



## Plantinga, Naturalism, and Defeat

WILLIAM ALSTON

### I

Plantinga's "evolutionary argument against naturalism" has attracted much attention in the last few years. It received two simultaneous publications, in Radcliffe and White's *Faith in Theory and Practice* and in Plantinga's *Warrant and Proper Function*.<sup>1</sup> My references will be to the latter (hereinafter *WPF*). Having received many criticisms, Plantinga replies to them in a still unpublished paper, "Naturalism Defeated" (hereinafter *ND*). Despite all the critical fire, there are some serious problems with the argument that, so far as I know, have not been published to date. I begin with a quick sketch of the argument.<sup>2</sup> Further details will be supplied in the course of the discussion.

Plantinga first argues that the conditional probability of the (by and large) reliability of our cognitive faculties (R) on the conjunction of naturalism (N), current evolutionary theory (E), and the description of our faculties (C) is either low or inscrutable. He doesn't argue for this directly, but instead first argues for a like thesis with respect to "creatures a lot like ourselves on a planet similar to Earth. . . . Suppose . . . these creatures have arisen by way of the selection processes endorsed by contemporary evolutionary thought. . . . What is  $P(R/N\&E\&C)$ , specified not to us, but to them?" (*WPF*, 223).

<sup>1</sup> Elizabeth S. Radcliffe and Carol J. White, eds., *Faith in Theory and Practice: Essays on Justifying Religious Belief* (Chicago: Open Court, 1993); Alvin Plantinga, *Warrant and Proper Function* (New York: Oxford University Press, 1993).

<sup>2</sup> I omit any consideration of what Plantinga calls "a preliminary argument against naturalism" (*WPF*, 228–229) in order to concentrate on what he calls the "main argument".

The argument proceeds by a consideration of five possibilities concerning the relations of beliefs and behavior for the population in question. They are:

1. No causal connection of belief and behavior.
2. Beliefs are the effects of behavior but are not among the causes of behavior.
3. Beliefs have causal impact on behavior but not by virtue of their content.
4. Belief is causally efficacious, in terms of content and otherwise, but is maladaptive.
5. Like 4., except that belief is adaptive.

In the case of all these except (5), Plantinga suggests that the probability of R on N&E&C, given that alternative, is low. Then, assuming that the alternatives are mutually exclusive and jointly exhaustive, he takes  $P(R/N\&E\&C)$  to be an average of that probability on each of the five alternatives, with those probabilities weighted for each alternative  $A_i$  in accordance with  $P(A_i/N\&E\&C)$ . He then presents reasons for supposing that this weighted average will be “relatively low, somewhat less than  $\frac{1}{2}$ ”, or at least that it is “very unlikely” that the average is “very high”.

Plantinga then confesses that his “estimates of the various probabilities involved in estimating  $P(R/N\&E\&C)$  were (of necessity) both extremely imprecise and also poorly grounded. You might reasonably hold, therefore, that the right course here is simple agnosticism: one just does not know (and has no good way of finding out) what  $P(R/N\&E\&C)$  might be.”

Because of complexities in the presentation of the final stages of the argument in *WPF*, I will rely here on the crisper formulation in *ND*. There, after the argument that  $P(R/N\&E)$ <sup>3</sup> for the hypothetical population is either low or inscrutable, he goes on to infer a parallel conclusion for *us*.

... if this is the sensible attitude to take to  $P(R/N\&E)$  specified to that hypothetical population, then it will also be the sensible attitude towards  $P(R/N\&E)$  specified to us. We are relevantly like them in that *our* cognitive faculties have the same kind of origin and provenance as *theirs* are hypothesized to have. And the next step in the argument was to point out that each of these attitudes—the view that  $P(R/N\&E)$  is low and the view that this probability is inscrutable—gives the naturalist-evolutionist a *defeater* for R. It gives him a reason to doubt it, a reason not to affirm it.

Thus one who accepts N&E&C has a defeater for R and, therefore, for every belief formed on the basis of our faculties, i.e., for all our beliefs, including N and E. N&E is *self-defeating*. He finally adds that since if naturalism is true, then so, in all probability is evolution, naturalism *simpliciter* is self-defeating.

<sup>3</sup> In *ND* Plantinga omits C from that on which he argues that the probability of R is low. I will henceforward follow him in this.

## II

I have two major objections to this argument, laying far and away the most emphasis, and giving the most space, to the second. (1) Not all the probability assignments are adequately defended. (2) Even if  $P(R/N\&E)$  were either low or inscrutable, it would not follow that  $N\&E$  is a defeater for  $R$ , possibly on Plantinga's own conception of defeaters, on one understanding of what that is. At least, I will argue, this doesn't follow on a maximally epistemically significant conception of defeaters. The first step in laying all this out is to consider how Plantinga understands conditional probability. That determination is crucial for both of the above criticisms and also for the evaluation of the inference from the hypothetical population to ourselves.

In chapter 12 of *WPF* there is a footnote telling us how to understand the probability claims.

We could think of this probability in two ways: as a conditional *epistemic* probability, or as a conditional *objective* probability. Either will serve for my argument, but I should think the better way to think of it would be as objective probability; for in this sort of context epistemic probability, presumably, should follow known (or conjectured)<sup>4</sup> objective probability. (*WPF*, 220 n. 7)

But how does Plantinga understand 'epistemic probability' and 'objective probability'? He presumably did not feel called upon to explain them at this point because he had addressed that issue in chapter 9 of *WPF*, "Epistemic Conditional Probability: The Sober Truth." To be sure, that chapter doesn't completely solve our problem. For one thing, Plantinga distinguishes there *within* epistemic probability between an "objective component" and a "normative component", rather than simply contrasting epistemic and objective probability. With respect to the normative component he says the following.

In asking after the normative component of such a probability judgment [of  $P(h/e)$ ], we are asking what someone of "sound understanding," someone whose rational faculties are functioning properly, would believe [vis-à-vis  $h$ ] . . . given that evidence. (*WPF*, 163).

Moreover, although he holds that the objective component can always be construed as a logical relation, in cases where  $e$  does not entail  $h$  this requires that  $e$  includes other propositions, in addition to those on which the probability is explicitly conditional. But what additional propositions? Plantinga gives us at least a rough indication.

<sup>4</sup> Plantinga needs to qualify 'conjectured' with something like 'rationally', as he does with 'believed' in *WPF*, 162. For it would be a matter of luck if genuine objective probability should "follow" irrationally conjectured objective probability.

... the objective probability in question is indeed a logical probability, but it isn't one conditional just on the evidence. It is also conditional on other propositions: such propositions, perhaps, as that *The future will relevantly resemble the past* (the world is not, for example, a grue world) and perhaps *Simpler theories are more like to be true than complex ones*. Alternatively, the relevant set of possible worlds is not just the worlds in which the evidence is true, but some narrower class of worlds, perhaps specifiable in part in terms of similarity to what we think the actual world is like. (*WPF*, 162)

I take the general drift of these suggestions to be that the additional propositions in that on which the probability of *h* on *e* is (objectively) conditional are restricted to very general ones having to do with basic structural features of the world, rather than any that would ordinarily be thought of as evidence for or against *h*, or any other relatively specific proposition.

Returning to the presentation of the evolutionary argument against naturalism in chapter 12 of *WPF*, what I take to be the best way to align that presentation with the account of epistemic conditional probability in chapter 9 is to equate 'epistemic probability' in chapter 12 to 'the normative component' in chapter 9. As to Plantinga's statement in note 7 of chapter 12 of *WPF*, quoted above, that we should think of the conditional probabilities that figure in the epistemological argument against naturalism as objective rather than epistemic, I have no objection to that, provided we remember that in practice objective probabilities are decided by using the recipe quoted above from chapter 9 for determining epistemic probabilities. The rationale for this is that epistemic probabilities give us our best access to objective probabilities.

There is yet another large issue to be addressed concerning how to understand Plantinga's probability assessments. When asking about  $P(h/e)$  in the epistemic sense, are we asking what degree of credence one should give *h* on the basis of *e* alone, ignoring any further knowledge (beliefs) that are relevant to the epistemic status of *h*, or are we presupposing (the possibility of) other such items in the basis? I will term the first construal *parochial* and the second, *global*. Clearly we do make use of both construals. Suppose I ask, "If all you know about a woman is that her favorite novelist is Proust, what level of education would you suppose she has attained?" I am asking for the parochial probability of 'She has attained level *n*' on 'Her favorite novelist is Proust' (where the latter, of course, takes into account what Proust's novels are like). But when we make explicit probability judgments, they are almost always more global. Thus if we say something like "Given the meteorological conditions in the vicinity, the probability of rain in the next twenty-four hours is very high", we mean the basis to include not only the particular conditions in this vicinity now, but also some relevant body of general meteorological knowledge and hypotheses.

In Plantinga's argument we can find indications of both these construals. The parochial interpretation is suggested by the following.

1. The inference from the hypothetical population to us seems to presuppose this. In ND, we find Plantinga supporting it by pointing out that we are relevantly like that population “in that *our* cognitive faculties have the same kind of origin and provenance [according to N&E] as *theirs* are hypothesized to have” (11). A similar point is made in *WPF*, 229. But why should Plantinga think that this provenance is the only thing relevant to that probability? The obvious answer is that he is presupposing a parochial construal of  $P(X/Y)$  so that the support  $Y$  gives to  $X$  is the only thing relevant. Any additional information we have about us (and there will be a great deal of it) would make no difference.

2. The only thing Plantinga considers in arguing for evaluations of  $P(R/N\&E)$  for the hypothetical population is the bearing of N&E on  $R$ .

3. The apparent reason for beginning the discussion with the hypothetical population is that it forces us to ignore other things we know about us. This would seem to be the force of the rationale, given in *WPF* (222), that he proceeds in this way “in order to avoid distractions”. These “distractions” presumably consist of various other matters relevant to the epistemic status of  $R$  that we think we know about ourselves.

4. In ND he considers the “perspiration objection”. It runs as follows. “The probability that the function of perspiration is to cool the body, given (just) N&E, is also low, as is the probability that Holland, Michigan, is 30 miles from Grand Rapids, given N&E. But surely it would be absurd to claim that these facts give the partisan of N&E a defeater for those beliefs” (13). Now it is noteworthy that Plantinga does not counter this objection by denying those probability assignments, as he would if global probability were under discussion. Clearly we have a great deal of evidence for both the above  $h$ 's, but that does not figure at all in Plantinga's response. Instead his point is that a very low probability of  $h$  on  $e$  is not sufficient to make  $e$  a defeater for  $h$ . Rather than saying “On my conception of probability, ‘Holland, Michigan, is 30 miles from Grand Rapids’ is not improbable on N&E”, he admits the improbability but denies that it makes the latter a defeater for the former. Indeed, he gives another example—that “You own an old Nissan” is improbable on “You own a Japanese car”—and denies that this improbability makes the latter a defeater of the former. And there are other examples that he treats in the same way. All this makes sense only if  $P(h/e)$  is being treated parochially.

So far as I can see, what indications there are on the other side are minuscule by comparison. The clearest example I can find is the following. In *WPF*, when presenting the extension to us with respect to an inscrutable probability, Plantinga writes: “Well, *if we have no further information*, then wouldn't the right attitude here, just as with respect to that hypothetical population, be agnosticism, withholding belief?” (229; italics added). This seems to assume that other information *is* relevant to the probability assessment; otherwise there would be no point in the italicized qualification.

It might seem, and it did seem to me earlier, that the treatment of proba-

bility in *WPF*, chapter 9, is a more important indication of the global interpretation. In particular, there is the point, noted earlier, that in typical judgments of the objective probability of *h* on *e*, the logical probability relation being claimed is not one that holds just between *h* and *e*, but between *h* and *e* plus “certain other propositions”. And since, as noted, Plantinga rules in chapter 12 that we should be thinking of objective probability when considering the probability of *R* and *N&E*, this would appear to militate against a local probability interpretation.

But things are not always what they seem. To find a way out of this puzzle I will further deepen the problem by noting an apparent conflict between the treatments of the “objective” and “normative” aspects of epistemic probability in chapter 9. In spelling out the latter, Plantinga says the following.

The rough initial idea, then, is that the normative component of the conditional epistemic probability of *A* on *B* is the interval containing the degrees of belief a rational person could have in *A*, provided she believed *B* and was aware that she believed *B*, considered the evidential bearing of *B* on *A*, *had no other source of warrant for B or its denial*, and had no defeater for the warrant. (167; italics added)

The italicized phrase seems to place the “normative component” squarely on the parochial side of the contrast, whereas the account of the “objective component” cited above makes it explicit that what the probability of *h* is conditional on, when we speak of  $P(h/e)$ , includes more than *e*, thus rejecting a parochial interpretation. How, then, can it be, as Plantinga also says in chapter 9, that “In the typical probability judgment . . . these two components *coincide*”? (162).

The solution is found in a closer reading of the italicized phrase in the account of the normative component. It is only “other sources of warrant for *B* or its denial” that are ruled out of court in the specification of what the rational person would be taking into account. There is no parallel bar to taking into account the very general assumptions that Plantinga is thinking of as affecting the objective component.

How does this affect our choice between parochial and global construals for the argument in chapter 12? Clearly, it indicates that those alternatives were too unqualified. What we have just seen is that the probability under discussion in chapter 12 can be understood as parochial in a qualified sense. It is limited to the evidence specified, so far as *what counts specifically for or against h or its denial* is concerned. But that leaves open the relevance of very general considerations not included in *e*, as in the explanation of the “objective component”. We could term this either a qualified parochial or a qualified global interpretation. But since it is considerations that count for or against *h* that are more salient in the local-global contrast, I come down on the side of a *qualified parochial* interpretation.

## III

I now turn to my first main criticism of the argument, viz., that the probability assignments Plantinga makes in arguing for the thesis that  $P(R/N\&E)$  is low, specified to the hypothetical population, are undersupported. Recall that Plantinga proceeds by distinguishing different alternatives for the relation of belief and behavior, alternatives he takes to be mutually exclusive and jointly exhaustive. He then takes  $P(R/N\&E)$  to be “the weighted average of the probabilities for  $R$  on each of the possibilities—weighted by the probabilities of those possibilities”. And given his estimates of both these sets of probabilities he takes  $P(R/N\&E)$  to be “relatively low, somewhat less than one-half”.

To evaluate Plantinga’s reasoning, then, we have to consider both  $P(R/N\&E\&P_i)$ , for each alternative  $P_i$ , and  $P(P_i/N\&E)$  for each  $P_i$ . I begin with the latter, the probability of each alternative on  $N\&E$ , since it can be quickly dispatched.

Plantinga has little to say about this. In *WPF* he says nothing beyond asserting of each  $P(P_1/N\&E)$  that “a substantial share of probability must be reserved for this option” or that it has “some probability”. That doesn’t help us in weighting the alternatives. In *ND* he only says that “the probabilities of  $P_1$  (epiphenomenalism) and  $P_2$  (beliefs causally affecting behavior but not in terms of their content) would be fairly high, given naturalism”, and that the probabilities of the four alternatives<sup>5</sup> cannot sum to more than 1”, thereby implying that the probabilities of the third and fourth must be fairly low. When we consider the task at hand, it is not surprising that Plantinga would say so little. Indeed, he has said too much. We know nothing about the population under consideration except that they are “*rational*: that is, they form beliefs, reason, change beliefs, and the like” and that “their cognitive systems have evolved by way of the mechanisms to which contemporary evolutionary theory direct our attention, unguided by the hand of God or of anyone else” (*ND*, 6). With no more than that to go on, how could we possibly make even a rough estimate of the likelihood that their beliefs do or do not have a causal influence on behavior, or that their beliefs are adaptive or maladaptive? How could we reasonably suppose such a likelihood to be high or low? If we could reason by analogy with what we know about ourselves, we might arrive at something, perhaps something that is pretty much the opposite of Plantinga’s suggestions, for example, that the likelihood of causal efficacy of beliefs on behavior is much higher than epiphenomenalism. But the rules of the game forbid that. I see no alternative to throwing up our hands and declaring these probabilities *inscrutable*. But then the project of weighting the  $P(R/N\&E\&P_i)$  for each  $P_i$  by the  $P(P_i/N\&E)$  breaks down.

<sup>5</sup> In *ND* the five alternatives of *WPF* are reduced to four, by omitting (1)—no causal connections of belief and behavior.

Suppose we ignore the weighting and simply take the average of the probabilities of  $R/N\&E$  on each of the four alternatives. Though I have my doubts about some of those estimates, I am not at all clear about the matter and I will pass them by, so as to consider what follows if those estimates are accepted. Will that enable us to reach an objective result as to the overall value of  $R/N\&E$ ? A simple averaging will do the job only if we are justified in supposing the alternatives to be equiprobable as well as mutually exclusive. But why should we suppose this? We could invoke a principle of indifference if we were forced to make some assumption about their relative probability on  $N\&E$ . But who or what is forcing us? In any event, a reliance on a principle of indifference on that basis could hardly be supposed to have the sweeping conclusion Plantinga draws from his argument.

Plantinga should be only mildly disturbed by this last criticism. For, as we saw above, after assessing  $P(R/N\&E)$  for each of the alternatives he acknowledges that they are “both imprecise and poorly grounded”, and hence “You might hold that the right course here is simple agnosticism” (ND, 11).

So the question is this. If we are unable to arrive at a confident estimate of  $P(R/N\&E)$ , and we assume  $N\&E$ , does that imply that we “have good reason for being agnostic about  $R$  as well [assuming  $N\&E$ ]” (*WPF*, 229)? That does seem a sound judgment, at least in this sense. If all we know about the population in question is that they are cognitive agents, and  $N\&E$  holds of them, and we are unable to say anything about the probability of  $R$  on  $N\&E$ , then we are in no position to either assert or deny  $R$  of them. But, of course, what we are really interested in is  $R$  vis-à-vis us. And, it would seem, we are not subject to the above restriction. Far from it. We know a great deal about ourselves. And so the fact that we can’t assess the probability of  $R$  on  $N\&E$  alone will be of little interest if there are other things we know about ourselves that give  $R$  strong support. Suppose I cannot determine the probability that my wife is trustworthy on the evidence that she makes a mean Osso Buco Milanese. That will be of little moment if I know lots of other things that strongly support the proposition that she is trustworthy. So the verdict on the claim of agnosticism about  $R$  for us awaits the discussion below as to whether we do have considerable information about us that supports  $R$ .

#### IV

Now I come to my most serious criticism, which concerns the interpretation and significance of the claim that  $N\&E$  is a defeater for  $R$ . Even if  $P(R/N\&E)$  is low or inscrutable for us, how is that related to the claim that  $N\&E$  is a defeater for  $R$ , and how is that latter claim related to the result to which all this is supposed to be leading, viz., that it is irrational for one who holds  $N\&E$  to hold  $R$ , and hence to hold  $N\&E$  itself. We will find that the attempt to deal with these questions takes us along a tortuous route. I shall be



spending most of my time on the “low probability” side of the disjunction of this with “inscrutable probability”. The latter disjunct will receive brief treatment at the end.

First of all, we must be clear that Plantinga does not suppose that the probability of A's being low on B is sufficient, in general, to show that B is a defeater for A. This comes out most explicitly in the discussion of two of the objections to his argument in ND—the “perspiration objection” and the “Austere Theism objection”. Briefly, the first of these objections points out that although “the probability that the function of perspiration is to cool the body, given (just) N&E is low, as is the probability that Holland, Michigan, is 30 miles from Grand Rapids, given N&E . . . surely it would be absurd to claim that these facts give the partisan of N&E a defeater for those beliefs” (ND, 13). Plantinga responds by denying the principle that a low or inscrutable probability of X on Y is a sufficient condition for Y's being a defeater of X. That, of course raises the question of what makes the difference between those cases, like the above, in which the low or inscrutable probability does not engender a defeater and those in which it does. I don't find that his response to these objections in ND does a very good job of answering that question. But in the presentation of his original argument, both in ND and, at somewhat greater length in *WPF*, he presents some features of the N&E–R relationship that can be seen to differentiate it from the “perspiration objection”. The crucial claim is that N&E involves a claim about the *origin* of our cognitive faculties. And because of this, 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, the claim that they are reliable. Whereas N&E is not related in that way to the claims about perspiration or the distance between Grand Rapids, Michigan, and Holland, Michigan.<sup>6</sup>

This point is spelled out by presenting some analogous cases in which the origin of a belief, or of beliefs of a certain type, is such as to make it improbable that the belief(s) would be true, cases in which, Plantinga claims, it is obvious that this provides a defeater for the supposition that beliefs with such an origin are generally true.<sup>7</sup>

Suppose I believe that I have been created by an evil Cartesian demon who takes delight in fashioning creatures who have mainly false beliefs (but think of themselves as paradigms of cognitive excellence): then I have a defeater for my natural belief that my faculties are reliable. Turn instead to the contemporary version of this scenario, and suppose I come to believe that I have been captured by Alpha-Centaurian superscientists who have made me the subject of a cognitive experi-

<sup>6</sup> In correspondence Plantinga has pointed out some other differences. We have strong independent evidence for both of the claims in the objection, but no strong (or any) evidence, in the context of the argument, that is not question begging, by reason of assuming the very point at issue, viz., that our cognitive faculties are generally reliable.

<sup>7</sup> I am leaving the inscrutability alternative for later treatment.

ment in which the subject is given mostly false beliefs: then, again, I have a defeater for R. (ND, 11)

Compare the case of a believer in God, who, perhaps through an injudicious reading of Freud, comes to think that religious belief generally and theistic belief in particular is almost always produced by wish fulfillment. Such beliefs, she now thinks, are not produced by cognitive faculties functioning properly in a congenial environment according to a design plan successfully aimed at truth; instead they are produced by wish fulfillment, which, while indeed it has a function, does not have the function of producing true beliefs. Suppose she considers the objective probability that wish fulfillment, as a belief-producing mechanism, is reliable. She might quite properly estimate this probability as relatively low. . . . But then . . . she has a defeater for any belief she takes to be produced by the mechanism in question. (*WPF*, 229–230)

In both *WPF* and ND Plantinga goes on to claim that one who believes that his cognitive faculties are produced according to N&E, and who recognizes that the probability of those faculties being generally reliable, given that provenance, is low, thereby has a defeater for the belief that those faculties are reliable, and, by derivation, for any belief produced by those faculties, including N&E.<sup>8</sup>

This looks to be a promising line of argument, but it needs further spelling out. Just what origin does N&E take our cognitive faculties, in particular our belief-forming faculties, to have? The impression given by Plantinga's exposition is that N&E implies that these faculties result from evolutionary mechanisms, in particular natural selection. For one thing, the argument for a low value of  $P(R/N\&E)$  is developed in opposition to the position of Popper and Quine that this origin makes it likely that these faculties are reliable. Moreover, in *WPF* Plantinga comes close to an explicit formulation of this reading. "Suppose we think N&E is true: we ourselves have evolved according to the mechanisms suggested by contemporary evolutionary theory, unguided and unorchestrated by God or anyone else" (229).<sup>9</sup>

But things are not always what they seem. Remember the five possibilities for the relation of belief and behavior distinguished by Plantinga. On only the fifth of these does belief causally affect behavior, by virtue of its content, in an adaptive fashion. On the other four either belief does not causally affect behavior at all (1 and 2), or it does but not by virtue of its content (3), or it does by virtue of its content but not so as to render behavior adaptive (4). It is only on the fifth possibility, on which (true) belief influences behavior so as to render it likely to be adaptive, that natural selection would favor reliable be-

<sup>8</sup> In *WPF* Plantinga also cites the case of an "undermining" defeater (one that is not even probabilistically incompatible with the target belief but that indicates that S's ground for that belief lacks force). But this is much less analogous to the N&E-R relationship. Plantinga does not present N&E as undermining the evolutionary naturalist's reason for accepting R. Indeed, he says nothing as to what such a reason might be.

<sup>9</sup> The first conjunct is supposed to come from E and the second from N.

belief-forming faculties. On the other four, as Plantinga says with respect to the first two, “they are invisible to evolution” (*WPF*, 223). But in that case, evolutionary mechanisms plays no role in their genesis—in our acquiring belief-forming faculties at all, or in our acquiring such faculties of one kind rather than another. And so, unless we can limit the relevant possibilities to (5) (and Plantinga’s argument essentially depends on not doing so), we have to say that N&E can’t be or imply a thesis about the origin of these faculties in that way. “The mechanisms suggested by contemporary evolutionary theory”, in particular natural selection in forms of various degrees of sophistication, can have nothing to do with it.

But all is not lost for Plantinga’s argument. The N component of N&E may come to the rescue. For, on alternatives 1–4, N guarantees that our acquisition of these faculties is “unguided and unorchestrated by God or anyone else”. So if their origin cannot be explained by evolutionary mechanisms, N&E is left with nothing to say about the origin that would serve to explain why we have cognitive faculties of the sort we do. In particular, N&E, on any of the first four alternatives, would have nothing to say that is relevant to the question of their reliability. It is conceivable that natural laws should be discovered such that the possession of whatever psychic factors influence behavior in an adaptive fashion would carry with it, by nomological necessity, the possession of beliefs with propositional content that themselves have no such influence. But even so, this would provide no answer to the likelihood of the reliability of the belief-forming faculties. So far as the question of reliability is concerned, we may as well say that the faculties arose by sheer chance, by some sort of random drift. And this is just as good a reason for holding that on (1) to (4) the probability of our belief-forming faculties being reliable cannot be higher than  $\frac{1}{2}$ , and would seem to be significantly lower.

Thus Plantinga has a strong case for the thesis that *if* there is a low probability of R on N&E, that has a significant bearing on the rationality of holding both N&E and R. They are in such conflict that it would be irrational to hold both. But to evaluate Plantinga’s full argument and the conclusion thereof, we need to ask whether this implies that one who holds N&E has a defeater for R (rather than vice versa), and if she does whether this makes it irrational for her to hold R (rather than irrational for her to hold N&E). And the first step in looking into that is looking at the concept of *defeat* being employed. To this I now turn.

## V

In both ND and chapter 11 of *Warranted Christian Belief* (hereinafter *WCB*),<sup>10</sup> entitled “Defeaters and Defeat”, there is extensive discussion of the issue. In *WCB* we find the following definition.

<sup>10</sup> New York: Oxford University Press, 2000.

(D\*) *D* is a purely epistemic 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). (*WCB*, 363)

Plantinga's reason for concentrating on the notion of a 'purely epistemic defeater' is that he wants to identify the kind of conflict between *D* and *B* that would make it a violation of the aim at truth to hold them both, rather than a violation of the aim at, for example, survival or psychological comfort. Even if believing *B* along with *D* would make one extremely uncomfortable, that would not (tend to) make *D* a purely epistemic defeater of *B*.

In this chapter of *WCB* Plantinga also makes a couple of further distinctions that are relevant to understanding the notion of a defeater he is using. First the kind of defeater picked out by *D*\* is a *rationality defeater*. "Given belief in the defeating proposition, you can retain belief in the defeated proposition only at the cost of irrationality" (*WCB*, 359). This is distinguished from a *warrant defeater* that prevents the defeated proposition from enjoying enough warrant to count as known (even if true), but is not such as to prevent one from rationally accepting the proposition. Plantinga's example is the familiar barn case. *S*, on driving through the countryside sees something that looks like a barn, and he forms the belief that it is a barn. And so it is. But, as it happens, the area in question contains numerous barn facades that would look just like a barn to *S*, given his distance, speed, angle of vision, etc. *S* is rational in taking what he sees to be a barn, but since it is just a matter of luck, given what is in the neighborhood, that what he saw at that point was a barn rather than a barn facade, he can't be said to *know* that it was a barn. Note that a rationality defeater is a *belief*, something that belongs to the subject's noetic structure, whereas a warrant defeater can be some fact external to that structure.

The other distinction is between *internal* and *external* rationality. "*Internal* rationality is a matter of proper function 'downstream from experience'. . . . *External* rationality, by contrast, is a matter of the proper function of the *sources* of experience . . ." (*WCB*, 365). Spelling this out a bit, I take it that internal rationality has to do with the internal economy of one's noetic structure—its coherence—while external rationality has to do with the way in which new items are added to that structure. Plantinga's example of the latter in this passage has to do with the role of pride in distorting one's evaluations of one's own accomplishments. In a moment I will explore the way this distinction bears on Plantinga's use of 'defeater', something that I find not entirely clear from the texts at my disposal.

To return to the definition,  $D^*$ , one thing on which we obviously need more light is the stipulation that the hypothetical subject's "faculties are functioning properly *in the relevant respects*" (italics added). How is the italicized phrase to be fleshed out? Plantinga does not tell us. A natural reading would be that these respects include anything that is relevant to the epistemic status of B and D and to their logical (including probabilistic) relationships, and hence that is relevant to what would be the most rational response to their incompatibility. This would at least include the kind and degree of incompatibility relation between B and D, what intrinsic positive epistemic status is enjoyed by each along with whatever support each receives from other components of S's noetic structure.<sup>11</sup> But why wouldn't the degree of warrant enjoyed by D and B also be a "relevant respect", even where determinants of degree of warrant lie, at least in part, outside S's noetic structure? Wouldn't that be highly relevant to determining whether it would be more rational to reject B or to reject D? Isn't it clear, from Plantinga's epistemological perspective, that in case of incompatibility it is most rational to stick with the more warranted contender? Here the question of whether it is internal or external rationality or both (or alternatively, the question of what kind of proper function) is involved in  $D^*$  is crucial. If degree of warrant is a relevant respect, then the rationality defined by what one whose cognitive faculties are functioning properly *in the relevant respects* would do would include external, as well as internal, rationality. But it is clear that Plantinga is not prepared to be that inclusive. This comes out most clearly in his insistence, in both ND and *WCB*, that a belief with no warrant can be a defeater. ". . . a belief D can defeat another belief B, for me, even if D has little or no warrant for me" (ND, 24). "Suppose I hold a belief B, but then come to accept a belief D that goes against B in some way, but this belief D I accept has no warrant. Can it still be a defeater for B? I should think so" (*WCB*, 364). In these passages Plantinga does not say that B does have warrant, but since he gives no hint that a greater warrant for B would prevent D's from defeating B, I shall take it that he means to be committing himself to the thesis that a belief with a lesser (down to no) warrant can defeat a belief with greater warrant. This does clearly indicate that the relative warrant of D and B is not included in the "relevant respects" specified in  $D^*$ . But, furthermore, Plantinga says "it is possible for a belief that is irrationally acquired to be a defeater, even for a belief that is rationally acquired" (*WCB*, 363). Here is one of his examples.

Suppose I've always thought you a genial sort who is rather well disposed to me. Unhappily, I start sinking into a paranoid condition; because of cognitive malfunction, it comes to seem to me that you are, in fact, trying to harm me by destroying my academic reputation. Because of the cognitive malfunction, this just

<sup>11</sup> If the rationality involved in  $D^*$  is restricted to internal rationality, then it is only support from within the noetic structure that is relevant. Otherwise support from outside is or might be relevant. I will return to this issue.

seems wholly obvious to me. . . . Can my belief D that you are trying to destroy my reputation serve as a defeater for my belief B that you are favorably disposed toward me? (*WCB*, 365)

He answers this in the affirmative and supports this answer by invoking the distinction between internal and external rationality. In terms of that distinction, he says that the belief that you are out to get him is *externally* irrational “because it arises from sources of doxastic experience that are not functioning properly”. (By ‘doxastic experience’ Plantinga refers to the *sense* of truthfulness or falsity that accompanies beliefs.)

By virtue of their malfunction, however, my experience is such that I am powerfully impelled to believe D, that you trying to ruin me. This now seems to me much more obvious than that you are favorably disposed toward me: the doxastic evidence for D is much stronger than that for B. What internal rationality calls for, under those circumstances, therefore, is my giving up B; I have a defeater for it in D, even though D is arrived at irrationally. I can therefore have a defeater D for a belief B, even where B is rationally held and D is irrationally acquired. (*WCB*, 365)

Thus, according to Plantinga, a less warranted belief, D, can defeat a more warranted belief, B, and an (externally) irrationally acquired belief can defeat an (externally) rationally acquired belief, provided that *internal* rationality in each case calls for retaining D rather than B. Note that this example, and other examples he gives, supports my conjecture that internal rationality concerns the internal economy of one’s noetic structure, what one can access concerning it “from the inside”. What it leaves out, *inter alia*, is the relation of that structure to outside factors, including, crucially, facts concerning the origin of its components, *at least so far as these facts do not concern relations of the new component to already existing components*. Thus, to a first approximation, we may think of the “relevant respects” mentioned in D\* as restricted to what it takes to maximize internal rationality.

Now we must examine another clause of D\*, the subclause of (2) that requires of the hypothetical subject, whose potential doxastic reaction to the D-B conflict determines whether D defeats B, that it be one “(c) whose noetic structure is N and includes B”. We will see that this clause reinforces the above conclusion that the “relevant respects” include those that are concerned with internal rationality but not those concerned with external rationality. Of course, one rationale for the clause is that it ensures that the model subject has the qualifications for the task, in particular has the right noetic structure. But a close examination of the way Plantinga handles certain examples suggests that it plays the further role of making the point that positive and negative probability relations among components of the doxastic struc-

ture are taken into account by our ideal subject in deciding how to resolve the D-B conflict.

I have thought for some time that you once spent a year in Aberdeen, Scotland; you tell me (soberly, with no hint of teasing or joking) that you have never been to Scotland, although you once planned to go but were prevented at the last moment. I thus learn that you have never been to Scotland; the noetic structure that results from adding this belief to the noetic structure I have at *t* is irrational. . . . What rationality requires is that I change belief. . . . But not just any old change. . . . What needs to be given up is my belief that you once visited Aberdeen. . . . You might think that other changes would be . . . consistent with rationality. . . . Perhaps I could give up the belief that you are now truthful, or the belief that you are mentally competent, or the belief that you are capable of distinguishing Aberdeen, Scotland, from Aberdeen, South Dakota. *And perhaps these changes would be rational with respect to some noetic structures—ones, perhaps, in which I have enormously powerful evidence for your having been in Aberdeen . . . or structures in which I have good reason to doubt that you are telling the truth. But these are quite different from the structure I do in fact display at t, which involves my being quite properly sure that you are telling the truth, and also involves my having little more by way of support for the belief that you have been to Scotland, than a sort of vague memory to the effect that I once learned this. With respect to this noetic structure, these changes would not be rational. That is why the defeater is a defeater for my belief that you have been in Aberdeen, rather than for some other belief.* (ND, 33–34; italics added)

The italicized portion makes it crystal clear that Plantinga is prepared to take into account the relevant epistemic status of D and B in deciding whether D defeats B rather than vice versa, or rather than D defeating something else if anything. Is this incompatible with the restriction to *internal* rationality? Perhaps not, so long as the only sort of epistemic statuses in the picture are those that consist in the logical and probabilistic relations of the beliefs in question to other beliefs in the structure. And what is said here is compatible with that (though Plantinga's speaking of being *properly* sure that you are telling the truth evokes echoes of his concept of warrant, which is by no means confined to internal relations within the noetic structure).

But this example, and other examples, do raise one question about the understanding of internal rationality that I have not yet addressed. When Plantinga speaks of "powerful evidence for your having been in Aberdeen" and "good reason to doubt that you are telling the truth" and "little more by way of support", are these to be understood as referring to objective relations of *powerful evidence for* and *good reason to*, which obtain, where they do, regardless of whether S believes they do or is justified (warranted) in believing they do? Or are they to be understood as referring to what S believes about such relations (with or without justification)? Or are both required?<sup>12</sup> My previous in-

<sup>12</sup> We shall see later that Plantinga holds that a low probability of B on D does not render D a defeater of B unless S believes that this low probability relation obtains.

terpretation of internal rationality as concerned with the “internal economy” of the noetic structure is ambiguous between these interpretations. It could be “internal” by virtue of the components of the economy without having to include, among the components, the relations between the components that make the economy what it is. Furthermore, if we were to insist that beliefs in the logical and probability relations that make the economy what it is be among the components of the structure, we would be off on an infinite regress. For having included beliefs in these relations among the components, we would then need more relations between these new components and the others. And then, having added those relations . . . We would never get the complete economy specified. For this reason, and since a natural reading of epistemic terms like ‘powerful evidence’ would be in terms of objective relationships, and in the absence of any statement by Plantinga known to me on the point, I will assume that a reading in terms of objective evidential relations is compatible with the restriction to internal rationality.<sup>13</sup> That leaves the way in which Plantinga handles this and other examples compatible with the ignoring of the relative *warrant* of D and B in considering what a subject whose cognitive faculties are functioning properly *in the relevant respects* would do about the D-B conflict. For in Plantinga’s epistemology, warrant generally depends on more than evidential relations within a noetic structure. Moreover, it leaves the treatment of the examples compatible with allowing that an irrationally acquired belief can be a defeater of a rationally acquired defeater, provided what makes the former irrational is, at least in part, something outside the noetic structure.

But there is another consideration that seems to disturb this picture of sweet harmony. In ND, after saying that “a belief is rational in a certain set of circumstances when it is a healthy or sane belief to hold in those circumstances” (21), he continues:

The relevant circumstances have a two-tiered character. First, there is my *noetic structure*; an assemblage of beliefs and experiences (and other cognitive states such as doubts, fears and the like) together with various salient properties of these states and relevant relations obtaining among them. Let’s oversimplify and think just of beliefs and experiences. A description of a noetic structure would include a description of the strength of each belief, of the logical relations between that beliefs and others, and of the circumstances (crucially including experiences) under which the belief in question was formed and sustained. Not all beliefs are formed in response to experience (together with previous belief), and it may be that some beliefs are formed in response to experience, previous belief and still other circumstances: let’s use the ugly but popular term *doxastic input* to denote

<sup>13</sup> We should also note that in *WPF* the claim that N&E is a defeater for R, for us, is sometimes qualified by provisos like “If we have no further information [about R]” (231), or “if we have no independent evidence [for R]” (233). But unless the epistemic status of R vis-à-vis the rest of S’s noetic structure were relevant to the claim that N&E is a defeater for R, these provisos would have no point.



whatever it is that beliefs are formed in response to. Then we can say that a description of S's noetic structure would include an account of the doxastic input to which S has been subject, as well as an account of the doxastic responses thereto. (ND, 21–22)

This account obviously is in line with my earlier supposition that the relations that structure a noetic structure can do so whether or not the subject believes them to obtain. In speaking of relations obtaining between beliefs, Plantinga gives no hint that he thinks such relations are (must be) believed by the subject to obtain. But there is a further feature of the account that does not jibe with the developing picture. According to the passage just quoted, “a description of a noetic structure would include a description . . . of the circumstances (crucially including experiences) under which the belief in question was formed and sustained”. And it is made explicit that these circumstances are not restricted to other beliefs. Hence it would seem that a full description of a noetic structure would include a full account of the “circumstances” under which the belief is formed and sustained. And that, in turn means that the description would include the degree of warrant of each belief, since in Plantinga's epistemology warrant is a function of relevant facts about the forming and sustenance of the belief in question. But then, given that relevant facts about S's noetic structure are among those that are relevant for the response of our paradigmatic subject (in  $D^*$ ) to the D-B conflict, how can the relevant degree of warrant fail to be included in those considerations? And hence how can it be rational for our ideal subject to prefer a less warranted to a more warranted belief? And, to put it in terms of rationality, how, given the relevance of the noetic structure as so described, can the ideal subject fail to make the response that is most rational, externally as well as internally?

We seem to have a flat incompatibility between these texts. For a while I thought that I was missing something in the account of noetic structures just quoted, something that neutralized the apparent implication that the full description of a noetic structure would include a specification of the degree of warrant enjoyed by each constituent belief—though I had not been able to find any such something. But in a recent correspondence Plantinga writes in response to a question I posed, “As for a noetic structure, I was thinking of it as including something like an index of degree of warrant for each belief”. And so the incoherence won't go away. For the moment I will simply live with two versions of ‘defeat’; in one, relative degree of warrant is relevant to what defeats what, and in the other it is not relevant. The first version takes the noetic structure to include everything specified in the above quotation, including degree of warrant for each belief, and takes all that as relevant to the hypothetical subject's response to the D-B conflict. Whereas the second version places a constraint of only internal rationality on the response of the hypothetical subject, where internal rationality does not require taking into account the “external” features of how the belief is formed and sustained. In this

chapter I will go with the second, leaner version, which is more strongly supported by the texts.

## VI

So much, for the moment, for Plantinga's notion of a defeater. Now we need to relate the claim that N&E is a defeater for R to the conclusion that Plantinga is primarily concerned to establish—namely, that one who holds N&E cannot rationally hold R (and, as a further conclusion, cannot rationally hold any belief at all, including N&E). Though Plantinga does not sufficiently underline the point in the published statement of the argument in *WPF* (and in its predecessor in Radcliffe and White's *Faith in Theory and Practice*), N&E's being a defeater for R is not by itself sufficient for the further conclusion. For one thing a defeater can itself be defeated (*WPF*, 233; ND, 12). Defeaters, including defeaters for defeaters, can be divided into *rebutters* and *undercutters*. The former is what is defined by D\*. 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. Since a defeater is a belief, it might, like any belief, be itself defeated in either of these ways. If it is, its efficacy as a defeater is thereby nullified.

Plantinga has another way of allowing that N&E's being a defeater for R may be insufficient to render it irrational for S (who accepts N&E) to accept R. S might have such strong support for R as to outweigh the negative impact of N&E (*WPF*, 233). Plantinga is sufficiently alive to the necessity of both these conditions to devote several pages of *WPF* to arguing that one can't have non-question-begging support for R, and similarly that any supposition that N&E is itself defeated is also question begging, on the grounds that any such suppositions presuppose R, that is, assume that our faculties are generally reliable.

By virtue of treating *no defeater defeaters* and *no sufficient independent evidence for R* as requirements that are additional to N&E's being a defeater for R, Plantinga runs into further internal difficulties. For note that where such requirements are not satisfied, it is by virtue of features of S's noetic structure. And, as we have seen, D\* implies that features of the noetic structure are relevant to whether our ideal subject would resolve the D-B conflict by rejecting B, i.e., are relevant to whether D is a defeater for B. So how can they be requirements additional to D's being a defeater for B? I don't take this to be nearly as serious a problem for Plantinga as the earlier problem as to whether relative degree of warrant counts as to what the most rational resolution of the conflict would be. It is almost just a matter of bookkeeping. If they are extra requirements, then the definition of 'defeater' will have to be modified in such a way that they are not requirements for D's being a defeater for B. Otherwise, we can just bill them as extra requirements.

But there is a bit more than convenience of bookkeeping involved in this choice, which gives us reason to prefer the alternative of taking these requirements to be necessary for defeater status, as  $D^*$  (on my reading) implies. There are three advantages of this choice.

1. As just noted, this permits the retention of  $D^*$  as a definition of 'defeater'.
2. Furthermore, it fits better the intuitive notion of a defeater. Plantinga formulates this in several places.

... the basic but rough answer is that defeaters are reasons for changing one's beliefs in a certain way. (ND, 20)

... the basic idea is that when S acquires a defeater for B, she acquires a reason for modifying her noetic structure in a certain way. (ND, 31)

Acquiring a defeater for a belief puts you in a position in which you can't rationally continue to hold the belief. (WCB, 359)

The first two formulations are very unspecific. Modify beliefs in *what* way? And how strong a reason? But the third formulation ties things down. The change is a deletion of the belief for which the defeater is a defeater. And the reason is strong enough to make it irrational to continue to hold that belief. To fit that specification we need the "extra" requirements.

Apart from Plantinga's delineation of the intuitive idea, it must be noted that 'defeat' is a term of considerable epistemic strength. It seems unintuitive to call  $D$  a defeater for  $B$  if its defeating efficacy is overridden by a higher-level defeater or by a stronger support for  $B$ . In that case, we would have what might more felicitously be termed a '*prima facie* defeater' or a 'candidate defeater'. We need the extra requirements to get the full epistemic force the term 'defeater' suggests.

3. Reserving 'defeater' for what satisfies the "extra" requirements enables us to avoid an uncomfortable shift if one or another of these requirements are not satisfied. Suppose being undefeated is not required for  $D$ 's being a defeater of  $B$ . Then it turns out that  $D$  is itself defeated. Is it still a defeater of  $B$ , albeit a defeated one, or does it lose that status? If the defeater of  $D$  is added to the noetic structure after  $D$ , we can think of that addition as changing  $D$ 's status from being a defeater to not being a defeater. But what if the defeater of  $D$  was there all along, but unnoticed? Or it was noticed but its defeating of  $D$  was not? In that case, did  $D$  defeat  $B$  prior to the noticing? We have a similar problem with strong support for  $B$  that turns up or comes to be realized at a later point. Building the "extra" requirements into what it takes for a belief to be a defeater, as in  $D^*$  on my reading, will avoid these problems.

It is time for a review of my interpretation of Plantinga's account of defeaters. First, I will ignore indications that Plantinga does take relative warrant of  $D$  and  $B$  (and perhaps other "external" factors as well) to be relevant to

the question of whether D defeats B, except where explicitly noted to the contrary. I will be working with the reading of  $D^*$  that restricts “relevant respects” and features of the noetic structure to matters that are relevant to “internal rationality”. But, as noted, I will take the noetic structure to be organized by objective logical and probability relations between beliefs, relations that may or may not be noted by S. This means, inter alia, that positive and negative “support” for D and B from other beliefs and from experience are relevant to the reaction of our ideal subject to the D-B conflict, and hence to whether D is a defeater of B. Moreover, this support is to be evaluated not only by the number and content of supporting beliefs but also by the epistemic status of these beliefs, *insofar as it is a function of the relation of the belief to other beliefs and to experience*. Thus an (internally) rational response to a conflict is determined by the *internal economy* of the noetic structure, so understood. Note that this implies that although a defeat relation so understood is of undoubted epistemic significance (it does constitute a reason for abandoning a belief), its epistemic significance is limited, so long as we ignore the relative degrees of warrant of constituents of the noetic system and other external influences on epistemic status. I will return to this last point when I move to critical remarks on this account of defeat and on Plantinga’s use of it in his evolutionary argument against naturalism.

## VII

It is high time to move away from the general notion of defeat and look more specifically at Plantinga’s claim that anyone who holds N&E thereby has a defeater for R (and, by derivation, for any belief whatever, including N&E). Remembering Plantinga’s disjunction of two views as to the probabilistic relation between N&E and R, I will be working with the *low value of  $P(R/N\&E)$*  alternative, adding some brief remarks on the *inscrutability of  $P(R/N\&E)$*  alternative at the end of the chapter. Remember also that Plantinga argues that what he takes to be the extra requirements of *N&E’s not itself being defeated* and of *the lack of (sufficient) independent evidence for R* are satisfied. Later I will suggest some doubts about the latter claim, but for now I will assume Plantinga is right in both cases, thus allowing him those parts of what it takes, on my preferred bookkeeping, for N&E to defeat R.

To be sure, if I were to concede to Plantinga in this case everything that  $D^*$  requires for a defeater relation, I would pass on to the question of the significance of N&E’s defeating R, in this sense of the term. But I am by no means prepared to do that. The main worry is this. For this stage of the argument I also concede a low value of  $P(R/N\&E)$ , despite the doubts to which I alluded in section 3. But the considerations that support that result equally support a low value for the reverse relation,  $P(N\&E/R)$ . For if the naturalistic evolutionary story of the development of human cognitive faculties makes it un-

likely that they are reliable, then, by the same token, if they are reliable that makes it unlikely that they arose in the naturalistic evolutionary way. If certain meteorological conditions make a thunderstorm unlikely, then if a thunderstorm occurs it is unlikely, in the absence of other relevant information, that those meteorological conditions prevailed. So the question is: Why take it that N&E defeats R, rather than that R defeats N & E? Or in terms of D\*, why suppose that a subject working under the conditions specified in D\* would withhold R, rather than withhold N&E? There could be other grounds for Plantinga's preference. R could be defeated by other parts of the noetic structure and N&E not. But Plantinga gives no hint of this, and let's assume that neither is otherwise defeated. Or one or the other could have much stronger independent support within the noetic structure than the other. As noted, Plantinga argues that it would be question begging in this context to suppose that R enjoys strong independent support. And though Plantinga does not discuss the issue in these texts, he does not believe that the N conjunct at least of N&E enjoys strong independent support. But rather than turn my argument into an ad hominem, let me just say that unless N&E enjoys significantly stronger independent support than R, there does not appear to be, within the terms set by Plantinga, any reason to prefer taking N&E to defeat R rather than taking R to defeat N&E.

## VIII

But perhaps the game Plantinga is playing has been misconstrued. So far I have not cited passages like the following (italics added).

A defeater for a belief *b*, then, is another belief *d* such that, given my noetic structure, I cannot rationally hold *b*, *given that I believe d*. (WCB, 361; italics added)

... you have a defeater for one of your beliefs B just if you acquire another belief D such that, *given that you hold that belief*, the rational response is to reject B. (WCB, 366; italics added)

... no noetic structure to which S can rationally move (*given that she accepts D*) will contain B. (ND, 33; italics added)

What these formulations suggest is that the question of what defeats what (in terms of D\*, what the most rational response to the D-B conflict would be) is one that is raised in a certain context, a context that is defined by S's holding D. Since that is a presupposition of the question being raised, part of the content of the question itself, the rejection of D is not one of the options between which a choice is to be made. Applying this point to our central concern here, the question of what rationality requires is raised for the person

who accepts N&E. The question is as to what rationality requires for a person who continues to satisfy that description. Hence there is no problem as to what reasons there are for a preference between D's defeating B and B's defeating D. The latter alternative has been blocked out before we get started.

Perhaps that is the way Plantinga is thinking of the matter. But if so, this markedly affects the epistemic significance of the question he is raising and hence the epistemic significance of the answer to the question. The question is not without significance. It is sometimes relevant to consider what a person is committed to, on pain of irrationality, given that he accepts a certain proposition. This can be useful in persuading him to give up the proposition in question. Once he sees how the proposition would render his belief system seriously incoherent or inconsistent, he may give it up. Presumably something like this is involved in Plantinga's argument. Addressed to the proponent of N&E, it can be construed as an appeal to such a person to take seriously the dismal situation that confronts him if he continues to hold fast to N&E. Either he must deny that his cognitive faculties are reliable, in which case he is without adequate rational basis for any beliefs whatever, or he must suffer crucial incoherence in his belief system.

But how about defeat, as defined by D\*? Does this "rationality relative to holding N&E fixed" interpretation enable us to conclude that for S (an advocate of N&E) N&E is a *defeater* of R in the sense defined by D\*? Again, this depends on how "relevant respects" are specified, and on what is included in the description of S's noetic structure. What's the best we can do along this line? Well, suppose we go along with Plantinga in supposing that N&E itself is undefeated and that there is no significant non-question-begging support for R. (Remember that we have decided to include that in what is required for N&E's defeating R.) But now what about the hypothetical subject's cognitive faculties functioning proper in the *relevant respects*. I have already pointed out that there would seem to be no sensible alternative to allowing the relative epistemic status of N&E and R to be relevant to determining which, if either, defeats the other. How could their status not be relevant? Even if we restrict ourselves to the "internal rationality" alternative, we will at least have to count as relevant the intrinsic evidence of each (proper basicity) and the confirmation and disconfirmation each has from within the subject's noetic structure. But then it seems that we are still driven to the former conclusion. If N&E is no better supported internally than R, the situation still fails to tell in favor of N&E's defeating R, rather than the reverse, if each has a low conditional probability on the other. Otherwise put, if that is the case, then unless N&E is significantly better internally supported than R, we cannot conclude that a subject whose cognitive faculties are properly functioning in the relevant respects would reject R, rather than rejecting N&E.

By the way, I can easily imagine Plantinga's welcoming this conclusion. For it provides the basis for a strong argument against N&E. "Either N&E is woefully undersupported by evidence, or it leads to self-refuting conse-

quences. In either case, it is irrational to hold it.”<sup>14</sup> If he would like to appropriate this in the next version of his argument, I make him a free gift of it. But my concern in this chapter is to critically evaluate the argument as he has presented it.

To continue, if my conclusions thus far are sound, the only way Plantinga can derive the conclusion that the ideal cognitive subject will reject R, assuming that N&E lacks significantly stronger internal support, is to “cook” the account of relevant respects in such a way as to reflect the “retain no matter what” status of N&E. That would amount to presenting the above argument as to what the internal coherence of one’s belief system commits one to if one accepts N&E, in the guise of an argument for what an ideally rational subject would do about the N&E–R conflict in the light of all features of its noetic structure that are relevant to that resolution. But those clothes simply do not fit the “internal coherence of belief system with N&E fixed” argument. They are quite different arguments, quite different trains of reasoning, suitable for different purposes, and it is false advertising to present the former as if it were the latter. Whereas the former is suited to convincing (or trying to convince) an N&Eer of the error of his ways, the latter is suited to consider how a subject, ideally rational in the relevant respects would respond to a probabilistic inconsistency between N&E and R. To repeat, if we give a sensible reading of “relevant respects”, we simply can’t conclude that N&E defeats R (in the sense of ‘defeat’ given by D\*) unless N&E is significantly better supported internally than R.

Here is another way of seeing this. Consider a favorite example of Plantinga’s, Descartes’s reference to a madman, M, who thinks his head is made of glass. Plantinga uses the case of M to illustrate how a radically irrational belief can render certain responses rational. If M also believes that wearing a football helmet at all times (like B.D. in *Doonesbury*) will protect his glass head from shattering, it is rational for him to wear a football helmet at all times. There is a sense of ‘rational’ in which the rational thing for M to do, *given that he has those beliefs*, is to wear a football helmet at all times. And in that same sense, given the assumptions Plantinga is making, it is rational for one who holds N&E to give up R. But this sense is too subjective to have any important epistemic significance. This can be seen if we consider this non-contextually delimited question (i.e., not limited by the context of what M in fact believes): Is it rational for M to believe that whenever he takes off his helmet his head is in danger of shattering? Of course not. It is a paradigm of irrationality, and just because of the extreme irrationality of the belief from which it was derived. Similarly, if N&E is no better supported for S than R,

<sup>14</sup> The attentive reader of *WPF*, chapter 12, will recognize that this argument comes close to being a combination of Plantinga’s “preliminary argument” (228–229) and “main argument” (beginning on 229), combining the strong points of each and avoiding their weaknesses.

then there is no rational basis for a preference for withholding R, rather than withholding N&E.

I should note that Plantinga says things that might be taken to support his raising the question “What defeats what?” in a context defined by the acceptance of N&E. In particular, he says, “In the typical case of defeat, I will first believe *b* and then later come to believe the defeater *d*” (*WCB*, 361). If he supposes this supports his procedure, he would be assuming that the temporal order of acquisition of beliefs makes an important difference to the rationality of one or another response to their incompatibility. But why suppose that? If I assumed anything along this line, which I don’t, it would be that the older belief has the right of prior possession and, by a principle of conservatism, that if everything else is equal preference goes to the already established. In any event, I can’t see that the fact that D is the new boy on the block gives it the right to displace B rather than vice versa, if there is no reason in comparative epistemic status to chose one response rather than another.

Finally, at the risk of undue repetition, we have already seen that Plantinga does allow comparative epistemic status (in the Aberdeen case and elsewhere) to affect the choice as to what to give up, at least where there is a significant difference in epistemic status. And reading “relevant respects” in D\* in the light of those cases, why then should this (approximately) equal epistemic status not also have a bearing—in this case leaving us without rational grounds for preference, and hence leaving it ambiguous which, if either, “defeats” the other. And if it is necessary to make a choice nevertheless, why adopt the principle of always holding the newcomer fixed? Flipping a coin would have more to commend itself than that!

I think I have complained enough about Plantinga’s overly subjective notion of rationality, overly subjective because too tied to what S believes, whatever the objective epistemic status of those beliefs. Now I want to point out some severe limitations of his argument vis-à-vis the conclusion he is aiming at, even if we work with his “given that S believes that D” notion of rationality. The conclusion aimed at is that an advocate of N&E cannot rationally accept R, since the former is a defeater for the latter. We can see some roadblocks in the path to this conclusion by reminding ourselves of some of the things that Plantinga recognizes as necessary for the conclusion.

1. First, the conclusion will follow only if the defeater is itself undefeated. (Or on my preferred reading, the alleged defeater will be a defeater only if undefeated.) Of course, if N&E is defeated, that will be no consolation to its advocates; and Plantinga can use this as one horn of the dilemmatic argument against N&E I suggested for him earlier. But it will affect the conclusion that one who holds N&E thereby has an (undefeated) defeater for R. And if, as Plantinga holds and as I hold, N&E is woefully undersupported, this is a live possibility, given Plantinga’s views. For Plantinga holds that “with respect to some sorts of beliefs, what gives me a defeater for a belief of that sort is just



the fact that I don't have a reason for it; that realization is itself a defeater for the belief" (ND, 27).<sup>15</sup> I take it 'have a reason' here is to be understood as 'have a sufficient, adequate, strong . . . reason'. Moreover, N&E is certainly a thesis for which one needs sufficient *reasons* if it is to be rationally held. It could hardly be considered to be self-evident or anything like that. And, if Plantinga and I are correct, N&E is defeated in this way (or could be if we add Plantinga's requirement that S realizes the lack). And so this is one way in which N&E could fail to defeat R, on Plantinga's own showing.

2. Plantinga opines that even if B has a low probability on D, this is not sufficient for defeat unless S sees this D-B relationship (ND, 36–37). By imposing this requirement Plantinga makes the (alleged) probability relationship useless for reaching his intended conclusion. For presumably very few N&Eers will think that  $P(R/N\&E)$  is low. Indeed, Plantinga's argument is set up in opposition to some very prominent N&Eers who affirm and argue for the opposite thesis. So with the above requirement N&E will defeat R only for those N&Eers (presumably a distinct minority) who agree with Plantinga's probability assessments. Note that if Plantinga were working with a more objective conception of rationality and hence did not make so much hang on what beliefs S actually has, he could take the objective probability relation to yield the defeat relation (if it does so) without the need to require that S believe that the probability relation holds.

3. Then there is the question of whether R has considerable support within typical noetic structures. Plantinga argues that any argument for R, indeed any argument for anything, will presuppose R, and therefore will be question begging (in a way I term 'epistemic circularity') (*WPF*, 233–234). But there are two complications here. First, I have argued elsewhere that one can have strong reasons for a reliability of faculties claim even if those reasons are the output of faculties for whose reliability they are claimed to be reasons.<sup>16</sup> But I won't pursue my reasons for this obviously counterintuitive claim here. Instead I will defend the possibility that R enjoys "basic warrant", to use Plantinga's term of choice. 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. The claim that R has this status is the core of one of the objections Plantinga considers in ND, the one labeled "R Beyond Defeat". His response to that objection is not that R doesn't have warrant in this way,<sup>17</sup> but that it doesn't follow from R's having basic warrant that I cannot acquire a defeater for it. I agree with

<sup>15</sup> Note that, in accordance with his usual practice in this discussion of placing a lot of weight on what the subject in fact believes, Plantinga specifies the defeater as the *realization* that one lacks an adequate reason. I would prefer to be more objective and say that the lack of an adequate reason, whether the subject realizes it or not, could be a defeater. But let that pass.

<sup>16</sup> See Alston, "Epistemic Circularity," in *Epistemic Justification* (Ithaca, N.Y.: Cornell University Press, 1989).

<sup>17</sup> "Let's also agree that R does have warrant and perhaps a great deal of warrant, when it is taken as basic" (ND, 51).

Plantinga about that. But the crucial point is that if R does have “a great deal of basic warrant”, then one would have to show that there is a defeater for it in spite of that, in order to clear the way for the conclusion that N&E defeats R. And Plantinga has not done that.<sup>18</sup>

## IX

The only item remaining on the agenda is the other part of Plantinga’s low probability–inscrutability alternative. Suppose that we are unable to determine the value of  $P(R/N\&E)$ . Is that sufficient, along with other relevant features of the situation, to make N&E a defeater of R?

Before tackling this question, I need to precisify the inscrutability alternative. If the claim is only that we are unable to arrive at a well-grounded precise value, or even an informative range of values, that is too obvious to mention. Presumably what is intended is that we are much more at a loss than that, that we are unable to decide even between very low, somewhere in the middle half, or very high.

So what about the alternative, thus construed? Since this chapter is already so long, since the bearing of the inscrutability alternative raises many difficult questions, and since I do not at the moment see my way clear through all of them, I will limit myself to two points.

First, on the inscrutability assumption we don’t have even the *prima facie* case for defeat that we have on the low probability assumption. If I don’t know what to say about how likely or unlikely N&E makes R, how does that engender even a *prima facie* tension between a belief in N&E and a belief in R? Suppose I believe both a certain version of quantum mechanics (Q) and general relativity theory (T), and I have considerable evidence for each. I am inclined to think that Q has some evidential bearing on R, but I am unable to determine what that is. (This may be because one or the other is insufficiently developed, or because the links between what they apply to has not been sufficiently worked out, or because I am not capable of spelling out the connections, or . . .) In those circumstances would I have even a *prima facie* case that Q is a defeater for T? I can’t see that I would. That would entail that unless my independent evidence for T outweighs my independent evidence for Q, then rationality requires that I give up the former. But surely that is not the case. Since I am unable to assess the probability of T on Q, I have no basis for thinking that the tenability of T depends on my having more evidence for it than for Q. What does that have to do with the rationality of accepting T?

<sup>18</sup> There is also the *ad hominem* point that Plantinga’s “preliminary argument against naturalism” uses R as a premise in a Bayesian argument against naturalism (*WPF* 228–229). The fact that he feels free to do this, despite the epistemic circularity involved in any argument for R, indicates that he does not take that epistemic circularity to prevent R itself from occupying whatever favorable epistemic status is required for being properly used as a premise in arguments.

If the probability of T on Q were low, then that requirement would make sense. But when we are unable to say what that probability relation is, it would seem sheerly arbitrary to say that T has to be given up if the evidence for it is less strong than the evidence for Q.

Second, even if there were such a *prima facie* case for rejecting T, to make it more than *prima facie* it would have to be the case that T does not enjoy strong independent support, contrary to the way the case was described. But, Plantinga might well say, that is just where the N&E–R relationship differs from this case. Indeed, in giving his argument in *WPF* on the inscrutability alternative, he explicitly makes it dependent on the assumption of no independent support for R. After arguing for rationality's requiring giving up R for his "hypothetical population", he continues:

But now suppose we again apply the same sort of reasoning to ourselves and our condition. Suppose we think N&E is true: we ourselves have evolved according to the mechanisms suggested by contemporary evolutionary theory, unguided and unorchestrated by God or anyone else. Suppose we think furthermore, that there is no way to determine  $P(R/(N\&E\&C))$  (specified to us). . . . If we have no further information, then wouldn't the right attitude here, just as with respect to that hypothetical population be agnosticism, withholding belief [in R]? (*WPF*, 229)

Thus Plantinga himself makes his conclusion on the inscrutability alternative depend on the claim that there is no (sufficient) independent support for R. And, again, as pointed out in discussing the low probability assumption, he has not come close to establishing that. Hence here too the argument fails.

## X

By way of summary: (1) The claim that the  $P(R/N\&E)$  is low is poorly supported, as Plantinga himself admits. Taking that value to be "inscrutable" seems more reasonable, but there is no clear relevance of that to the defeat claim. (2) Even if the value of  $P(R/N\&E)$  is low, it doesn't follow that N&E is a defeater of R, in any epistemically important sense of 'defeater', unless R fails to enjoy greater warrant than N&E. And it seems plausible to suppose that it does, and in any event Plantinga has failed to show that it does not.

All this leaves Plantinga's argument, as he has presented it, without the intended force. It would still be possible for him to argue that one who insists on holding on to N&E cannot preserve internal rationality except by giving up R. And perhaps that would give him the heart of what he is after. But it

doesn't give it to him in the way in which he tried to reach that result, by arguing that N&E is a defeater of R for one who holds both, nor does it yield the epistemically objective conclusion that internal rationality requires for anyone (regardless of their subjective preferences) that a conflict between N&E and R be resolved by giving up R.