Can Teleosemantics Deflect the EAAN?

Brian Leahy

Abstract Alvin Plantinga's Evolutionary Argument Against Naturalism aims to show that the conjunction of contemporary evolutionary theory (E) with the claim that there is no God (N) cannot be rationally accepted. Where R is the claim that our cognitive faculties are reliable, the argument is:

P1.   The probability of R given N and E is low or inscrutable.
P2.   Anyone who sees (1) and accepts (N&E) has a defeater for R, and this defeater cannot be defeated or deflected.
P3.   Anyone who has an undefeated, undeflected defeater for R has an undefeated, undeflected defeater for everything she believes.
 C.   Therefore she has an undefeated, undeflected defeater for (N&E).

Plantinga (2011) defends the second premise. It examines and rejects several candidate defeater defeaters and defeater deflectors. One candidate is Millikan's teleosemantics. I show that Plantinga's motives for rejecting teleosemantics as a defeater deflector are inadequate. I then show that teleosemantics is not on its own an adequate defeater deflector. Then I offer an additional premise that constitutes a defeater deflector in conjunction with teleosemantics.

Keywords Evolutionary Argument Against Naturalism (EAAN) · Plantinga · Teleosemantics · Millikan · Productivity

B. Leahy (✉)
Zukunftskolleg and Departments of Linguistics and Philosophy, University of Konstanz, Konstanz, Germany
e-mail: brian.leahy@uni-konstanz.de

Introduction

Let metaphysical naturalism (N) be the claim that God does not exist, or if there is a God, this being plays no causal role in the world's transactions. The Evolutionary Argument Against Naturalism (EAAN) is concerned to show that N is rationally unacceptable. Let R stand for the claim that our cognitive faculties are reliable—that they produce a good deal more true beliefs than false beliefs in nearby possible worlds. Let E stand for the claim that our cognitive faculties are the product of evolution, as described by current evolutionary theory. Then the EAAN is:

P1. The probability of R given N and E is low or inscrutable.

P2. Anyone who sees (1) and accepts (N&E) has a defeater for R, and this defeater cannot be defeated or deflected.

P3. Anyone who has an undefeated, undeflected defeater for R has an undefeated, undeflected defeater for everything she believes.

C. Therefore she has an undefeated, undeflected defeater for (N&E).

Alvin Plantinga's article "Content and Natural Selection" (Plantinga 2011) defends P2. It examines and rejects several candidates for defeater defeaters or defeater deflectors. One of the candidates is offered by Millikan's teleosemantics. I will show that Plantinga has overlooked relevant aspects of Millikan's theory, and so has not shown that Millikan's theory cannot deflect the defeater for R. I will then ask whether Millikan's theory can deflect the defeater for R. I will show that on its own it can not. But supplemented with a further premise it can.

   The next section briefly describes the EAAN. The third section presents Plantinga's two arguments for why teleosemantics does not provide a reason to reject P2. The following section outlines Millikan's view. The fifth section shows how Millikan's account avoids Plantinga's arguments. The sixth section argues that teleosemantics cannot, on its own, defeat or deflect the defeater for R. It then introduces and additional premise that, combined with teleosemantics, can deflect the defeater for R, demonstrating that P2 is false. The final section concludes.

The Evolutionary Argument Against Naturalism

The EAAN is developed at length in Plantinga (1993, 2002). In Plantinga (2011), he establishes P1 thus: start with a hypothetical species such that

(N&E) is true for that species. What would p(R|N&E) be for this species? Assume materialism or physicalism, and assume that a belief is some sort of neurological structure with both neurophysiological properties and propositional content. Thus understood, it is plausible that beliefs play a causal role in generating behaviour. But it is not clear that their propositional content plays a causal role: for if a belief had the same neurophysiological properties and a different content, it would still play the same causal role in generating behaviour. Plantinga concludes that semantic epiphenomenalism (SE) is likely, given (N&E): p(SE|N&E) is high.

If SE is true, then false beliefs do not cause maladaptive behaviour because they are false, and true beliefs do not cause adaptive behaviour because they are true. So the truth or falsehood of a belief is causally irrelevant to behaviour, and so causally irrelevant to fitness. So it is unclear how natural selection can select for true belief or against false belief. Writes Plantinga, "If SE were true, it would be an enormous cosmic coincidence, a stunning piece of not-to-be-expected serendipity, if modification of behaviour in the direction of fitness also modified belief-production in the direction of greater reliability" (p. 437). So R is unlikely given N&E&SE: p(R|N&E&SE) is low.

But then, by the theorem of total probability, we have:

$$p(R|N\&E) = [p(R|N\&E\&SE) \times p(SE|N\&E)]$$
$$+ [p(R|N\&E\&{\sim}SE) \times p({\sim}SE|N\&E)]$$

This equation ensures that when p(SE|N&E) is high and p(R|N&E&SE) is low, p(R|N&E) must also be low. This can be checked by substituting arbitrary values for the terms of the equation: for example, if p(SE|N&E) is 0.8 and p(R|N&E&SE) is 0.3, p(~SE|N&E) is 0.2 and p(R|N&E) is at most 0.44—that is, if p(R|N&E&~SE) is 1. While it is unreasonable to assign precise values to these terms, our estimation of the value of p(R|N&E) should be guided by this equation.

Plantinga concludes that p(R|N&E) is low for our hypothetical population under the assumptions of materialism/physicalism and that a belief is some sort of neurological structure with both neurophysiological properties and propositional content such that p(SE|N&E) is high. And since the human population is not relevantly different from our hypothetical population, p(R|N&E) is low for us as well.

Now we turn to P2, the claim that anyone who accepts both P1 and (N&E) has a undefeated, undeflected defeater for R. Plantinga argues that there is no admissible additional proposition X that the naturalist may accept such that p(R|N&E&X) is high. His argument cannot be exhaustive; rather, he reviews and rejects several reasonable candidates. One of these, labelled T, is the proposition that expresses a Millikan-style teleosemantic theory of content.

In this paper I will assume that T is a candidate defeater deflector, not a defeater defeater. Any proposed defeater deflector must be both *admissible*

and *adequate*. Plantinga provides some necessary conditions on admissibility. First, no proposition that entails R or entails R in conjunction with N or E is admissible, since if such defeater deflectors were admitted, every defeater would be deflected (R itself would be a defeater deflector for every defeater of R). Second, no proposition that is equivalent to R is admissible (this is a consequence of the first condition). Third, no contradiction or obviously irrational belief (Plantinga's example is the denial of a simple truth of arithmetic) may be a defeater deflector. What the EAAN aims to show is that there is an irrationality to believing [(N&E) and p(R|N&E) is low]; even if an irrational belief X redeems that pair of beliefs from irrationality it does so at the expense of introducing irrationality elsewhere. A proposed defeater deflector X is adequate only if p(R|N&E&X) is neither low nor inscrutable (p. 440).

Plantinga, in denying that teleosemantics provides a defeater deflector, must deny either that T is admissible or deny that it is adequate. He denies its admissibility.

Plantinga's Arguments

Plantinga argues that if T were true, there would be no such thing as belief in N. As a result, conjoining T with N to evaluate the conditional probability of R on (N&E&T) introduces irrationality into the belief system. So T is inadmissible.

The argument for why N and T cannot be conjoined is this. Suppose that T is true and that N is a teleosemantic signal. Then there is (1) a producer of tokens of N, (2) a consumer of tokens of N, (3) a response A from the consumer to tokens of N such that (4) there is a state of affairs S that must obtain if the consumer(s) of N are to accomplish their functions by the normal mechanism in responding A-wise. Then the content of a token of N is S.

Plantinga allows that (1) and (2) may be fulfilled. The problems he sees involve (3) and (4). What is the adaptive response from consumers of N? And can the state of affairs that God does not exist be necessary for consumers to perform their proper functions normally by responding that way?

Plantinga argues that there is no adaptive response by consumers of N to tokens of N. Of course, tokens of N evoke a wide range of actual responses. But none of these do much if anything to increase the reproductive chances of their consumers; and if they do occasionally increase the reproductive chances of their consumers, they would do so whether or not God exists. Furthermore, even if there were a normal response, that God does not exist could not be a necessary condition on the consumer performing their proper functions normally by responding that way. For whether God exists or not, God does so invariantly, and so no sign can covary with variations in that state of affairs, and so no sign can carry information about that state of affairs. Plantinga concludes that if teleosemantics is true, then there can't be any such belief as naturalism.

I now want to show how Professor Millikan's view is not addressed by this challenge.

A Presentation of Millikan's Teleosemantics

I think Professor Plantinga's arguments support this conclusion: a signalling system where the only type of signal was 'God does not exist,' where the only variable aspects of that sign were the time and place of its utterance, and where the variable system of environmental conditions responsible (in Millikan's sense) for the success of the signalling system was God's nonexistence, could not develop. For God's nonexistence is not a variable state of affairs, and so signal producers cannot have as function to make their sign co-vary with variations in God's existence.

However, a signal that means that God does not exist can be part of a richer signalling system, that is, a system whose token signals are articulated into more variable aspects. Further, facts about God need not have played any explanatory role in the historical success of the system. To understand this we need a presentation of Millikan's view on productivity: that is, on how novel signals can have content determined by the selectional history of the signalling system of which they are a member. On this view a sign can have as content a state of affairs that has played no causal role in the selection of the signalling system.

Here I present sufficient conditions on the development of a productive Millikan-style signalling system. These are not all individually necessary. I add unnecessary restrictions that reduce the generality of the account presented because I think they ease presentation of the view. The account is fully developed in Millikan (1984a). I begin with basic pieces of Millikan's technical apparatus, gradually developing and assembling Millikan's complex account of the content of a representation. With one example I introduce and illustrate (reproductively established) characteristics, proper functions, and most proximate normal explanations. I use a different example to illustrate variable ranges of environmental conditions. A third example assembles introduces and illustrates the notions of signal producer, signal consumer, signalling system, and mapping function. Finally I show how these pieces determine the content of a signal.

Take a family I of devices that are reproducing under some sort of selection pressure, such as natural selection. Suppose that some members of I have a characteristic C, and having C sometimes results in them having effect F. Suppose that having effect F is beneficial for the organisms that have it, increasing their reproductive chances, and thereby contributing to the proliferation of C in I. When one must appeal to the fact that some members of I with C had effect F in order to explain the proliferation of C in I, we say that it is a proper function of descendent members of I to F; and that holds even for descendent members of I that lack C or are otherwise unable to F.

For example, let the lenses of human eyes be the family I; some members of this family have a characteristic C (a particular shape) that sometimes causes diffused light reflected from a single point in the environment to be projected onto a single point on the retina. Doing so was beneficial for human eyes with lenses, and increased their reproductive chances, making sure lenses with that shape remain widespread amongst humans. Thus we say that causing diffused light reflected from a single point to be projected onto a single point on the retina is a function of those lenses, including malformed lenses that are incapable of doing so.

A complete, general explanation of how members of I most often performed F on the occasions when they actually performed F is called a normal explanation for the performance of F. A normal explanation will appeal to characteristics shared by all members of I that did F on the relevant occasions, environmental conditions those members were in that were consistent across the relevant occasions, and, by adding natural laws, shows without gaps how this setup resulted in the performance of F. An explanation is *general* if it covers several cases by appeal to a generality, rather than a Boolean construction. Suppose members of I in fact sometimes performed F through process x, sometimes through process y, and sometimes through process z. Even if members of I in fact performed F through process (x or y or z) more often than they performed F by any other process or Boolean construction of processes, it is not the case that it is normal for members of I to perform F by process (x or y or z). Disjunctive explanations of this sort are not normal explanations. A normal explanation for how members of I perform F looks for the single process by which members of I most often performed F on the occasions when they in fact performed F. In our example, a normal explanation for a lens' ability to cause diffused light reflected from a single point to be projected onto a single point on the retina will appeal to internal features such as the shape and transparency of the lens, and so on; environmental conditions such as the transparency of the medium (air) that most often lies between the eye and observed objects, spectrum and intensity of ambient light, etc.; natural laws about the behaviour of light in those environmental conditions such as rates of diffusion over various distances, etc., all in enough detail to show completely how this arrangement leads to diffused light reflected from a single point in the environment being projected onto a single point on the retina.

When a member i of I has effect F by the mechanism appealed to in the normal explanation, we say that i has done F normally. When a member i of I has the internal characteristics C appealed to in the normal explanation, then we will say that i is normally constructed. Alternatively we may say that a member of I is normally constructed not across the board but relative to some specified functions. When environmental conditions are as they need to be for a member i of I to perform all F's normally, then we will say that i is in its normal conditions. Again, we may also say that a member of I is in its normal conditions not across the board but only relative to some specified functions. The least detailed normal explanation that is still a complete explanation is called the most proximate normal explanation.

Now suppose that the causal process specified by a normal explanation for a function F is in some ways costly for individuals engaged in that process; it requires resources that could be used elsewhere. And suppose that the process by which F is normally accomplished does so with higher probability under certain environmental conditions, but with lower probability under other, incompatible environmental conditions. That is, there is a natural variable range of environmental conditions E such that the conditional probability of completing the process by which F is normally accomplished is different on different members of that range. In that case I will do better if that causal process is initiated by its members more when environmental conditions determine a higher probability for successful F, and less when environmental conditions determine a lower probability for successful F.

Suppose further that members of I are not causally sensitive to the relevant environmental conditions. So whether the process is initiated or not is independent of whether environmental factors determine a higher or a lower probability for successful F. The only causal impact those environmental factors have on the process that results in F occurs after the costs associated with that process have been incurred.

A new (fictionalized) example. Some turtles bury their eggs in holes dug on beaches (a costly process). These holes protect the eggs until they hatch by hiding them from predators. But the effectiveness of this depends on whether the beach is home to a species of lizard that is capable of detecting buried turtle eggs. So the conditional probability of eggs being protected by burying given that there are egg-detecting lizards on the beach is lower than the conditional probability of eggs being protected by burying given that there are not egg-detecting lizards on the beach. Turtles as a family will do better if more turtles bury their eggs on beaches without these lizards instead of on beaches with them. But if turtles are not sensitive to whether a candidate beach has egg-detecting lizards, they will be unable to take advantage of this fact. Turtles would be more successful if they were consumers in a signalling system that indicates which beaches have egg-detecting lizards. But they are not.

Now suppose that there is a family of devices P (*producers*), again reproducing under natural selection, that also benefits when members of I (now *interpreters*, or alternatively *consumers*) have effect F. Suppose further that members of P are causally impacted by variations in E, the range of environmental conditions that impacts the probability of success of the process by which members of I normally have effect F. Suppose further that members of P have causal impacts on members of I through the modulation of some medium M, impacts that sometimes inhibit and/or sometimes stimulate the causal process by which members of I normally have effect F. Suppose that, as a result of natural selection, these causal processes become co-ordinated: the varying causal impacts that variations in E have on members of P result in variation in how members of P modulate M, thus modulating the causal impact members of P have on members of I; and in particular, to inhibit the causal process by which members of I normally perform F when the state of E

determines a lower probability of success and/or stimulate that causal process when the state of E determines a higher probability of success.

Another example (much simplified). Human ovaries (I) have as function to bring an egg to maturity at a certain point in a woman's menstrual cycle. Ovaries that bring an egg to maturity at the correct point lend to the success of all other organs in a woman that has such ovaries. Pituitary glands (P) are one such organ. Pituitary glands, unlike ovaries, are immediately sensitive to variations in the level of GnRH—gonadotropin releasing hormone—produced by the hypothalamus (E).[1] When the concentration of GnRH is sufficiently high, the pituitary gland secretes LH—lutenizing hormone—which (through several intermediates) stimulates the ovaries to bring an egg to maturity. This process has been carefully calibrated by natural selection. The varying causal impacts that variations in levels of GnRH have on pituitary glands result in the pituitary gland introducing different volumes of LH in the bloodstream, thus modulating the causal impact pituitary glands have on ovaries; and in particular, to stimulate the causal process by which ovaries bring an egg to maturity when the volume of GnRH secreted by the hypothalamus determines a higher probability of success.

How could this natural selection occur? For teleosemantic content to arise, it must be a result of members of I (ovaries), in being impacted in variable ways by members of P (pituitary glands) through the modulation of M (LH), doing better at performing F (bringing eggs to maturity at the correct point). That is, the causal impacts members of P have on members of I through modulation of M must improve the ability of members of I to F in a range of environmental conditions (across variations in the volume of GnRH). And the benefits to P must depend on the success of members of I in performing F. That way we can have the selection of a *signalling system*: the variation of some medium (M) by the producers of a sign (members of P) that has variable impact on consumers of the sign (members of I) such that how the producer varies the medium improves the success of I, which in turn improves the success of P, which in turn keeps that system of variations in M being reproduced and *stable*. That system of variations will continue to be reproduced because members of P that (unilaterally) respond to variations in the relevant environmental conditions by modulating M in some manner other than according to the historically established pattern will be selected against, as will members of I who unilaterally cease to respond to the form of M in accord with the historically established pattern. And that system of modulations of M—that system of signals—is in fact itself being selected.

In a signalling system, the most proximate normal explanation for how members of I perform the functions that keep the system of signals being stably reproduced will appeal to the historical pattern of relationships between modulations of M and variations in E. Remember that a normal explanation

---

[1] By *immediate* I mean that GnRH acts directly on cells in the pituitary. All biologically useful causal impact of GnRH on ovaries is mediated by the impact of GnRH on the pituitary.

for how members of I perform F is a general explanation for how members of I most often in fact performed F on actual past occasions of success. In order to do so, the normal explanation will need to appeal to a generalization over actual cases of past success that describes how variations in the medium co-varied with variations in some variable range of environmental conditions and accordingly adjusted the behaviour of members of I, contributing causally to their successful performance of F. This generalization can be described by a mathematical function that pairs (in the first instance) sets of actual values that the modulable elements of M took on the normal occasions with actual values that the relevant variable range of environmental conditions took on those same occasions. These mathematical functions are called *mapping functions*.

The content of a signal is one of the environmental conditions that must obtain if the interpreter of the signal is to perform its functions in accord with a normal explanation. But there are many such conditions, and not all are the content of the signal. The privileged environmental condition is the one that is a member of E, of the variable range of conditions such that it is a function of producers in the signalling system to produce signals that covary with variations in that range. It is the state of affairs that that modulation of the medium is mapped to by the mapping function for the system.

Mapping functions may be disjunctive and/or irregular. Speaking of mapping functions Millikan writes, "Isomorphisms can be defined by functions that are as bizarre, as gruelike, as you please" (Millikan 2004, p. 84). In the next paragraphs I describe first how signalling systems whose mapping functions are projectable display productivity. Then I will describe the more limited productivity signalling systems display when their mapping functions are gruelike.

Suppose that the generalization that must be appealed to in giving a normal explanation for how consumers perform those of their functions that keep the signalling system being stably reproduced is underwritten by a natural regularity, such that variations in the signal map variations in environmental conditions in accord with this natural regularity. Natural regularities project. Consequently, the mapping function will pair not only actual historical modulations of M with environmental conditions that must obtain if interpreters are to perform their proper functions normally, but also modulations that have not been produced before. For example, a honey bee, returning from a good source of nectar, performs a dance on the honey comb that should correspond to the location of nectar found. In addition to the time and place of the sign, the dancing bee can modulate the axis of the dance relative to the vertical of the honey comb, the speed of the dance, and the intensity of the dance. These latter elements correspond to variations in the direction of resources, the distance of the resources, and the abundance of the resources, respectively. To focus on one variable element of the dance: as orientation off the vertical axis of the honeycomb changes, the direction where nectar must be if interpreters are to perform all relevant functions in responding to the dance changes accordingly. This portion of the mapping function is specifiable by a projectable rule: if the orientation of the dance is x° off the

vertical axis of the honeycomb, there should be nectar x° off the line from the hive entrance to the sun. It may be the case that not every angle off the vertical has a historical instance; for example, perhaps no dance 42° off the vertical has ever been danced. Still there is a proper response to a dance 42° off the vertical, because the normal explanation for how bees should respond to dances appeals to a projectable rule that determines, for any orientation of the dance, a corresponding direction of nectar. Thus projectable rules ensure that novel signals can have mapping conditions, that there can be signals that map states of affairs that have historically played no causal role in explaining the success of the system.

Irregular, nonprojectable mapping functions can also enable the interpretability of novel signs, though in a more limited way. The speed of the bee's dance correlates with the distance of resources found. This is projectable; faster dances correlate with nectar at shorter distance. But let's pretend the rule that the normal explanation appeals to is irregular: if speed is 1 cm/s, nectar should be 10 m away; if speed is 3 cm/s, nectar should be 5 m away; if speed is 3.5 cm/s, nectar should be 15 m away; and so on for 15 or 20 different speeds. In order for a normal explanation to appeal to such an irregular rule, there must be an actual historical instance of a dance for each speed of dance in the domain of the mapping function. And the mapping function must pair speeds that dances actually took on normal occasions with distances where nectar was on those same occasions, where the fact that there was nectar at those distances contributes to explaining why bees that displayed exactly that pattern of responses to dance-speeds were successful.

In this example, dances with novel speeds are not interpretable by normal mechanisms, though dances with novel orientations are. Still, novel combinations of non-novel speeds with orientations and intensities are novel, interpretable dances. For example, perhaps there has never been a bee dance with orientation 42°, speed 3 cm/s, and intensity x, even if there has been a historically normal dance with orientation 42°, another with speed 3 cm/s, and yet another with intensity x. The novel dance will be interpretable, even though the mapping function that relates speeds to distances is nonprojectable. Most of the correspondences found in natural languages are of this more limited sort (Shea 2012; Millikan 2012).

I conclude this section with two examples that may clarify the purpose of Millikan's apparatus. A central virtue of teleosemantics is its ability to account for false representation as well as representations that are true by accident. Beavers signal predators to other beavers by slapping their tails sharply on water. Suppose a beaver is scared by a falling branch and slaps its tail on the water when there is no predator nearby. The mapping function for the tail slap requires that the time and place of the tail slap co-varies with the time and place of a predator. A signal is false if and only if it is produced and the state of affairs that it maps does not obtain; so the tail slap in our example is false.

On the other hand, no most proximate normal explanation for how consumers perform a function that keeps the system of tail-slaps stable needs to appeal to the antecedent causes of covariation between tail-slaps and

predators, though the most proximate normal explanations for some of those functions will appeal to the fact of covariation. So if a beaver is scared by a falling branch and slaps its tail on the water when there is an undetected predator nearby, then consumers that respond to the slap by diving under water perform the function of predator avoidance in accord with a fully normal most proximate normal explanation. The sign is true, but true by accident. The difference between an accidentally true sign and a nonaccidentally true sign lies on the production side: the most proximate normal explanation for some functions of producers that keep the system of tail-slaps stable requires that perceptions of predators cause producers to slap. In this example, the producer's functions in signal producing are performed, though abnormally. The relevant consumer functions are performed normally.

Response to Plantinga's Arguments

Millikan's account of productivity lets us address Plantinga's concerns. Again, I agree that a signalling system whose only type of signal was 'God does not exist,' whose only variable aspects were the time and place of utterance, and where the variable system of environmental conditions that explains the success of the signalling system was God's nonexistence, could not develop. For there is no behaviour on the part of interpreters that will be more successful depending on which environmental condition (God's nonexistence) from the range of possible variants is actual. There is no range of possible variants. God, existent or not, does so invariantly. However, richer signalling systems that include 'God does not exist' as a token can develop, if we have a broader range of variable aspects of the sign, and a corresponding broader range of variable aspects of the signifieds.

There are broad differences between simple signalling systems like tail-slaps and human languages. Some of the differences are spelled out in Millikan (2005). We don't need to go through them here; it will suffice to see some of the respects in which 'God does not exist' is part of a system that is richer than the beaver's system. In addition to the time and place of its production, producers in this signalling system can modulate the subject of the sign, replacing 'God' with 'Santa Claus' or 'Saint Christopher' or 'phlogiston' or 'George Bush'. And they can modulate whether the sign involves negation or not: they can say 'Phlogiston exists' or 'Saint Christopher doesn't exist'. The signalling system may include a great deal more variability still. We can neglect that additional variability for now. What matters is that a signalling system with just this degree of variability could arise, and could include 'N', a sign that is true just in case God does not exist. There being an adaptive response to N is not a necessary condition on the existence of such a signalling system, as long as there are adaptive responses to sufficiently many other members of the signalling system to result in the selection of a rule-bound (though the rule may be nonprojectable) relation between variations in the form of the sign and variations in the form of the environmental condition that must obtain if the

consumer of the sign is to perform its proper functions normally, and as long as N is a modulation in the domain of the mapping function for the signalling system.

Can the state of affairs that God does not exist be necessary for consumers of the sign N to perform their proper functions by the normal mechanism? Millikan argues at length that it is. She claims that the function of nonexistence claims is to cause the deactivation of a concept, to get hearers to stop wasting their time trying to use that concept to identify things. Similarly the function of existence claims is to cause the activation of a concept, as that concept will help hearers identify things. Details of this account are found in Millikan (2000) and Millikan (1984a, chapter 12).

Moreover, that God doesn't exist, though a necessary condition on consumers of 'God doesn't exist' performing all proper functions in accord with a normal explanation, need not provide any benefits for consumers of 'God doesn't exist'. As long as following the rules established by the selection of the entire signalling system pays off for the consumer and thereby for the producer often enough, there can be a sign that means that God doesn't exist, and it might be true, even if consuming tokens of 'God doesn't exist' never provides any useful benefits to their consumers.

And finally, the invariance of God's existence or inexistence causes no problems for the teleosemanticist, if the signalling system that contains 'N' is sufficiently articulate. For it is various tokens in a signalling system that must covary with variations in a variable range of environmental conditions, not tokens of the same signal 'N' that must covary with variations in the enironmental condition that 'N' maps. Millikan argues that existence and nonexistence claims are signs of relationships between words and the world or between words and other words (Millikan 1984a, p. 203). For example, the environmental condition that must obtain if consumers of 'George Bush exists' are to perform their proper functions normally is that the name 'George Bush' has a referent. 'George Bush exists' is subject to transformations: we can replace the subject with 'Don Quixote' or with 'Calamity Jane' or with 'God'; we can negate the sentence, and much more. Each of these transformations corresponds to transformations of the environmental condition that must obtain if consumers of the transformed sentence are to perform their functions normally. Each environmental condition that results from such a transformation is a member of a variable range of environmental conditions. It is this range that must be variable. The invariance of God's existence or nonexistence is not relevant.

So, on Millikan's theory, if a signalling system is sufficiently articulate, it may contain a signal whose truth condition is that God does not exist. So contra Plantinga's argument, if T (teleosemantics) were true, there could be a belief that that God doesn't exist. So the conjunction (T&N) can be rationally believed, and so T is an admissible defeater-deflector.

It has been objected that 'God does not exist' could be a necessarily true or false claim about an entity that exists outside our natural environment, as are mathematical claims such as "There is no highest prime number".

Whether teleosemantics can explain the content of mathematical statements is an interesting and open question too intricate to enter into here. But if teleosemantics cannot explain the content of mathematical statements, then it could also have trouble with 'God does not exist' when this is also understood as a claim about an entity that exists outside of our natural environment.

Millikan discusses similar issues in Millikan (1995, 2006). However, even if her responses there are inadequate, this issue raises no problem for the strategy of this paper. For Plantinga characterizes naturalism not as the proposition that God does not exist, but as the logically weaker proposition that either God does not exist, "...or if there is [a God], this being plays no causal role in the world's transactions" (Plantinga 2011, p. 435). So to argue that teleosemantics can't capture the content of 'God does not exist' because God exists only outside the natural environment concedes naturalism as Plantinga defines it. On the other hand, if God does play a causal role in the world's transactions, then this being is not the sort of being that raises challenges for the teleosemanticist. So my position may be qualified: if teleosemantics is true, either there can be a belief whose content is that God does not exist, or else God plays no causal role in the world's transactions.

Is T An Adequate Defeater Deflector?

In this section we will see that p(SE|N&E&T) is low. As a result, Plantinga's argument (presented in the second section of this paper) cannot be extended to show that p(R|N&E&T) is low. Then we will see that we cannot establish that p(R|N&E&T) is high. I then offer a premise S, acceptable to the teleosemanticist, such that p(R|N&E&T&S) is high.

If T is true, a belief is a member of a signalling system that mediates between two families of reproducing devices. The members of the family of belief interpreters are variously impacted by beliefs to generate action (in conjunction with other beliefs, desires, and so on); members of the family of belief producers modulate beliefs so that those actions are more effective given variations in some variable range of environmental conditions. The mechanisms that produced a given belief caused it to have the NP properties it has because those mechanisms have a selectional history of causing beliefs with NP properties that cause interpreters to (interact with other mental states and thereby) act more effectively, as a result of the historical covariation between variations in the form of the belief and variations in some variable system of environmental conditions. The content of the belief is a consequence of this selectional history. (Again, the content of a given signal is one of the environmental conditions that must obtain if the consumer of the sign is to perform its functions in the normal way, and in particular, the environmental condition that is an element in the variable system of environmental conditions such that it is a function of the producer to make the signal covary with variations in that variable system of environmental conditions.) As a result, Plantinga's premise that "a given belief...would have had the same causal

impact on behavior if it had had the same NP properties, but different content" (p. 436) is false. The content property of a belief is determined by the historical properties of the signalling system. So if a belief had had the same NP properties but a different content, the signalling system of which the belief is a member would have to have had a different history.[2] But then, if the signalling system had had a different history, it's very unlikely that signals in the signalling system would have had the same effects: the devices that are effected by the signalling system would have evolved to respond to signals in different ways. To use the jargon of possible worlds, the closest worlds where a belief has the same neurophysiological properties and different content is a world where the devices that respond to beliefs are different. So if a belief had had the same NP properties but a different content, it would have had a different causal impact on behaviour.

So p(SE|N&E&T) is low. So there is no argument that p(R|N&E&T) is low. But can the teleosemanticist ensure that p(R|N&E&T) is high?

First, it is clear that no low value for p(SE|N&E&T) in the equation below is on its own sufficient to determine a high value for p(R|N&E&T). In the equation:

$$p(R|N\&E\&T) = [p(R|N\&E\&T\&SE) \times p(SE|N\&E\&T)]$$
$$+ [p(R|N\&E\&T\&{\sim}SE) \times p({\sim}SE|N\&E\&T)]$$

if p(SE|N&E&T) is close to 0, p(R|N&E&T) is low if p(R|N&E&T&~SE) is. If p(SE|N&E&T)=0, the equation reduces to:

$$p(R|N\&E\&T) = p(R|N\&E\&T\&{\sim}SE)$$

Nothing yet said ensures that p(R|N&E&T&~SE) is high.

Moreover, the truth of (N&E&T) is consistent with the falsity of R. If teleosemantics is true we know that, historically, enough beliefs were true relative to the strength of the consequences of error that continuing to respond to our beliefs in the normal way is a better strategy than unilaterally changing to any alternative. Consumers of signs in a teleosemantic system cannot improve their expected payoffs by unilaterally changing their pattern of responses in the signalling system. This can result in signalling systems such that consumers respond to signals in a way that maximizes expected value even though signals are unreliable. For example, beavers have quite unreliable predator-detecting mechanisms. These mechanisms most often cause representations of predators nearby—and induce diving under water—when there are no predators nearby. But the mechanisms persist because the costs of responding to false representations of predators nearby by diving under water are sufficiently low, given the proportion of false representations, and the benefits of responding to true representations of predators nearby by diving under water are sufficiently high,

---

[2]We are here forced to reckon with *backtracking* counterfactuals, as described by Lewis (1979).

given the proportion of true representations. Responding to a given signal with predator avoiding behaviour is a better strategy—has higher expected value—than any other available strategy, even though the probability that the signal is false is greater than the probability that the signal is true.

This establishes that p(R|N&E&T&X) is low, where X is the proposition "responding to false representations as though they were true has low costs and responding to true representations as though they were true has high benefits". Of course it is still possible for reliable signalling systems to arise when payoffs are arranged as described by X. But such signalling systems tend to be unstable, for if producers manage to generate a few extra true representations at the cost of not too many extra false representations, the value of the whole system increases. As the difference between the cost of responding to false representations as though they were true and the benefit of responding to true representations as though they were true broadens, the number of extra false representations that are worth generating if they bring along even one true representation increases.

If beavers were sufficiently reflective, or if humans had belief-generating mechanisms that shared this feature of unreliability, then they (we) would have to admit, for any belief generated by the system, that it was unlikely to be true. So on Plantinga's view that belief is defeated.

Millikan's position on knowledge is spelled out in detail in Millikan (1984b). She argues that a belief counts as knowledge when it is produced by a belief-producing mechanism that is performing its proper functions in its normal conditions. On this account some of the outputs of unreliable mechanisms can still count as knowledge. But I won't engage in that debate here. Rather, I will offer an admissible proposition S that the teleosemanticist may accept such that p(R|N&E&T&S) is high. (T&S) is an admissible, adequate defeater deflector. So not only is Professor Plantinga's argument for P2 flawed; P2 is false as well.

Before proceeding I want to set aside two sceptical challenges. I will argue that (N&E&T&S) provides a reason to accept R. The classical sceptic may challenge that this is subtly circular: if the goal is to show that our cognitive faculties are reliable, we cannot accomplish that goal by offering an argument. For any such argument, one way or another, will need to assume that our cognitive faculties are reliable.

I am not concerned here to address the classical sceptical challenge. As Plantinga notes, the classical sceptical worry traps anyone who needs an argument for the reliability of her cognitive mechanisms (Plantinga 1993, p. 236–237). He points out that the non-naturalist "cannot sensibly *argue* that in fact our beliefs are mostly true" (p. 236). But the non-naturalist is not compelled to doubt the reliability of her cognitive mechanisms anyway; and classical sceptical arguments that we should doubt the reliability of our cognitive mechanisms do not compel the naturalist to doubt the reliability of her cognitive mechanisms any more than they compel the non-naturalist to do so. With respect to the classical challenge, naturalist and non-naturalist are in the same boat.

But there is a subtler sceptical challenge. The first premise of the EAAN provides the naturalist, but not the non-naturalist, with a special reason for doubting the reliability of her cognitive faculties. Plantinga's sceptical challenge arises only once it has been established that $p(R|N\&E)$ is low, and only arises for those who accept N and E. Thus the non-naturalist is exempt from Plantinga's sceptical challenge.

If the naturalist needn't accept that $p(R|N\&E)$ is low, then she is exempt from Plantinga's sceptical challenge. In arguing that $p(R|N\&E)$ is low, Plantinga appeals to the premise that $p(SE|N\&E)$ is high. He establishes this by arguing that "a given belief...would have had the same causal impact on behavior if it had had the same NP properties, but different content" (Plantinga 2011, p. 436). Since the teleosemanticist rejects this conditional, she needn't accept P1. So the naturalist has been given no reason unshared with the non-naturalist to suspect that her cognitive faculties are unreliable.

I now offer an argument that our cognitive faculties are reliable. These arguments are not addressed to the classical sceptic, who will be unimpressed. They are addressed to she who, having noticed that teleosemantics admits of unreliable signalling systems, wonders whether teleosemantics only admits of unreliable signalling systems, or whether there are conditions under which teleosemantic signalling systems are usually reliable. Then we may check if these conditions hold for human cognitive faculties.

Teleosemantics guarantees that some proportion of tail-slaps are true. So there is some integer k such that for any set of k-many tail-slaps, the probability that one member of that set is true is high (the probability of drawing the ace of spades on one draw from a fair deck is low, but the probability of drawing the ace of spades at least once out of 70 draws from 70 fair decks is high). Moreover, any single tail-slap is unreliable, but if k-many beavers independently perform tail-slaps at the same time and place, the probability of error is lower. For it is unlikely that all k beavers are moved by different events to perform tail-slaps at the same time and place; it is more likely that their slaps are all or mostly all motivated by perceptions of the same state of affairs. But if they are all responding to the same state of affairs, as observed from different perspectives, then the probability of error is lower. Many kinds of misperceptions are ruled out. Not all of them, surely; perhaps k beavers may all be scared by the same large branch falling to the ground. But a falling twig, barely sufficient to scare one beaver by falling close to its head, is insufficient to scare several beavers at diverse locations.

The idea here, central to Millikan's theory, is that coherence is a test of truth (Millikan 1984a, chapter 18, 2000, chapter 7). "Highly consistent convergence of independent methods to the same judgements serves as strong testimony to the objective univocal sources of those judgements" (Millikan 2000, p. 103). Agreement in judgements about the same enduring state of affairs at different times, from different perspectives, through different media, using different tools or perceptual apparatus, correlates with the truth of that judgement. As Millikan writes, "...the primary evidence for our hold on the objective world is

that we do not find ourselves wanting to affirm and negate the same sentence" (Millikan 1984a, p. 308).

The vast majority of our perceptual beliefs in fact are checked many times over, and since we almost always find agreement amongst our judgements, we have good reason to think that most of those judgements are accurate. Almost never do I find that my chair, or my desk, or my shirt changes its colour, shape, or texture from moment to moment or encounter to encounter. Given that some of my judgements are true, and that almost all of them are coherent, we have a good reason to think that most of them are true. "There are, in most cases, many more ways for a sentence not containing the negative to be false than for it to be true. And when a whole fabric of agreeing positive judgement appears, covering a domain such as that of the colors and shapes of mundane physical objects, the evidence that we are mapping the same domain of affairs out there with our independent judgements, hence that each of us is mapping *a* domain of affairs out there with his or her judgements, is overwhelming" (Millikan 1984a, p. 309). So $p(R|N\&E\&T\&S)$ is high, where S is the proposition that the large majority of outputs of beliefs produced by our cognitive systems are checked many times over, at different times, from different perspectives, and through different perceptual apparatus and media, and we almost always find no contradictions amongst our judgements.

This makes (T&S) an adequate defeater deflector. So if (T&S) is an admissible defeater deflector, Plantinga's second premise is false. (T&S) does not entail R, either on its own or in conjunction with (N&E). It is possible, however minutely, that most of our double and triple and multiple checkings yield coherent but false results. It could be that our true beliefs are few and that these are all amongst those that do not cohere with the majority, or that our true beliefs are few and are all logically independent of the rest of our beliefs, so that they cannot create a contradiction. So (N&E&T) is consistent with ~R. (N&E&T&S) is not obviously irrational. Perhaps it could be argued that (N&E&T&S) is subtly irrational. One reason for thinking so is Plantinga's sceptical challenge, which we have already set aside. I cannot offer a general argument that this combination of beliefs is rational.

Conclusion

In this paper I have shown that Plantinga's reasons for rejecting Millikan's teleosemantics as a defeater deflector are misguided. His arguments do not recognize how novel signals can have mapping conditions, and hence how signals can map environmental conditions that have played no role in generating the success of the signalling system. Once T is properly described, we see that Plantinga's arguments do not show that T is inadmissible as a defeater deflector. We also find that T is not an adequate defeater deflector. But (T&S) is both admissible and adequate.

References

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Reprinted in Philosophical Papers, 2*, 32–52.

Millikan, R. (1984a). *Language, thought, and other biological categories*. Cambridge: MIT.

Millikan, R. (1984b). Naturalist reflections on knowledge. *Pacific Philosophical Quarterly, 65*, 315–334.

Millikan, R. (1995). A bet with Peacocke. In C. MacDonald (Ed.), *Philosophy of psychology: Debates on psychological explanation*. Cambridge: Blackwell.

Millikan, R. (2000). *On clear and confused ideas*. New York: Cambridge University Press.

Millikan, R. (2004). *Varieties of meaning*. Cambridge: MIT.

Millikan, R. (2005). *Language: A biological model*. Cambridge: Cambridge University Press.

Millikan, R. (2006). Useless content. In G. Macdonald & D. Papineau (Eds.), *Teleosemantics*. Oxford: Oxford University Press.

Millikan, R. (2012). Reply to Nick Shea. In J. Kingsbury, D. Ryder, & K. Williford (Eds.), *Millikan and her critics*. Oxford: Blackwell (forthcoming).

Plantinga, A. (1993). *Warrant and proper function*. Oxford: Oxford University Press.

Plantinga, A. (2002). Introduction: The evolutionary argument against naturalism. In J. Beilby (Ed.), *Naturalism defeated?* Ithaca: Cornell University Press.

Plantinga, A. (2011). Content and natural selection. *Philosophy and Phenomenological Research, 83*(2), 435–458.

Shea, N. (2012). Millikan's isomorphism requirement. In J. Kingsbury, D. Ryder, & K. Williford (Eds.), *Millikan and her critics*. Oxford: Blackwell (forthcoming).