(R/N\&E)$  is the weighted average of a set of probabilities, including  $\Pr(R/N\&E\& \text{common-cause scenario})$ , therefore,  $\Pr(R/N\&E)$  must be inscrutable as well.

(5)  $F_2$  will do the job only if (i) all possible relevant cases are taken into account, (ii) those cases are mutually exclusive, and (iii) all possible cases are equally probable (i.e., every case which is conceivably possible is essentially treated as being equally probable until additional information indicates otherwise- it is a version of the Principle of Indifference.) As long as Plantinga worked with  $F_1$ , we had no reason to think (i) and (ii) are met, and he never provided any reasons to think so. Maybe he switched to  $F_2$  to fix this problem. However,  $F_2$  or  $F_1$  cannot do the job unless they assume (iii) or the Principle of Indifference. But it is not clear at all why we should assume  $A_1$ -  $A_4$  are initially (and regardless additional information we may receive later on) equally probable. Besides that the Principle of Indifference itself is philosophically a matter of controversy: it claims to obtain probabilities from ignorance, and it is not clear at all that the ignorance of probabilities is a good guide to belief.<sup>39</sup>

Anyway, whether or not we are legitimate to use  $F_2$ , Plantinga himself confesses that the assignment of specific real numbers to the various probabilities involved in ( $F_2$ ) is

"laughable", and his estimates of those probabilities are "both imprecise and poorly grounded". And therefore, "the right course here is simple agnosticism: one just doesn't know what that probability is.... Then this probability is *inscrutable* to you."<sup>40</sup> After all, if we are legitimate to use F2, then this claim seems more plausible than the claim that  $\Pr(R/N\&E)$  is low.

**Objection to the Probability Thesis as a premise in the main argument:**

**3. The Problem of Conditionalization:**

According to the main argument the fact that the probability of R *conditionalized on* N&E is low or inscrutable, is a defeater for R and hence for every thing else he believes. But here there is an important question: Is the conditional probability of R given N&E the appropriate probability to work with? What is so special about this particular conditional probability? Suppose it is true that the probability of R given N&E is low. But there are many propositions P that naturalist believes (or are consistent with his naturalistic position) such that  $\Pr(R/N\&E\&P)$  is high. Why can't naturalist just add P to N&E, and raise the probability of R given this body of propositions?

We may call it "the problem of conditionalization". This is the problem of determining which conditional probabilities should be used in the main argument.<sup>41</sup>

Plantinga's argument seems to ignore other information we have that would make R likely. There might be some additional relevant evidence such that if we take it into consideration, we find that the probability of R conditional on it is high.

What could that piece of information be?

One possible option could be R itself: even though  $\Pr(R/N\&E)$  is low, but we *know* that R is true, and if we add this piece of knowledge to the conditional probability in question, we raised the value of it drastically (in fact, by adding R, the probability comes out 1). Then why isn't it rational for a naturalist to do so?<sup>42</sup>

Another option could be a proposition which either entails R or in some way raises the probability of R given the new set of evidence. For example, we might add the proposition O to N&E, where O is "simply a general proposition to the effect that the initial conditions of the development of organic life and the sum total of evolutionary process (including ones as yet known or only dimly understood) were and are such as to render  $\Pr(R/N\&E\&O)$  rather high."<sup>43</sup>

But there is a problem here: In the main argument, Q (i.e., PT&N&E) is a defeater for R. In other words, in this context the debate is over whether a certain source or origin of our beliefs or faculties would be reliable. Because of this we cannot conditionalize on anything that presupposes or is based upon that source being reliable. Such a probability statement would not be useful in this context, because it would presuppose what is being questioned. If we were to conditionalize on R itself, or any proposition such as O, then we could come up with a high conditional probability, but in doing so we would have begged the question in an important sense, because it assumes the source in question is reliable.

To avoid the problem of begging the question, we can conditionalize R on (N&E&R), or (N&E&O) only if we have some *independent* evidence for R or O, and Plantinga, as we saw before, argued that it cannot be the case.

But is he right? Here we need to ask two important questions:

- (1) Why cannot R be supported by its deliverances?
- (2) What if we accept R in the basic way, i.e., we accept it, but do not accept it on the evidential basis of other propositions? It is true that all our *propositional* evidence are dependent upon R, but it seems quite plausible to say that we have also *nonpropositional* evidence for R, and this sort of evidence is not based on R.

The first question is the question of "epistemic circularity": you cannot come to believe that a belief source is trustworthy by relying on the trustworthiness of that very

source. But epistemic circularity at least in some cases seems acceptable. For example, suppose one gives a simple inductive argument for the reliability of sense perception. She lays out a large, carefully chosen sample of perceptual beliefs, and report in each case that the belief is true. Thus the argument looks like this:

1. At  $t_1$ ,  $S_1$  formed the perceptual belief that  $p_1$ , and  $p_1$ .
2. At  $t_2$ ,  $S_2$  formed the perceptual belief that  $p_2$ , and  $p_2$ .

.....

Therefore, sense experience is a reliable source of belief.

Or perhaps I have never considered the question whether my perceptual faculties are reliable; I immediately conclude that they are, on the grounds that, as far as I can remember, they have produced beliefs of which the vast majority were true. Alston has argued in detail that epistemically circular argument of these kinds can be acceptable.<sup>44</sup> Plantinga also seems willing to concede that sometimes epistemically circular arguments are acceptable. For Plantinga, it could be the case only if there has been no reason to question R. But as he pointed out, sometimes one has a serious question as to whether R. If I have thought about whether R is true and have a serious doubt as to whether it is, according to Plantinga, I can't properly assuage that doubt by giving myself an epistemically circular argument.<sup>45</sup> Consider the unreliability drug XX example: suppose I take a good dose of XX. I believe that 95 percent of the population has the blocking gene; but I have no belief as to whether I myself have that gene. Since Pr (one's cognitive faculties are reliable/one has taken XX) is low or inscrutable, and I am aware of this fact, according to Plantinga, I have a defeater for R. Now suppose I came to believe that my physician has telephoned me and told me that I am among the lucky 5 percent whose reliability is unimpaired by ingesting XX. Can I rely on this new piece of information? Plantinga's answer is "No".<sup>46</sup> (We will get back to this point later in our discussion of "Rationality Defeaters".)

Suppose Plantinga is right to think that when we have a serious reason to question R, R cannot be supported by its deliverances. But now suppose R has a great deal of warrant in the basic way. It may be that the general reliability of our cognitive faculties is a fundamental presupposition that we are all warranted in accepting in the absence of any reasons or evidence whatever.<sup>47</sup> Or, if we employ Plantinga's language, one might say R has a special status: it might be a part of our *design plan* that we cannot abandon R. In fact, Plantinga himself explicitly says: "we do indeed, just as Reid said, believe R noninferentially, and this belief, held in that basic way, has a good deal of warrant for us."<sup>48</sup>

If it is so, then (a) it means we do have some *independent* evidence for R; and (b) R has a good deal of warrant for S, and that might commit S to believing tacitly that the probability of R on all the relevant evidence S possesses is high, and in so doing, she does not seem to be the question.

### III.

#### About the Main Argument

We may classify objections to the main argument into three major categories:

(1) Objections to "The Crucial Claim", (2) Questions about "The Rationality of Defeated R", and (3) The problem of "Atheist's Knowledge".

#### **1. Objections to "The Crucial Claim"**

The heart of the main argument is the move from PT to the defeater claim. This crucial step can be stated in terms of the following claim:

(CC) If  $\Pr(R \text{ specified to oneself}/N\&E)$  is low or inscrutable, then one who believes N&E has a defeater for R specified to herself or himself.

(1) As we saw, Plantinga's argument for (CC) is the argument from analogy. His basic strategy takes the following form: X is relevantly like Y; Y has F, therefore X has F.

One strategy for disarming such arguments is to show that X is *not* relevantly like Y. The first analogy Plantinga suggested was the analogy between the case of being created by evil demon (a demon whose creations have mainly false beliefs), and the situation of the

reflective naturalist. But there is an important difference between these cases. In evil demon case, the claim is that if the person is created by evil demon, her cognitive faculties are *actually* unreliable. But Plantinga does not suggest that his arguments show that the evolutionary naturalist must believe that the products of N&E are not just probably but *actually* unreliable. In other words, in evil demon scenario,  $\Pr(R/S \text{ being created by evil demon})$  is zero, but in naturalist case,  $\Pr(R/N\&E)$ , at worst, is low. Therefore, these two cases are *not* analogous.<sup>49</sup> In the first scenario, the conditional probability *clearly* functions as a defeater for R, but it is *clear* because the value of its conditional probability is zero. We may change the evil demon scenario in a way that the conditional probability comes out low, but by doing so, it won't be clear any more that the conditional probability does function as a defeater for R. The same point applies to the second scenario, i.e., the brain in a vat case.

Another strategy for disarming arguments from analogy is to suggest that X is more relevantly like Z than X is relevantly like Y, that Z lacks F, and therefore there is no reason to suppose that X has F, and indeed the more reasonable conclusion is that X lacks F. Thus if we can suggest an alternative to Plantinga's third analogy (i.e., the unreliability drug XX case), which is more relevantly like the reflective naturalist's situation, in fact, we may get a very different result. For example, imagine a pharmacologist who discovers that he was given a sizable dose of XX in utero. Whatever effects XX had in his cognitive faculties have been present for his entire life. Little is known about XX, so the pharmacologist sets about to study it. He studies its chemical composition and concludes that, when taken, XX will cause some sort of process involving the subject's cognitive faculties to occur- possibly resulting in total and permanent cognitive unreliability. He understands the basic principles of this process, and suspects that, based on what he knows about the process,  $\Pr(\text{one's cognitive faculties are reliable/one has taken XX})$  is low or inscrutable. But he also believes that the precise effects of XX depend on certain unknown factors. He realizes he needs more

information about the effects of XX. One good way to gain information about the effects of a drug is to administer the drug to test subjects and observe what happens to them. Suppose the pharmacologist sets up the appropriate drug trials and observes the vast majority of the test subjects in fact remain perfectly reliable.<sup>50</sup> Given these results, it is surely appropriate for him to revise his estimation of  $\text{Pr}(\text{one's cognitive faculties are reliable/ one has taken XX})$ -in light of the results of his trials, this probability turns out to be quite high. Of course, in all this process the pharmacologist relied on her cognitive faculties, but there seems to be a strong intuition behind the claim that the epistemic circularity in this case is harmless, and the pharmacologist has no sufficient reason to doubt the reliability of her cognitive faculties in this scenario.

The reflective naturalist's situation seems more like this scenario than Plantinga's third case. If it is so, it is plausible to say that the naturalist, like the pharmacologist, has no reason to doubt R specified to herself.

(2) And finally, as it was mentioned before, N&E will function as a defeater for R only if it meets some other conditions ( $C_1$ - $C_3$ ) as well (p.14). Among those conditions,  $C_3$  requires that R has no sufficient independent evidence. But as we discussed before (p.25 & 26),  $C_3$  is not met, i.e., R, when it is accepted in the basic way, has a great deal of independent warrant. But if it so, then we cannot say N&E defeats R.

## **2. Questions about the Rationality of Defeated R:**

According to the main argument, if one believes that the conditional statement of the (CC) is true, and also she affirms the antecedent of the conditional (i.e., PT), then she has to take N&E as a rationality defeater for R. And therefore, R is defeated for her, and it is no longer *rational* for her to hold R.

(3) But it is not clear what Plantinga means by "rational" in this context. There are some cases in which it seems rational to continue to believe the defeated belief, especially when the belief is R. For example, suppose again I take a good dose of XX. And I believe

that XX is a drug that causes widespread cognitive disorder, and it prevents the victim from detecting her unreliability. I also believe that 90 percent of those who take XX are no longer reliable, although of course unable to detect their unreliability. In this condition, I will continue to form perceptual beliefs, memory beliefs, and the like; and I also continue to assume that my cognitive faculties are reliable. Am I *irrational* in doing so? It doesn't seem so.<sup>51</sup> Even though I have a defeater for R, but I can't help forming belief in the ordinary way, and it might be a part of my *design plan* (using Plantinga's terminology) that I cannot abandon R. Abandoning R might lead to an epistemic/cognitive disaster in my every day life.<sup>52</sup>

Therefore, on some occasion a person might acquire powerful evidence against a given belief B but fail to change belief, due, not to dysfunction, but to the fact that the processes producing B are aimed not at the production of true belief (that is, not at the maximization of true beliefs and the minimization of false beliefs) but at the production of belief with some other property. In such a case, the design plan may dictate no change of belief in face of some sort of defeaters.

Plantinga believes there is a point here. In the XX case, *in a sense*, I am rational to continue to believe R, and in some *other* sense, I am not.

In order to make this point clear, Plantinga modifies his notion of "rationality defeater". He distinguishes between two different kinds of "rationality defeater": "*proper function* rationality defeater", and "*purely epistemic* or *purely alethic* defeater". The proper function defeater is the kind of defeater he defined and employed in the main argument. But now he thinks it is not the appropriate notion of defeater we should employ in the main argument. The purely alethic defeater is the one we need to make the argument work. He defines "purely alethic defeater as follows".<sup>53</sup>

(D\*) D is a purely epistemic [alethic] defeater of B for S at t iff (1) S's noetic structure N at t includes B and S comes to believe D at t, and (2) any person S\* (a) whose

cognitive faculties are functioning properly in the relevant respects, (b) who is such that the bit of the design plan governing the sustaining of B in her noetic structure is successfully aimed at truth (i.e., at the maximization of true belief and minimization of false belief) and nothing more, (c) whose noetic structure is N and includes B, and (d) who comes to believe D but nothing else independent of or stronger than D, would withhold B (or believe it less strongly).

The person who believes she has ingested XX has a purely alethic defeater for R.

And now Plantinga claims that the naturalist who comes to see PT also has a purely alethic rationality defeater. She might continue to assume R; nevertheless she has a defeater of this sort. She won't be able to help believing or at least assuming R; but if she reflects on the matter, she will also think that what she can't help believing is unlikely to be true.<sup>54</sup>

(4) We cannot simply conclude that N&E defeats R unless N&E is significantly better supported than R. If N&E is no better supported for S than R, then there is no rational basis for a preference for withholding R, rather than withholding N&E. So the question is "What defeats what?" Why does Plantinga think in this context N&E defeats R, and not *visa versa*?<sup>55</sup> At any rate, it seems quite clear that we believe that R is supported much more than N&E, therefore the natural thing to say is that N&E cannot defeat R, and if anything defeats anything, it must be R which defeats N&E.

Of course, it shouldn't be a serious objection to Plantinga's position. It is because either N&E is not supported by evidence enough to defeat R, and therefore it would be defeated by R, or it can defeat R, and so it would be a self-defeating position. In either case, it is irrational to hold N&E.

(5) But there is a problem here. Plantinga claims that even if  $\Pr(R/N\&E)$  is *inscrutable*, N&E still will be a defeater for R. But what is the reason for this claim? Here again, the heart of Plantinga's argument is the argument from analogy (see, for example, his case of "the widget under the red light, p. 17). But here again we may disarm or weaken his argument from analogy by suggesting another analogy: Suppose I take a good does of XX

by mistake. I run to the pharmacy to ask the pharmacist about XX. The pharmacist tells me that XX causes severe hallucinations within one hour of taking it. But then the manager of the store comes along, and tells me that the man in the pharmacy is not the pharmacist but a new technician, and he is not reliable on drug information. Still the so-called manager himself does not look wholly reliable. So I do not know *what* to believe about the alleged hallucinogen drug. It means I am agnostic about  $\Pr(R \mid I \text{ ingested a good dose of XX})$ . Does it mean as soon as I found out that the conditional probability in question is *inscrutable* I have a defeater for my cognitive faculties being reliable, and I am not rational in believing that I have hands and so on? It seems we have very strong intuition that it is not the case. In this scenario, it is surely rational for me to hold R, and rely on my cognitive faculties. If it so, then we have good reason to doubt about Plantinga's claim that if  $\Pr(R/N\&E)$  is inscrutable, N&E will be a defeater for R.

### **3. The Problem of "Atheist's Knowledge":**

Now suppose the reflective naturalist (who is also, by definition, an atheist) has indeed an undercutting defeater for R. Does it mean she has a defeater for everything else she believes? Plantinga's answer is clearly "yes": An atheist, if she is a reflective naturalist, has no knowledge. Of course Plantinga believes that "a naturalist [/atheist] who has never raised the question of reliability, or seen its connection with the Probability Thesis...will have no defeater for R or anything else he believes."<sup>56</sup> Presumably, we can say that a naturalist who has not been exposed to Plantinga's argument can know all sorts of things. She can further know that she knows that she knows those things, and even she can know that her cognitive faculties are reliable. But as soon as she is exposed to Plantinga's argument, she knows nothing and even if she wants, she is in no position to abandon naturalism in favor of theism, and return to God!

(6) More precisely, Plantinga's response to the question of the knowledge of the atheist is based on two assumptions:

- (i) If S has a defeater for R, then she has a defeater for everything else she believes.
- (ii) If S has a defeater for B, then she has no knowledge of B.

Suppose we granted (ii), but why should one believe (i)?

One possible reason is that (i) is based on the following *implicit premise*:

(IP) Whenever S accepts any deliverances of a faculty, S's doing so rests on the implicit premise that that faculty is reliable.

Since R is an implicit premise behind any belief, if S has a defeater for R, she has a defeater for every thing else she believes.

But (IP) is false. For example, a child may be innocent of any belief in the reliability for her faculties, yet it is hard to say she is irrational in her simple beliefs about food and shelter.<sup>57</sup>

### **Conclusion**

Plantinga's argument against naturalism, whether or not sound, may provide important insight into the epistemic shortcomings of naturalism. It seems plausible to me to make the following claims about Plantinga's argument against naturalism:

- (1) The Probability Thesis is plausible only if  $\Pr(R/N\&E)$  is considered *inscrutable*.
- (2) But if  $\Pr(R/N\&E)$  is inscrutable, the Preliminary Argument will fail. If  $\Pr(R/N\&E)$  is low, then the second version of the Preliminary argument will survive, but the conclusion is not an important achievement.
- (3) The strength of the Main Argument to a large extent depends upon the intuition behind Plantinga's analogies. For many, it is not easy to share the same intuition Plantinga has, and there are some ways to defend other intuitions which might lead to entirely different results. In particular, if we assume that  $\Pr(R/N\&E)$  is inscrutable, it

is hard to see how N&E can be a defeater for R. The Crucial Claim, when  
Pr(R/N&E) is inscrutable, does not seem true at all.

### Notes

\* I am indebted to Professor William Forgie, and Professor Anthony Brueckner not only for many helpful discussions of the issues involved in this paper, but also for their helpful criticisms of earlier draft of it.

1 Plantinga sometimes calls it "Darwin's Doubt", because this doubt seems to have (at least occasionally) crept into the mind of Darwin himself. See, Plantinga (1993) p.219

<sup>2</sup> Plantinga (1994), p.1

3 Plantinga (2002,1) P.2

4 For example, Patricia Churchland(1987), p.548.

5 Plantinga, (1993), pp223-225, introduced five accounts, but later on he reduced them into four accounts, see Plantinga (2002,1) Pp.6-9.

6 Plantinga (1993) p.225. Plantinga, hereafter, for simplicity, assumes that our behavior is a causal product just of our beliefs and desires.

7 Plantinga (1993), p.225.

8 Plantinga (1994), p.9

9 Branden Fitelson and Elliot Sober, (1998), p.417. And Fodor also seems to make a similar point; see, Fodor, (2002), p.39

10 Plantinga, (2002,2), p.260

11 Ibid, p.213; also Plantinga (2000), p.236

12 Plantinga (2002,1), p.6.

13 Plantinga (1993), p.228

14 Plantinga (1994), p.10

15 Ibid. .p.10

16 The first version appeared in Plantinga (1991), pp.38-39, and reappeared in Plantinga (1993), p.228-229

17 Plantinga (1993), p.228

18 Plantinga, (2000), p.229

19 This formula is the Bayes's Theorem we already had, but this time we are concerned with the probability of each item ( in the previous standard formula) given our background knowledge(B).

20 For some other difficulties of the second premise, see; Brand Fitelson and Elliott Sober, (1998), pp.414-416

21  $\Pr(N/R)=\Pr(N)\Pr(R/N)/\Pr(R)$

and also

$$\Pr(T/R)=\Pr(T)\Pr(R/T)/\Pr(R)$$

Therefore;

$$\Pr(N/R)/\Pr(T/R)=\Pr(N)\Pr(R/N)/\Pr(T)\Pr(R/T)$$

22 The second version appeared in Plantinga (2000), p.230-231

23 Plantinga (1994), p.29; and later on ,Plantinga (2000), p.362

24 Plantinga (1994), p.37

25 Plantinga (2002,2), p.240

26 For these conditions, see: Plantinga (1993), p. 233

27 Ibid. p.234

28 Ibid. p.233

29 Plantinga (2002), Pp.269-271

30 Plantinga (1993), pp. 230-231

31 We can get this formula as follows:

According to the General Conjunction Rule of Probability(GCR), we have:

(1)  $\Pr(A\&B)=\Pr(A/B)\Pr(B)$ ; or,

(2)  $\Pr(A/B)=\Pr(A\&B)/\Pr(B)$

Now we want to compare the following probabilities:

(F):  $\Pr(T/R\&O)/\Pr(N/R\&O)$

---

If we apply GCR(2) to (F), we will get:

$$\begin{aligned} \Pr(T/R\&O)/\Pr(N/R\&O) &= [\Pr(T\&R\&O/\Pr(R\&O))]/[\Pr(N\&R\&O)/\Pr(R\&O)] \\ &= \Pr(T\&R\&O)/\Pr(N\&R\&O) \\ &= \Pr(R\&T\&O)/\Pr(R\&N\&O) \\ &= [\Pr(R/T\&O)\Pr(T\&O)]/[\Pr(R/N\&O)/\Pr(N\&O)] \text{ ---\{GCR(1)\}} \\ &= [\Pr(R/T\&O)\Pr(O\&T)]/[\Pr(R/N\&O)/\Pr(O\&N)] \\ &= [\Pr(R/T\&O)\Pr(O/T)\Pr(T)]/[\Pr(R/N\&O)\Pr(O/N)\Pr(N)] \text{ ---\{GCR(1)\}} \end{aligned}$$

32 See, Richard Otte (2002), pp.138-141

33 Fitelson and Sober, (1998), p.416

34 For a similar kind of reasoning, see, Paul Draper, (1996), pp. 12-29

35 Fodor, (2002), p.38

36 Ibid., p.34

37 William Ramsey made this point; see, his article (2002), p.27

38 Plantinga, (1994), p.8, fn. (16).

39 William Alston, (2002), p.183. Fitelson and Sober also have some critical points on the use of the Principle of Indifference in Plantinga's argument, see their article, (1998), p. 422-423.

40 Plantinga, (1994) p.10

41 See; Richard Otte (2002), pp.141-142. Also, Fales (2002), p.53. Also, O'Connor (1994), pp.534-538. Also the similar point, even though in the different context, can be found in Fitelson and Sober (1998), p.418.

42 See, Carl Ginet (1995), p.407.

43 O'Connor, (1994), p.535

44 For a detailed discussion on "epistemic circularity", see, Alston (1986), pp.1-30. Also the following authors raised some questions on epistemic circularity in the main argument: Bergmann (2002), pp.76-82; Van Cleve (2002), pp.109-111.; Talbott (2002), pp.153-165.

45 Plantinga (2002,2), p.242

46 Ibid., pp.226-227.

47 The following philosophers believe that one can believe R in a "basic way" (if we use Plantinga's jargon), and it can provide a reason to deny Plantinga's conclusion in the main argument: Bergmann (2002), esp. pp.66-68, Bergmann tries to give a Reidian response to Plantinga, and his main point is based on the basicity of R. O'Connor even claims that "strictly speaking, there *couldn't* be a defeater for R, for any creatures in any possible world", see O'Connor (2002), p.131. Also see, Alston (2002), p. 200-201.

48 Plantinga (2002,2), p. 230.

49 Merricks (2002), p.172

50 This scenario is suggested by Wielenberg (2002), p.94.

51 Talbott (2002), esp.155-159. The case of "Purloined Letter" in O'Connor's article is employed to make the same point, see his article (2002), p.130-131. Also Bergmann made the similar point, see his article (2002), p.71-72.

52 O'Connor (2002), p.131

53 This definition appears for the first time in Plantinga (2000), p. 363; and then in Plantinga (2002,2), p. 209

54 See: Plantinga (2002,2), p.205-211

55 Alston (2002), pp.196-199

56 Plantinga (2002,2), p.243.

57 Sosa (2002), p.99; and also see, Van Cleve (2002), esp. pp.121-123.

---

**References**

- Alston, William (1986), "Epistemic Circularity", *Philosophy and Phenomenological Research*, Vol. 47, Issue 1, Sep, 1-30
- \_\_\_\_\_ (2002), "Plantinga, Naturalism, and Defeat", in *Naturalism Defeated?*, 176-203
- Beilby, James (ed.) (2002) *Naturalism Defeated? (Essays on Plantinga's Evolutionary argument against Naturalism)*. Cornell University Press.
- Bergmann, Michael (2002), "Commonsense Naturalism", in *Naturalism Defeated?* Pp. 61-90
- Churchland, Patricia(1987), "Epistemology in the Age of Neuroscience", *Journal of Philosophy*, 84, Oct.
- Draper, Paul (1996), "Pain and Pleasure: An Evidential Problem for Theists", in *The Evidential Argument from Evil*, ed. by Daniel Howard-Snyder, Indiana University Press, pp.12-29.
- Fales, Evan (2002), "Darwin's Doubt, Calvin's Calvary" in *Naturalism Defeated?*, pp.43-58
- Fitelson, Branden & Sober, Elliot, (1998) "Plantinga's Probability Argument against Evolutionary Naturalism", *Pacific Philosophical Quarterly* (1998, Vol.79, pp.115-129), republished in: Pennock, T. Robert *Intelligent Design Creationism and Its Critics (Philosophical, Theological and scientific Perspectives)*, Massachusetts Institute of Technology, 2000, pp.411-427.
- Fodor, Jerry (2002) "Is Science Biologically Possible?", in *Naturalism Defeated?*, pp.30-42
- Ginet, Carl (1995), "Comments on Plantinga's Two-Volume Work on Warrant", *Philosophy and Phenomenological Research*, Vol.55, Issue 2, Jun., pp.403-408.
- Merricks, Trenton (2002), "Conditional Probability and Defeat", in *Naturalism Defeated?*, 163-275.
- O'Connor, Timothy (1994), "An Evolutionary Argument against Naturalism?", *Canadian Journal of Philosophy*, Vol.24, No.4, Dec., pp.527-540.
- \_\_\_\_\_ (2002), "A House Divided against Itself Cannot Stand", in *Naturalism Defeated?*, pp.129-134.
- Otte, Richard (2002), "Conditional Probabilities in Plantinga's Argument", pp.135-149
- Plantinga, Alvin (1991), "An Evolutionary Argument against Naturalism", *Logos: Philosophic issues in Christian Perspective*, Vol. 12, pp.27-49.
- \_\_\_\_\_ (1993) *Warrant and Proper Function*. Oxford: Oxford University Press.
- \_\_\_\_\_ (1994) "Naturalism Defeated", unpublished paper. See the following link: [www.homestead.com/philofreligion/files/alspaper.htm](http://www.homestead.com/philofreligion/files/alspaper.htm)
- \_\_\_\_\_ (2000) *Warranted Christian Belief*. Oxford: Oxford University Press.
- \_\_\_\_\_ (2002,1), "The Evolutionary Argument against Naturalism: An Initial Statement of the Argument", in *Naturalism Defeated?*, 1-12.
- \_\_\_\_\_ (2002,2), "Reply to Beilby's Cohorts", in *Naturalism Defeated?*, pp.204-275
- Ramsey, William (2002), "Naturalism Defended", in *Naturalism Defeated?* Pp.15-29.

---

Reiter, David (2000), "Plantinga on the epistemic implications of naturalism", *Journal of Philosophical research*, Vol. XXV, pp.141-147.

Sosa, Ernest (2002), "Plantinga's Evolutionary Meditations", in *Naturalism Defeated?*, pp.91-102.

Talbott, W.J. (2002), "The Illusion of Defeat", in *Naturalism Defeated?*, pp.153-164.

Van Cleve, James (2002), "Can Atheists Know Anything?" in *Naturalism Defeated?*, pp.103-125.

Wielenberg, Erik J. (2002), "How to be an alethically rational naturalist?", *Synthese*, 131:81-98.