Plantinga's 'Defeat'

Edward Wierenga

University of Rochester

(draft . . . January 24, 2011 . . . not to be cited . . . comments welcome and appreciated)

Alvin Plantinga accepts much of the standard view about defeaters. For example, he endorses John

Pollack's distinction between a *rebutting* defeater and an *undercutting* defeater (Pollock 1974 and 1986).

The former, roughly speaking, is a proposition reporting new evidence that counts against a

proposition one believes. The latter is a proposition reporting new evidence that undermines one's

justification or evidence for a proposition one believes. And Plantinga accepts Pollock's further

claim that defeaters can themselves be defeated, that is, there are defeater-defeaters. Indeed,

Plantinga holds that "this defeater structure is to be found across the length and breadth of our

cognitive structure," and he adds that "the defeater system works in nearly every area of our

cognitive design plan and *is a most important part of it . . .*" (Plantinga, 1993*b*: 41; emphasis added). In

several places Plantinga even gives formal and somewhat complicated definitions[1] of various related

concepts of defeat (1994; 2000: 359ff.; 2002*b*: 208f.). Despite the intuitive description above, which

glosses defeat as the acquisition of new evidence resulting in the loss of evidential support for a

belief, Plantinga's definitions are explicitly tied to his own *proper-functioning* account of warrant.

Interestingly, defeat defined in this way seems to play only a minor role in Plantinga's epistemology.

Instead, the main application he makes of the concept is in his "Evolutionary Argument against

Naturalism" (see his 1993*b*: 216-237; 1994; 2000: 227-240; 2002*a*; 2002*b*), where he argues that

---

[1] Strictly speaking, what Plantinga offers are statements of necessary and sufficient conditions for one belief defeating another. But he describes these statements as *definitions* (2000: 363; 2002*b*: 208f.), so I will follow him in this.

evolutionary naturalism yields a defeater for anyone who accepts it (and appreciates a certain argument) for the proposition that his or her cognitive faculties are reliable. A full assessment of Plantinga's account of defeat, accordingly, will have to consider its use in the anti-naturalism argument. I will begin by summarizing Plantinga's proper-functioning account of warrant, and then I will present his definition of defeat with its tie to proper functioning. After identifying a few problems with this understanding of defeat, I will ask whether we can get a better understanding of the concept by considering its use in the Evolutionary Argument against Naturalism. I will conclude that not only does its use in this argument fail to illuminate Plantinga's concept of defeat, but the Evolutionary Argument against Naturalism appears less convincing when defeat is construed in terms of proper functioning than it would on a more traditional account.

1. Warrant

Since Plantinga's account of defeat is couched in terms of his concept of warrant, we should begin with a brief review of warrant. Plantinga's initial description of warrant is often something like, "it is that which distinguishes knowledge from true belief" (1993*a*: 3), although he usually goes on to say, somewhat more precisely, that it is "that property–or better, *quantity*–enough of which makes the difference between knowledge and mere true belief" (2000: xi). Stated in this way, it is tempting to think that Plantinga is joining a familiar debate in the history of epistemology in the wake of the Gettier problem (Gettier, 1963). That problem was taken to show that something must be added to justification, truth, and belief in order to give an adequate account of knowledge. Thus, responses

to the Gettier problem offered analyses of knowledge that took the following form:

(D)   *S* knows that *p* iff

(i) *S* believes *p*,

(ii) *p* is true,

(iii) *S* is justified in believing *p*, and

(iv) [the fourth condition].

Various proposals were made as to how to specify the fourth condition, including counterfactual conditions, causal conditions, and, for present purposes the most relevant, *defeasibility* conditions.[2] On this last approach, the fourth condition would be something like

(iv) *p* is undefeated for *S*.

We need not consider the finer details of this earlier controversy, however, to notice that Plantinga's account of knowledge may be represented in a similar way:

(W)   *S* knows that *p* iff

(i) *S* believes *p*,

(ii) *p* is true, and

(iii) *S* has (sufficient) warrant for *p*.

---

[2] For a nice survey of this literature, see (Shope, 1983).

So construed, it might be thought that Plantinga's *warrant* is simply the combination of justification and whatever additional requirement is needed to avoid Gettier counter-examples, that warrant has a dual role to play.[3] That would be mistaken, I think, for two reasons. First, Plantinga takes himself to be offering an alternative to accounts of knowledge that appeal to justification, holding instead that "justification is a fine thing . . . but it is neither necessary nor sufficient for warrant" (1993a: 45). In his view justification has a deontological component, one that is too easily satisfied by one who does her best to flout no intellectual duties (2000: 101; 1993*a*: 6ff.), and warrant involves no such obligation. Second, his account of warrant in terms of proper functioning seems not to be closely connected to the concept of evidence central to many treatments of justification. Here is one of Plantinga's summaries of the theory:

> Put in a nutshell, then, a belief has warrant for a person *S* only if that belief is produced in *S* by cognitive faculties functioning properly (subject to no dysfunction) in a cognitive environment that is appropriate for *S*'s kind of cognitive faculties, according to a design plan that is successfully aimed at truth. We must add, furthermore, that when a belief meets these conditions and does enjoy warrant, the *degree* of warrant it enjoys depends on the strength of the belief, the firmness with which *S* holds it. (2000: 156) (See also (1993*b*: Chs. 1, 2), (1995), and (1996).)

---

[3] Cf. Ryan (1996: 184) who holds that "an account of warrant should include a justification condition and a condition that rules out Gettier cases."

A belief has warrant for a person just in case that belief is produced by a cognitive process, for example, perception or memory or a faculty that responds to testimony, provided that the process is functioning properly, according to a design that is successfully aimed at producing true beliefs. In addition, the process in question must be operating in the sort of circumstances for which it was designed to operate; for example, in the case of perception the circumstances include proper lighting on medium-sized objects in the near vicinity. A fuller presentation of the account would develop each of these listed conditions (and add a few more to avoid objections).

For our purposes, however, the following capsule summary will suffice: a belief is warranted for a person provided that it is produced by that person's properly functioning cognitive faculties successfully aimed at truth, operating in circumstances where they were designed to operate.[4]

2. Defeat

Plantinga distinguishes two types of defeaters, *warrant* defeaters and *rationality* defeaters (2000: 359), although he does not spend much time discussing the former. It is clear that he thinks that a warrant defeater for a belief can leave the rationality of that belief unscathed. His only example suggests that a warrant defeater need not be believed and that it is a proposition to the effect that

---

[4] Since warrant is defined on *beliefs*, it is not quite accurate to say that it is whatever must be added to true belief to obtain knowledge; more simply, it is whatever must be added to truth to get knowledge. At least one commentator (Merricks, 1995) thinks that warrant entails truth. If that were right, then warrant is not something that is added to truth, either–it would just *be* knowledge.

one of the more recondite conditions on warrant (one we left out of the capsule summary above) is not satisfied.

On the other hand, he says that undercutting and rebutting defeaters are rationality defeaters (2000: 359), so let us focus on them. Plantinga introduces rationality defeaters informally as, "a defeater for a belief *b* . . . is another belief *d* such that, given my noetic structure, I cannot rationally hold *b*, given that I believe *d*" (2000: 361). The relativity to "noetic structure" is important; it is possible for a belief in a proposition to be a defeater for one person's belief but not for the same belief held by another. In Plantinga's example

> . . . you and I both believe that the University of Aberdeen was founded in 1495; you but not I know that the current guidebook to Aberdeen contains an egregious error on this very matter. We . . . both read it; sadly enough, it contains the wholly mistaken affirmation that the university was founded in 1595. Given my noetic structure (which includes the belief that guidebooks are ordinarily to be trusted on matters like this), I thereby acquire a defeater for my belief that the university was founded in 1495; you, however, knowing about this improbable error, do not. The difference, of course, is with respect to the rest of what we know or believe: given the rest of what *I* believe, I now have a reason to reject the belief that the university was founded in 1495; the same does not hold for you. (2000: 360)

In his (2000), where an extended discussion of defeaters is found, Plantinga says very little about noetic structures. The most detailed published discussion of this concept is found in his (1983).

There he holds that "a person's noetic structure is the set of propositions he believes, together with certain epistemic relations that hold among him and these propositions" (1983: 48). Three such epistemic relations are identified (1983: 49-52). The first is the relation of believing one proposition *on the basis of* one or more other propositions. Typically, some propositions a person believes are believed on the basis of other propositions the person believes, and a person's noetic structure thus includes information about how his or her beliefs are related in this way. The second feature of a noetic structure is that it includes an index of *degree* of belief, that is, for each belief a measure of how *firmly* the person holds that belief. Finally, a noetic structure includes an index of *depth of ingression*. The idea here is that some beliefs are on the "periphery" of a person's noetic structure and could be given up without requiring much of a change in the rest of the structure. Other beliefs, however, are so centrally held–they have such a depth of ingression–that they could not be reasonably abandoned without requiring wholesale revision of much of the rest of a person's beliefs. Later, in his (1994), Plantinga adds that noetic structures include, in addition to beliefs, experiences and other cognitive states, such as doubts and fears. Furthermore, noetic structures include information about what various beliefs are formed in response to. He suggests leaving the notion at an "intuitive level", although the more detail that is added to the account, the less plausible it becomes, as I hope will become clear below, that two people ever have the same noetic structure.

Plantinga's example given above of how a belief in a proposition could be a defeater for one person's belief but not for another's depends on people differing with respect to which other propositions they believe. But with this sketch of noetic structures, we can see that if it is possible for two people to believe the very same propositions but have different noetic structures (if they differ in their basing relations, the firmness with which they hold their beliefs, or the depth of

ingression that their beliefs have for them), then there is another way in which a belief in a proposition could be a defeater for one person but not another. For example, if two people believe the same propositions, but one of them believes p on the basis of q and the other believes p on the basis of r and s, then if they both learn that q is not a good reason for p, the former would seem to have a defeater for p, but the latter would not.

As we have seen, Plantinga's informal characterization of a defeater for a belief b is that it is another belief d such that, given one's noetic structure, one cannot rationally hold b, given that one believes d. He expands this with a formal definition of a *purely epistemic defeater*:[5]

(D*) D is a purely epistemic defeater of B for S at t if and only if

(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 beliefs and minimization of false beliefs) 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).[6] (2000: 363)

---

[5] In (2002*b*: 209) Plantinga uses the term "alethic defeater" for what he had earlier called a purely epistemic defeater.

[6] I have changed the formatting, but not the text, in order to display the various clauses of (D*) more explicitly.
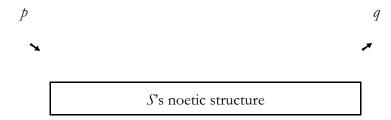
Jonathan Kvanvig (2007) calls this account of defeat a "back door" approach. He contrasts this account with his preferred conception, according to which defeater relations between propositions depend upon the evidential relations holding among them, and thus hold prior to being added to one's store of beliefs. If the defeater relations among propositions obtain antecedently of one's belief in them, then defeat enters the "front door" when one acquires the right (or wrong) beliefs. By contrast, on Plantinga's proposal, whether one proposition is a defeater for another depends on what happens when one adds it to one's stock of beliefs. As Kvanvig puts it, "the account characterizes defeat in terms of epistemically appropriate responses to the presence of a defeater in a noetic system: We insert the defeater into the noetic house, and see which belief gets expelled out the back door" (2007: 111).

I prefer to think of the back door as the output of a black box. On this way of construing Plantinga's account of defeat, whether p is a defeater for q (for a person at a time) depends on what would happen if p were added to that person's noetic structure (and it was functioning properly). If q comes flying out, then p is a defeater of q, as in the following figure:

Noting that Plantinga allows defeaters to be experiences as well as propositions, Kvanvig (2007: 111) revises (d) to

(d*)   who comes to *experience D or to* believe D but nothing else independent of or stronger than D, would withhold B (or believe it less strongly)

and suggests a parallel modification of clause (1).

If the required adjustment in noetic structure is to believe B less strongly, then D is a *partial defeater* of B. In (1994: n. 40) and (2000: 362) Plantinga suggests that being a defeater is a special case of being a partial defeater, on the assumption that coming to withhold is a special case of believing less strongly.

p q

↘   ↗

S's noetic structure

I propose the model of a black box because it is difficult to be sure of exactly what the proper functioning does *inside* a noetic structure.[7] One would hope that it would be factoring in the evidential relations holding among the propositions included in the structure, together with some attention to their basicality, degree of firmness, and depth of ingression features. Elsewhere Plantinga describes a case of someone for whom "an objective look at the evidence would [dictate] . . . a change of mind" (2000: 150), and perhaps proper functioning within a noetic structure would involve such assessment of evidence. Without further details, however, I think we will have to try to understand proper functioning in an intuitive way, to the extent that is possible.

Various objections can be raised against this account of defeat. But first I want to consider an objection that does not seem successful. Considering it will help us understand how the theory is supposed to work. Kvanvig (2007: 11-115) thinks that Plantinga cannot account for defeater-

---

[7] Plantinga has extensive discussions of the role of proper functioning in the *acquisition* of belief, whether through perception, memory, testimony, the *sensus divinitatis*, or another faculty, but, as far as I know, he does not provide the same depth of treatment to the *revision* of belief. One hint, however, is suggested by his remark that "a defeater (thought of my way) is in essence a reason for withholding a belief" (1994).

defeaters, that is, for defeaters that are themselves defeated. Let us consider an example of defeat

Plantinga himself gives:

> I see what looks like a fine bunch of tulips on your table; a whimsical bystander tells me that
> as a matter of fact there aren't any tulips there but only a cleverly contrived laser image; you
> tell me that there really *are* tulips there and that the whimsical bystander is only having his
> (rather lame) little joke. Then my circumstances include my being appeared to in that
> characteristic tuliplike way, having a defeater for the belief I initially form (that is, that I see
> some tulips), and a defeater for that defeater. The defeater-defeater isn't part of my *evidence*
> for the proposition that there are indeed tulips there; it is nonetheless a proposition I believe,
> which is such that if I hadn't believed it (or something similar), then in those circumstances
> the belief that I see tulips would not have had warrant for me. (1993*b*: 185)

At the outset, Plantinga is appeared to in a tuliplike way and believes, as a result of the proper

functioning of his perceptual faculty,

(p) I see some tulips.

Then, at a later time, say, t_1, Plantinga hears the testimony of the whimsical bystander and, through

the proper functioning of his faculty of credulity[8], adds

(q) There are no tulips on the table but only a laser image of tulips

---

[8] See his (1993*b*: 77ff.) for Plantinga's discussion of testimony and credulity.

to his set of beliefs. At this point, presumably, his properly functioning cognitive faculties detect a contradiction in his beliefs, and, accordingly, toss p out of his noetic structure. So far, so good. What we have is that q is a defeater of p for Plantinga at t_1. Next, at t_2, you say that there really are tulips there and that the bystander is joking. Again, Plantinga's properly functioning credulity faculty produces in him the belief that

(r) There are tulips on the table and the bystander is merely joking.

This time his cognitive faculties go to work and expel the belief in q. With q out of the way, S's cognitive faculties, whether in response to S's continuing tulip-like experience, S's memory of that experience, or your testimony that there really are tulips, restore S's belief that p. So at t_1 q is a defeater for p, and at t_2 r is a defeater for q. Thus, r is a defeater-defeater, because it defeats a belief that previously was a defeater. This application of (D*) is diachronic: apart from the acquisition and jettison of belief in these propositions, there is no standing relationship between them that determines which are defeaters and which is a defeater-defeater. Adding and discarding beliefs takes time, so it is natural to expect, on this account, that being a defeater is relative to temporal factors. Indeed, the *order* in which some beliefs are acquired could influence which are defeaters (for a person at various times). These features may (or may not) be objectionable, but it does not seem to be the case that Plantinga is unable to accommodate defeater-defeaters.

A second objection, however, might require a modification of (D*). Recall that it requires that if D is a defeater for S's belief B at t, then (1) S's noetic structure N at t includes B, and S comes to believe D at t, and (2) any person $S*$ whose cognitive faculties are functioning properly in the

12

relevant respects and whose noetic structure is *N* and includes *B* and who comes to believe *D* would withhold *B*. I suspect that the quantifier in clause (2) is vacuous, or nearly so. If *S*'s noetic structure is at all typical, it no doubt includes propositions that no one else believes, for example, first-person propositions about her own past, propositions about personal matters that no one else knows about, propositions true of *S* uniquely and of which others are unaware. In other words, if *S* is like most of us, she believes some propositions that no one else believes. So no one else has *S*'s noetic structure *N*. In that case, the only available instantiation for the quantifier in clause (2) will be *S* herself. So whether *D* is a defeater of *B* for *S* depends simply on what *S* does, on whether *S* herself stops believing *B*, at least if *S*'s faculties are functioning properly. But if they are not, a worse consequence follows. If that bit of the design plan sustaining *S*'s belief in *B* is wishful thinking or excessive optimism or some other faculty not successfully aimed at truth, then no one, not even *S*, instantiates the quantifier in (2). In that case, clause (2) is trivially true. If *S* believes that she will win the lottery because of wishful thinking, any new belief at all will be a defeater for her. From an epistemic point of view, it would be a good thing to disabuse her of this belief, but it should not be this easy to find a defeater.

This objection looks for instances of 'any person *S**' whose cognitive faculties *are* functioning properly and whose noetic structure *is N*, etc. But that is to ignore the 'would' in the final clause of (D*). Presumably, then, the relevant condition is to be understood as counterfactual:

> For any person *S**, if *S**'s cognitive faculties were functioning properly, and *S** were to have *N* for a noetic structure, and *S**'s belief in *B* were to be sustained by a cognitive faculty successfully aimed at truth, and *S** were to come to believe *D*, then *S** would withhold *B*.

13

So understood, the fact that no actual person (or no actual person other than S) has N for a noetic structure and whose faculties are functioning properly, and all the rest, is irrelevant to whether D is a defeater of B for S. What is relevant is what any person *would* do, were that person to satisfy these conditions.

Emphasizing the counterfactual nature of (D*), however, gives rise to a different objection.[9] Plantinga allows that a belief a proposition can be a defeater even though it is not warranted or does not arise by proper functioning, and even if it is acquired irrationally (2000: 364-365). So suppose that S's belief in B is sustained by a cognitive process successfully aimed at truth but that S comes to believe D by some other kind of cognitive process, one not successfully aimed at truth. To be more specific, suppose that S has received a letter from England in the mail and believes, by way of a properly functioning cognitive faculty successfully aimed at truth,

(B') The envelope is unopened.

S is then gripped by a powerful feeling of certainty, inspired perhaps both by wishful thinking and an inflated opinion of his talents, and comes to believe

(D') The letter contains an invitation to me to deliver the Thomas Locke Lectures[10].

---

[9] Compare Kvanvig's (2007: 112) remark: "Those familiar with the sad track record of counterfactual theories will know that this account is not going to work; those who believe that counterfactual theories hardly ever work may despair or salivate, depending on their attraction to the backdoor approach. . . ."

[10] Somewhat less distinguished lectures than the John Locke Lectures, named after John Locke's somewhat less distinguished younger brother, Thomas.

Now any person who had a noetic structure containing B', who came to believe D' and whose cognitive faculties were functioning properly, would give up the belief that B'. A person with properly functioning faculties would come to believe D' only after opening the envelope and reading the letter it contained, and it would then be obvious to the person that the envelope was opened. Thus, according to (D*), D' is a defeater of B' for S. But surely it is not. D' gives S no reason at all to abandon B.[11]

Another objection exploits a different way in which counterfactual circumstances can change the very features we are interested in. Suppose, as before, that S's belief that she will win the lottery (B) is produced and sustained by the cognitive faculty of wishful thinking, a faculty notorious for not being successfully aimed at truth, but this time she comes to believe that wishful thinking is an unreliable source of true beliefs (D). To determine whether D is a defeater of B for S, we are to consider what someone with S's noetic structure, properly functioning cognitive faculties, and a belief in B that was sustained by a truth-conducive cognitive faculty would do upon coming to believe D. Well, someone whose belief in B was sustained by a faculty successfully aimed at truth would be unlikely to give up that belief upon acquiring a belief in the unreliability of some other source of beliefs, a source that she is not using. So it turns out that D is not a defeater of B for S after all, although one would expect it to be.[12]

---

[11] Thanks to my student, Yaroslav Shtengrat, who devised a structurally similar counterexample. This objection assumes that it is possible for someone with properly functioning cognitive faculties to have the same noetic structure as S. If that is not possible, clause (2) of (D*) is trivially satisfied.

[12] This objection also assumes that it is possible for someone to have the same noetic structure as S while having the belief that B sustained by a different faculty. If this is not possible, then we are back to a version of the original problem. If no one could have the same noetic structure but with the belief in B sustained by a faculty successfully aimed at truth, every newly

Thus, (D*) seems to be flawed as a definition of defeat. How serious is this? Perhaps the best we can do is retreat to the black box model: a defeater for a belief *b* is another belief *d* such that, given properly functioning cognitive faculties and my noetic structure, I no longer hold *b*, given that I believe *d*? But exactly what are those faculties doing before they expel *d*?In the next section we will look at two (unsuccessful) approaches to spelling out the details of defeat.

3. Two Dead Ends

We might be in a better position both to assess Plantinga's account of defeat and to develop a more precise statement of it if it played a larger role in his epistemology. One possible avenue would be to explore the role of defeat in the analysis of knowledge. If Plantinga were using the concept of defeat to give a more traditional theory of knowledge, we might be able to see how defeat would have to be conceived. That is, if Plantinga were analyzing knowledge as true belief that is justified and *undefeated*, then the constraints of an adequate account of knowledge might force some features of defeat.[13] But, as we saw above, this is not Plantinga's approach to knowledge; his concept of warrant as proper functioning is intended to be an alternative to that traditional approach.

---

acquired belief defeats a belief that is not sustained by a faculty successfully aimed at truth, even beliefs wholly unrelated to the defeated belief.

[13] Inexplicably, Beilby (2007: 127) adds a "no defeater" condition in his statement of Plantinga's account of knowledge, but there is no textual justification for this interpretation. It does seem to follow (subject to a qualification in n. 15) from the claim that a belief is warranted that it is not defeated, but as I claim below, this is not a condition that needs to be satisfied separately in order for a belief to count as knowledge.

A second, potentially more promising, avenue is to examine Plantinga's contention that Christian belief (at least for the typical Christian believer) is not defeated. Perhaps we could learn more about defeat by examining this claim. Plantinga holds that one way of producing a defeater for one of his beliefs is by giving a sound argument for its falsehood, an argument with premisses he both accepts and which are such that "once I see the conflict, rationality requires that I give up the prospective defeatee rather than the premisses" (2000: 367). He adds that a second way of providing a defeater is to "put me in a position where I have *experiences* such that, given those experiences (and given my noetic structure), the rational thing to do is to give up the purported defeatee." Plantinga then goes on to argue that neither projective theories of theistic belief (such as Freudian or Marxist claims about the origin of religious belief), nor the results of contemporary Biblical criticism, nor postmodern theories of truth, nor facts about religious pluralism, nor facts about the nature and distribution of suffering and evil in the world yield a successful defeater for theistic or Christian belief (2000: 367-499). In each case, Plantinga holds that the believer remains *warranted* in theistic or Christian belief because his or her cognitive faculties, especially the *sensus divinitatis* or the inner instigation of the Holy Spirit, are functioning properly in producing and maintaining that belief. In the face of evil, for example, "the very fact that [the person] continues in theistic belief is evidence that the *sensus divinitatis* is functioning properly to at least some degree in her, and in such a way that knowledge of the facts of evil does not constitute a defeater" (2000: 491). More generally, these putative defeaters are not actual defeaters for typical theistic or Christian believers. Such believers come to have their theistic belief through the proper functioning of their cognitive faculties and are thus warranted in them, and if they reasonably come to believe the

putative defeater, they nevertheless remain warranted in their theistic belief.[14] So the putative

defeater can be added to the noetic structure of someone with properly functioning cognitive

processes without the belief in God being ejected.

I have no quarrel with this account. The chapters I have summarized in this most cursory

fashion are full of insightful detail. Instead, I want to make two points. First, it is no part of

Plantinga's epistemology that one must first tell whether a belief is undefeated in order to tell

whether it is a case of knowledge. Rather, the crucial test is whether a belief is warranted, whether it

is produced and maintained by properly functioning cognitive faculties, and it if is, it *follows* that it is

not defeated. On this view, if a belief is defeated, it does indeed follow that it is not known; but that

is because if it is defeated either it is no longer believed or the subject's cognitive faculties are not

functioning properly. The concept of defeat does not play a role in securing knowledge.[15] So we

learn no more about exactly how it works by looking at Plantinga's claims that theistic belief is

undefeated. On the other hand, defeat does have a more prominent part to play in Plantinga's

evolutionary argument against naturalism, to which we turn our attention in the next section.

---

[14] I am simplifying here. In the case of some of the putative defeaters, Plantinga sees no reason to accept them, and in the case of the example in the text, Plantinga is careful to qualify the claim that the theist is warranted *at least if theism is true*.

[15] Michael Sudduth (2009: 89) claims in his presentation of Plantinga's view that "there is what we might call an 'internalist no-defeater condition' for warrant: S's belief that p is warranted (to degree N) only if S does not have an internalist type defeater for the belief that p." See also (Sudduth 1999: 171). My interpretation, in contrast, is that there is a "no-defeater *consequence*", but not a *condition* that must first be satisfied in order for a belief to count as warranted. Even this claim that being undefeated is a consequence of being warranted needs to be qualified: a *partial* defeater might lower the warrant of a belief but leave the amount of warrant above the threshold for knowledge. In this case, a proposition known to be true might still have a (partial) defeater. I have been ignoring the complications posed by partial defeat.

4. Defeat and the Evolutionary Argument against Naturalism

In several places Plantinga develops a remarkable objection to the twin theses of naturalism and standard evolutionary theory (1993*b*: 229-237; 1994; 2002*a*; 2002*b*). According to the former, there is nothing supernatural and, so, no such person as God or anyone like him. According to the latter, human beings, and thus their cognitive faculties, have developed by such mechanisms as natural selection and genetic drift working on random genetic mutation. The combination of evolutionary theory with naturalism means that the evolutionary forces are *unguided*. The sources of our cognitive faculties are thus, to use Sosa's term, *brute forces* (2002, 2007), not under the control of a being who "oversees" their development or "orchestrates" the course of evolution (Plantinga, 1994). Plantinga's claim is that naturalism and evolution are *self-defeating*. More exactly, Plantinga holds that evolutionary naturalism "furnishes one who accepts it" (Plantinga, 1993 CHECK CIT.) with a defeater for the belief that our cognitive faculties are reliable. And if it is a defeater for that belief, it is also a defeater for any belief that arises from those faculties. In particular, then, it is defeater for belief in evolutionary naturalism itself.

It will facilitate our discussion to introduce some abbreviations. Let

R = Our cognitive faculties are reliable.

N = Metaphysical naturalism is true.

E = Our cognitive faculties have developed by way of the (brute) mechanisms to which

contemporary evolutionary theory appeals.

Exactly how Plantinga's argument proceeds, however, is not entirely clear. Plantinga has a detailed and elaborate strategy for arguing for the truth of

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

This is to be read as the claim that the conditional probability of R, given N&E, is low or inscrutable. We need not consider Plantinga's defense of (1) in detail. His basic idea is that the evolutionary forces that would have produced our cognitive faculties if N&E were true would have been geared towards producing adaptive behavior, regardless of whether that includes acquiring beliefs that are true (2000: 231). I will not challenge this claim (although I can imagine others doing so). We can also leave the threshold for a probability's being low imprecise, but let us assume that it is below, or significantly below, .5, and a probability is *inscrutable* for someone, just in case that person has "no definite opinion at all as to what that probability might be" (Plantinga, 1994).[16]

A natural assumption, given Plantinga's extended defense of (1), is that it is a premiss of the argument against evolutionary naturalism. I do not think that this is right, however. The reason is that, as I understand the argument, (1) functions as a defeater for the evolutionary naturalist who accepts it. But a defeater need not be *true*; rather, it must be *accepted* by the person for whom it is a defeater. So Plantinga's extended defense of (1) is aimed, I think, at persuading the naturalist to

---

[16] There is a suggestion in the fuller text from which this passage is quoted that Plantinga might require believing that a probability is not high for its being inscrutable, but in a parallel passage in (1993*b*: 231) he equates inscrutability simply with agnosticism.

accept it, for then, Plantinga believes, the naturalist has a defeater for R, for N&E, and for everything else the naturalist believes. Here, then, is how I think that argument proceeds:

> (2)   For anyone who believes N&E and who comes to believe (1), (1) is a defeater for R.
>
> (3)   For anyone for whom (1) is a defeater for R, (1) is a defeater for any proposition he or she believes on the basis of his or her cognitive faculties.
>
> ∴   (4)   For anyone who believes N&E and who comes to believe (1), (1) is a defeater for N&E. (2), (3), and the plausible assumption that anyone who believes N&E believes it on the basis of cognitive faculties.[17]

This reconstruction, which takes the alleged defeater to be (1), fits well with Plantinga's claim that "the view that P(R/N&E) is low and the view that this probability is inscrutable [i.e., (1)] gives the naturalist-evolutionist a defeater for R" (1994; 2002*a*: 11).

On the other hand, this construal fits less well with Plantinga's frequent remark that N&E is *self-defeating*. If N&E defeats N&E itself, then *it*, and not some other claim about the conditional probability of R on it, is a defeater (of R and of any proposition acquired through one's cognitive faculties). Plantinga writes that N&E "furnishes one who accepts it with a *defeater* for the belief that our cognitive faculties are reliable–a defeater that can't be defeated. But then this conjunction [of N and E] furnishes a defeater for any belief produced by our cognitive faculties, including, in the case

---

[17] This assumption that anyone who believes N&E does so by way of cognitive faculties is dispensable. The argument could just as well be explicitly addressed only to such persons, rather than, say, people who acquire beliefs by way of post-hypnotic suggestion or brain tumor. Plantinga's target is someone whose beliefs are *prima facie* respectable; there is no need to develop an intricate objection to beliefs that are obviously defective.

of one who accepts it, N&E itself: hence its self-defeating character" (2002*a*: 2). If N&E is the

defeater, rather than (1) being the defeater, then the way it "furnishes" a defeater is simply by being

one. However, taking N&E to be the defeater does not fit either the formal definition (D*) of a

defeater or of the various informal statements, according to which a defeater is a *newly acquired* belief.

Plantinga's objection is directed at the person who already believes N&E. The belief that such a

person acquires, or at any rate, the belief of which Plantinga hopes to persuade such a person, is the

belief that (1). So I will take (1) to be the alleged defeater and the argument above to be a summary

of the Evolutionary Argument against Naturalism.[18]

---

[18] A compromise interpretation: perhaps the alleged defeater is the conjunction of (1) and
(N&E). This interpretation is suggested by Plantinga's account of Frege's noetic condition before
and after receiving Russell's famous letter of 1902 (Russell, 1967). Before receiving the letter
(Plantinga assumes) Frege believed both
      (a)   For every property or condition, there exists the set of just those things that have the
            property or display the condition, and
      (b)   There is such a property or condition as being non-self-membered.
After reading the letter, presumably Frege believed
      (c)   (a) and (b) together entail a contradiction.
Plantinga concludes, "we should say that Frege's defeater for [(a)] was not just [(b)], but [(b)]
together with the newly acquired belief that [(c)]. . . . Although rationality doesn't require seeing the
connection between (a) and (b), failure to recognize that [(b)] and [(c)] together imply the falsehood
of [(a)] would of course be irrational, pathological. But failure to see that [(c)] by itself is not" (1994;
cf. 2000: 361). If we were to adopt this suggestion, we could say that the alleged defeater of R is the
conjunction, (1)&(N&E). This proposal at least has the merit of making the proposed defeater
something the adherent of N&E *comes to believe*. The adherent does not believe this conjunction
prior to coming to believe (1). It also accords well with such remarks as, "one who accepts [E&N]
(and also grasps the argument for a low or inscrutable value of [P(R/N&E)]) has a *defeater* for R"
(2002: 231), and the earlier, "But . . . why think that N&E (together with the proposition that
P(R/N&E) is low or inscrutable offers a defeater for R? (1995: 442, n. 8). (This latter paper (p. 440)
also includes statements apparently affirming each of the other two interpretations, as well, namely,
that the defeater is (1) and that the defeater is N&E). Nevertheless, I will stick with the
interpretation that (1) is the putative defeater, since whatever I have to say about it can, I think, be
translated into a comment about the conjunction of (1) and (N&E).

There is one other possibility, one that requires taking (1) as a premiss after all:
      (1)   P(R/N&E) is low or inscrutable.
      (2')   If (1) then for anyone who believes N&E and comes to believe (1), N&E is a
            defeater for R

So let us look at the argument (2)-(4). I am dubious about both of the premisses, (2) and (3).

First, I am not persuaded that

(2)   For anyone who believes N&E and who comes to believe (1), (1) is a defeater for R.

Suppose you do not believe that for just any pair of propositions, p and q, there is such a thing as the objective conditional probability, $P(p/q)$ of the one on the other,[19] or that if there is such a value, we do not in general have a way of telling what it is. Then you might regard many conditional probabilities as inscrutable. Realizing that it follows that if a probability is inscrutable, it is also low or inscrutable, you might infer of many such probabilities that they are low or inscrutable. You might entertain any proposition you believe, say,

(M) The moon is round

(3')   For anyone for whom N&E is a defeater for R, N&E is a defeater for any
        proposition he or she believes on the basis of his or her cognitive faculties.

∴   (4')   For anyone who believes N&E and who comes to believe (1), N&E is a defeater for
        N&E. (2), (3)

As I indicate in the text, this interpretation does not square with the official definition of a defeater as a newly acquired belief. But if this is supposed to be the argument, the objections I raise against (2) and (3) can be adapted with equal force against (2') and (3').

[19] When I first thought of this example, I assumed that it was purely hypothetical. Then I remembered that I had myself expressed such a skepticism about probability in (Wierenga, 1978). Compare a conjecture of Plantinga's: "Perhaps the probability relation (the relation in which A and B stand when there is a number n such that $P(A/B) = n$) isn't *connected* among propositions; perhaps it doesn't hold between A and B when A is contingent and B a tautology" (1979: 21), where, however, the clause following the semi-colon suggests that he did not mean to endorse a general skepticism about probability.

and then, given your general agnosticism about probabilities, conclude that P(R/M) is low or inscrutable. Surely that does not give you a defeater for R, and if parallel agnostic reasoning leads you to think that P(R/N&E) is low or inscrutable, you have not thereby acquired a defeater for R. It is simply hard to see why finding some such conditional probability low or inscrutable should give you a defeater for your belief in the reliability of our cognitive faculties

It might be thought that a flaw in this objection is that there is no reason to think that there is any interesting evidential or rational connection between M and R in the first place, so doubts about the conditional probability of R given M have no bearing on whether R is defeated . It is easy enough, however, to construct other examples that avoid this feature. Perhaps, impressed by the work of William Alston, you come to believe

(W) I have no non-circular evidence that my cognitive faculties are functioning properly.[20]

You then, by way of another judicious application of your agnosticism about probabilities, come to believe that P(R/W) is low or inscrutable. Does it follow that you thereby have a defeater for R? I think not. Failing to endorse an assignment to P(R/W) does not give one a reason to give up belief in R.[21] Similarly, there seems to be little reason to accept (2); failing to endorse a high enough assignment to P(R/N&E) also does not give one a reason to give up belief in R.

---

[20] See, for example, (Alston 1993), where Alston argues that there is no non-circular argument for the reliability of sense perception.

[21] This is perhaps Alston's point when he writes, "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&R and a belief in R?" (Alston 2002: 201).

But suppose that (2) were true. Suppose that the inscrutability of P(N&E) rationally requires one who accepts it (and N&E) to give up belief in R. Should we also think that such a person is now rationally required to give up belief in every proposition produced by his or her cognitive faculties? In other words, is (3) true? Here, too, I think the answer is that it is not. But to see this point it will be helpful to make a distinction and consider the question under two different scenarios.

Plantinga asserts that the Evolutionary Argument against Naturalism "does not presuppose this [that is, warrant as proper functioning] or any other specific epistemology" (2002*b*: 205, n. 2). This suggests that he thinks that this argument does not appeal to the unique features of his account of defeat. I think that (3) does not seem plausible, whether defeat is understood in the traditional way as based on evidentiary relations holding among propositions (Kvanvig's "front door" approach) or whether it is understood in Plantinga's way as deriving from what properly functioning cognitive faculties would require one to jettison (Kvanvig's "back door" approach). There are different ways of arguing for this claim, however, depending on the account of defeat one employs. Let us start with an intuitive version of the front door account, according to which a defeater for a proposition is new evidence which is either decisive evidence against that proposition or which sufficiently undermines the evidence one has for it. Then let us ask, if a person, *S*, has a defeater for R, does that person have a defeater for every belief produced by *S*'s cognitive faculties? If *S* believes N&E and comes to believe that P(R/N&E) is low or inscrutable, does *S* have a defeater not only for R but for every belief produced by *S*'s cognitive faculties? Perhaps *S* believes the proposition *S* would express by saying, "I have a hand" (call this proposition *h*). It seems clear that *S* does not have a *rebutting* defeater for *h*. Even if we concede that the belief that the probability of R on N&E is low or inscrutable is a defeater of R for *S*, it is not evidence *against h*. So if *S* has a defeater for *h*, it

can only be an undercutting defeater. Now presumably *S* has a good deal of evidence in favor of

h–the experiences of seeing her hand, of feeling it move in response to her efforts, of remembering

having used her hand, etc. These various pieces of evidence remain good evidence for *S*'s belief that

she has a hand, even if *S* does not believe R and even if she has a defeater for R. No doubt *S* need

not believe *h* on the basis of an argument that has R as one of its premises in order for her to be

justified in believing it, but, if I am right, she need not even believe *R* in order to be justified in

believing *h*. A defeater for the general thesis that her cognitive faculties are reliable can leave intact

her justification for specific propositions such as *h*. So, even if we concede that *S* has a defeater for

R, it is not obvious that she has a defeater for everything else she believes–at least if defeat is

characterized in a vague but intuitive way that appeals to relations of evidential support. On this

interpretation, then, (3) seems not to be true.

This is not, of course, how Plantinga construes defeat. As we have seen, on his view one

belief defeats another belief just in case proper functioning cognitive faculties toss out the second

belief upon acquiring the defeater. If a person whose cognitive faculties are functioning properly

stops believing R, as a result of acquiring a belief in (1), that P(R/N&E) is low or inscrutable, would

it also be the case that properly functioning cognitive faculties jettison all of that person's other

beliefs, as well? This seems incredible, but it is difficult to know what to say, especially if the defeat

system works, as I suggested above, as a sort of black box. Exactly which beliefs are ejected from a

person's noetic structure when the person comes to believe (1) and the person's faculties are

functioning properly? If the proper functioning proceeds by assessing relations of evidential

support, then the answer in this case should be the same as in the last: a person can be justified in

believing lots of ordinary propositions even if the person gives up the belief in R as a result of
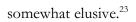
acquiring a defeater for R, and so proper functioning cognitive faculties need not expel all beliefs after acquiring a defeater for R.

Interestingly, Plantinga's account of defeat, coupled with his appeal to noetic structures, seems to offer an even stronger reason for this conclusion. Recall that for Plantinga a person's noetic structure includes not only that person's beliefs, but it also includes information about them, including the strength with which they are held and their "depth of ingression" (how central they are to the noetic structure). Typically a person's belief in *h* ("I have a hand") and other ordinary propositions is both exceptionally firm and central. For most of us, there are few beliefs more obvious than that we have hands, and it is one we could give up only with a wholesale revision of our beliefs. So what should properly functioning cognitive faculties do with a cognitive structure that includes belief in *h* but which, upon acquiring a belief in (1), abandons the belief that R? Will they further eject *all* of a person's beliefs, including those most firmly and centrally held? That seems like a poor way to function. Shorn of a belief in the reliability of our cognitive faculties themselves, one hopes that properly functioning faculties would attempt to hold on at least to the most obvious and fundamental beliefs a person has. My claim here is not merely that properly functioning faculties aimed at sustaining one's life and keeping one moderately happy will find a way to maintain many familiar beliefs in this situation; it is that properly functioning faculties *aimed at truth* will do this, too. For the typical evolutionary naturalist, belief in (1) or in N&E is significantly less firm and less central than belief in *h* or other fundamental beliefs. It would be odd for properly functioning cognitive faculties to ignore these differences within a noetic structure and to allow

27

more peripheral beliefs to be so destructive.[22] An easy solution would be for properly functioning cognitive faculties simply to ignore the effects of giving up R; they would not require belief in R in order to maintain other beliefs. More sophisticated cognitive faculties might isolate the rejection of R; they might reduce the strength of some beliefs in the periphery of the person's noetic structure, while allowing the firm and central ones to remain firm and central. However it is to be accomplished, it is hard to believe that cognitive faculties are functioning properly if they jettison all beliefs whatsoever. It is difficult to be sure about this, of course, because Plantinga has told us much less about how properly functioning cognitive faculty *adjust* noetic structures than he has about how they add beliefs to them. But it is tempting to think that on Plantinga's own account of defeat, it is even less plausible than on a traditional account to hold that acquiring a defeater for R yields a defeater for every belief. So, on his own account of defeat, (3) seems unpersuasive.

We initially turned to the Evolutionary Argument against Naturalism to see if we could come to a better understanding of Plantinga's account of defeat by seeing what that argument required. My discussion of the argument focused on interpreting it and, ultimately, criticizing it. But the original question remains: what must defeat be like in order for that argument to be convincing? The most we can say, I think, is that if (2) and (3) are true, then the proper functioning required for defeat must operate in some mysterious way to eject belief in R and all other beliefs. If the details of this operation are mysterious in this way, however, then Plantinga's concept of defeat remains

---

[22] My claim here is not available to the *dogmatic* evolutionary naturalist who holds N&E and (1) as firmly and as centrally as possible. Even someone who is this dogmatic about N&E, however, is less likely to be so committed to (1).

somewhat elusive.[23]

References

Alston, William (1993). *The Reliability of Sense Perception*. Ithaca, NY: Cornell University Press.

Alston, William (2002). "Plantinga, Naturalism, and Defeat" in Beilby (2002), pp. 176-203.

Baker, Deane-Peter (ed.) (2007). *Alvin Plantinga*. Cambridge: Cambridge University Press.

Beilby, James (2002). *Naturalism Defeated?: Essays on Plantinga's Evolutionary Argument against Naturalism*. Ithaca, NY: Cornell University Press.

Gettier, Edmund (1963). "Is Justified True Belief Knowledge?", *Analysis* 23: 121-123.

Kvanvig, Jonathan (ed.) (1996). *Warrant in Contemporary Epistemology*. Lanham, Maryland: Rowman and Littlefield.

Kvanvig, Jonathan (2007). "Two Approaches to Epistemic Defeat", in Baker (2007), pp. 107-124.

Merricks, Trenton (1995). "Warrant Entails Truth", *Philosophy and Phenomenological Research* 55: 841-855.

Plantinga, Alvin (1979). "The Probabilistic Argument from Evil", *Philosophical Studies* 35: 1-53.

Plantinga, Alvin (1983). "Reason and Belief in God", in Plantinga and Wolterstorff (1983), pp. 16-93.

Plantinga, Alvin (1993*a*). *Warrant: The Current Debate*. Oxford: Oxford University Press.

Plantinga, Alvin (1993*b*). *Warrant and Proper Function* Oxford: Oxford University Press.

Plantinga, Alvin (1994). "Naturalism Defeated", <http://philofreligion.homestead.com/files/alspaper.htm>, accessed April 28, 2010.

Plantinga, Alvin (1995). "Reliabilism, Analyses and Defeaters", *Philosophy and Phenomenological Research* 55(2): 427-464.

Plantinga, Alvin (1996). "*Respondeo*", in Kvanvig (1996), pp. 307-378.

Plantinga, Alvin (2000). *Warranted Christian Belief*. Oxford: Oxford University Press.

Plantinga, Alvin (2002*a*). "Introduction: The Evolutionary Argument against Naturalism, an Initial Statement of the Argument", in Beilby (2002), pp. 1-12.

Plantinga, Alvin (2002*b*). "Reply to Beilby's Cohorts", in Beilby (2002), pp. 204-275.

~~Plantinga, Alvin (2007). "Religion and Science", *The Stanford Encyclopedia of Philosophy* (Winter 2008 Edition), Edward N. Zalta, ed., URL = <http://plato.stanford.edu/archives/win2008/entries/religion_science/>.~~

Plantinga, Alvin and Wolterstorff, Nicholas (eds.) (1983). *Faith and Rationality: Reason and Belief in God.* Notre Dame, Indiana: University of Notre Dame Press.

Pollock, John (1974). *Knowledge and Justification.* Princeton: Princeton University Press.

Pollock, John (1986). *Contemporary Theories of Knowledge.* (Totowa, New Jersey: Rowman and Littlefield).

Russell, Bertrand (1967). "Letter to Frege", Beverly Woodward, trans., in Jean van Heijenoort, ed., *From Frege to Gödel.* (Cambridge, Mass.: Harvard University Press), pp. 124-125.

Ryan, Sharon (1996). "Does Warrant Entail Truth?", *Philosophy and Phenomenological Research* 56: 183-192.

Shope, Robert (1983). *The Analysis of Knowing: A Decade of Research.* Princeton: Princeton University Press.

Sosa, Ernest (2002). "Plantinga's Evolutionary Meditations", in Beilby (2002), pp. 91-102.

Sosa, Ernest (2007). "Natural Theology and Natural Atheology: Plantinga's Evolutionary Argument against Naturalism", in Baker (2007), pp. 93-106.

Sudduth, Michael (1999). "The Internalist Character and Evidentialist Implications of Plantingian Defeaters", *International Journal for Philosophy of Religion* 45: 167-187.

Sudduth, Michael (2009). *The Reformed Objection to Natural Theology.* (Farnham, Surrey, England: Ashgate Publishing Limited, and Burlington, Vermont: Ashgate Publishing Company.

Wierenga, Edward (1978). "Reply to Harold Moore's 'Evidence, Evil, and Religious Belief'", *International Journal for Philosophy of Religion* 9 (4): 246-251.