

Content and Natural Selection

ALVIN PLANTINGA
University of Notre Dame

Say that *metaphysical naturalism* (call it ‘N’) is the idea, roughly, that there is no such person as God or anything at all like God—or if there is, this being plays no causal role in the world’s transactions. Then conjoin N with E, the view that our cognitive faculties have come to be by way of the processes to which contemporary evolutionary theory direct our attention. I’ve argued elsewhere that N&E is incoherent or self-defeating.¹ In very rough overview, the argument goes as follows. (1) Where R is the proposition that our cognitive faculties are reliable (i.e., produce a substantial preponderance of true over false beliefs in nearby possible worlds), $P(R/N\&E)$ (the probability of R given N&E) is low; (2) anyone who sees that (1) is true and accepts or believes N&E has a *defeater* for R, a defeater that can’t be defeated, and (3) anyone who has an undefeated defeater for R has a defeater for any proposition she believes—including, of course, N&E itself. Therefore, (4) N&E is self-defeating; it is self-referentially incoherent, and hence rationally unacceptable. Call this “The Evolutionary Argument Against Naturalism” (EAAN). In what follows, I want to look critically at premise (2) of the argument. Suppose you accept N&E and see that (1) is true: can the looming defeater be deflected? Is there something else p you do believe or could believe such that your believing p as well as N&E protects R from defeat, for you? That’s the question I mean to investigate.

Assume you accept N&E and you also see that (1) is true. Perhaps in support of (1) you reason as follows. First, in order to avoid irrelevant distractions (not to mention species chauvinism), you reflect, not

¹ *Warrant and Proper Function* (New York: Oxford University Press, 1993), chapter 12; *Warranted Christian Belief* (New York: Oxford University Press, 2000), pp. 227ff; *Naturalism Defeated?* ed. James Beilby (Ithaca: Cornell University Press, 2002), pp. 204ff; most recently, “Naturalism vs. Evolution: a Religion/Science Conflict?” in *God or Blind Nature? Philosophers Debate the Evidence*, online book, ed. Paul Draper, 2007–2008 and *Knowledge of God* with Michael Tooley (Oxford: Blackwell, 2008), pp. 30 ff.

about us, but about some species (perhaps in one of those other cosmoi posited by inflationary scenarios) much like us: they form beliefs, change beliefs, reason, and the like. You assume that N&E holds for them; and you ask what $P(R/N\&E)$ is, specified, not to us, but to them. Then you note that materialism or physicalism with respect to human beings is *de rigueur* for naturalism (contemporary naturalism, anyway); so assimilate physicalism to naturalism. A belief, presuming there are such things,² will then be a physical structure of some sort, presumably a neurological structure.³ Such a structure, of course, will have neurophysiological properties ('NP' properties): the number of neurons and neural connections involved, the strength and rate of neuronal fire at different times and in various parts of the structure, the rate of change of strength and rate of fire in response to differential input, and the like. But it will also have a *propositional content*.⁴ It will be the belief that p for some proposition p : perhaps the belief that Marcel Proust is more subtle than Louis L'Amour. It is easy to see how beliefs thus considered can enter the causal chain leading to behavior; current science gives us a reasonably plausible account of the process whereby volleys of impulses propagated along the efferent nerves cause muscle contraction, motor output, and thus behavior. It is exceedingly difficult to see, however, how they can enter that chain *by virtue of their content*; a given belief, it seems, would have had the same causal impact on behavior if it had had the same NP properties, but different content. So you believe, first, that semantic epiphenomenalism (SE) is likely, given N&E (N construed as including physicalism): $P(SE/N\&E)$ is high.⁵

² A materialist might hold that belief-talk is to be paraphrased into talk about the property of believing; then we could say that human beings sometimes display the property of believing p , for some proposition p , while denying that there are any such things as *beliefs*. For what follows, this difference makes no difference. Eliminativism is also an option for the physicalist. In this paper, though, I'll be assuming that there really are such things as beliefs (or at any rate ways of believing), because what we are investigating is an argument for the claim that a certain belief (or way of believing), namely the belief that N&E, is rationally unacceptable.

³ Beliefs will therefore be mental events or structures, in that they exemplify such mental properties as *having such and such content*, but they remain physical events or structures in that only physical substances are involved in them.

⁴ It is far from obvious that a material or physical structure can *have* a content; see my "Against Materialism," *Faith and Philosophy*, January, 2006.

⁵ This issue, of course, has been heavily canvassed; see in particular Jaegwon Kim's "Precis of *Mind in a Physical World*" pp. 640–642, Barry Loewer's "Comments on Jaegwon Kim's *Mind and the [sic] Physical World*" p. 647, and Kim's "Responses" p. 675, all in *Philosophy and Phenomenological Research*, Vol. LXV, no. 3 (November, 2002).

Second, you also believe that R is unlikely on N&E&SE. For if SE is true, it will not be the case that a false belief causes maladaptive behavior by virtue of its having false content, and it will not be the case that a true belief causes adaptive behavior by virtue of having true content. The truth or falsehood of the belief will then be irrelevant to fitness and thus, so to speak, *invisible* to natural selection; but then it is hard to see how natural selection can promote or enhance or reward true belief (or reliable belief-producing processes) and penalize false belief (or unreliable belief-producing processes). If SE were true, it would be an enormous cosmic coincidence, a stunning piece of not-to-be-expected serendipity, if modification of behavior in the direction of fitness also modified belief-production in the direction of greater reliability. So $P(R/N\&E\&SE)$ is low.

By the theorem on total probability, however,

$$(5) \quad P(R/N\&E) = [P(R/N\&E\&SE) \times P(SE/N\&E)] \\ + [P(R/N\&E\&\sim SE) \times P(\sim SE/N\&E)],$$

i.e., the probability of R on N&E is the weighted average of the probabilities of R on N&E&SE and N&E& \sim SE—weighted by the probabilities of SE and \sim SE on N&E. Inspection shows that if $P(SE/N\&E)$ is high and $P(R/N\&E\&SE)$ is low, then $P(R/N\&E)$ is also low. (For example, if $P(SE/N\&E)$ is .9 and $P(R/N\&E\&SE)$ is .2, then even if $P(R/N\&E\&\sim SE)$ is 1, $P(R/N\&E)$ is .28.) Of course it is ludicrous to assign precise values to any of these probabilities; still our estimates of them can (and should) be guided by (5). So it looks initially as if $P(R/N\&E)$, specified to that hypothetical population and with N&E construed as including materialism, is low.⁶ But if this probability is low with respect to them, it is also low with respect to us.

I. Reductive Materialism and the Conditionalization Problem

I am not proposing here to defend this argument (although I believe it or something like it to be successful); instead, my aim is to address a problem with respect to premise (2), the thought that anyone who sees that (1) is true and accepts or believes N&E has a *defeater* for R. Suppose the argument for (1) is indeed successful: is there some way in which materialism can be reconciled with the rejection of SE? Can the partisan of N&E find some way to ward off semantic epiphenomenalism, thus evading the argument? Yes; she can adopt *reductive materialism*, or *reductionism*, or (given the closure of propertyhood under Boolean construction) *type identity of the mental with the physical*. She

⁶ There is a fuller development of this argument in *Naturalism Defeated?* pp. 211–15.

can then claim that the question asked above—wouldn't a belief, a neuronal structure with content, have the same physical effect, the same effect on nerves and muscles and glands, if it had the same NP properties but a different content?—she can declare that question wholly misleading: it *couldn't* have had the same NP properties but different content. For, she says, consider the property of having as content the proposition *Naturalism is all the rage these days*, and call this property 'C'. According to reductive materialism (hereafter 'RM'), C, like other mental properties, *just is* a certain combination of or Boolean construction on NP properties. It might be a disjunction of such properties; more likely a more complex Boolean construction, perhaps something like

$$(P_1 \& P_7 \& P_{28} \dots) \vee (P_3 \& P_{34} \& P_{17} \& \dots) \vee (P_8 \& P_{83} \& P_{107} \& \dots) \vee \dots$$

(where the P_i are NP properties).⁷ Assume that NP propertyhood is closed under Boolean construction: then for any given content C, there will be an NP property P such that having C is the very same thing as having P. Hence it won't be so much as logically possible that a structure have different content but the same NP properties. And hence the counterfactual *if it had had the same NP properties but different content, then it would have made the same causal contribution to behavior*, may be true, but it will be of dubious relevance. If content properties just are NP properties, there is no reason whatever for thinking content doesn't enter the causal chain leading to behavior. The specter of semantic epiphenomenalism is dispatched.

Just how does this consideration bear on P(R/N&E)? On the above argument for (1), SE was said to be likely; the present claim is that on RM, SE is *unlikely*. The claim is that P(SE/N&E&RM) is low, thus undercutting the argument for a low value for P(R/N&E). So suppose the partisan of N&E also accepts RM: isn't the impending defeater thus deflected? Granted; she believes N&E, and P(R/N&E) is low. But she also believes RM, and if P(R/N&E&RM) is high, doesn't that deliver her from defeat? Where A is *Sam is 40 years old* and B is *Sam lives in Cleveland*, my learning B and seeing that P(A/B) is low doesn't give me a defeater for A if I also believe C: *Sam's wife Suzy just told me that Sam is 40 years old*. True, P(A/B) is low; P(A/B&C), however, is high, and this means that I don't get a defeater for A in learning B. C, we might say, is a *defeater-deflector* with respect to A and B; it is a

⁷ If physical propertyhood is closed under Boolean construction, mental properties will just be physical properties. To accommodate wide content, we may have to suppose that some of the P_i aren't neurophysiological properties, but environmental properties of one sort or another. For ease of exposition I'll ignore this qualification in what follows.

belief C I have such that $P(A/B\&C)$ is high and such that this fact deflects the impending defeat offered by the low value of $P(A/B)$.

A. Is Reductive Materialism Admissible?

Of course not just any belief D I hold such that $P(A/B\&D)$ is high, is a defeater-deflector. Return to $N\&E$. Suppose I accept $N\&E$ and see that $P(R/N\&E)$ is low: there are presumably many propositions Q I accept, such that $P(R/N\&E\&Q)$ is high; but not all of those propositions Q are defeater-deflectors. In objecting to EAAN, Carl Ginet proposes that R itself is a defeater-deflector: “Why isn’t it . . . reasonable for the naturalist to take it as one of the tenets of naturalism that our cognitive systems are on the whole reliable (especially since it seems to be in our nature to have it as a basic belief)?”⁸ But that can’t be right—not, at any rate, as a general strategy. Consider the probabilistic argument from evil against theism: the claim is that $P(G/E)$ is low (where G is the existence of God and E some proposition about evil); the theist could hardly respond that while $P(G/E)$ is indeed low, she also believes G , and (naturally enough) $P(G/G\&E)$ is high, so that she has defeater-deflector for the proposed defeater. If this were sufficient for deflecting a defeater, there wouldn’t be any probabilistic defeaters at all.⁹ R , we might say, is not an *admissible* defeater-deflector. But what about RM ? Is *that* an admissible defeater-deflector? What sorts of beliefs *are* admissible? This is the *conditionalization problem*:¹⁰ which beliefs B are such that if $P(R/N\&E\&B)$ is high, then $N\&E$ and $P(R/N\&E)$ is low or *inscrutable* does not constitute a defeater, for the naturalist, of R ? Which beliefs are defeater-deflectors¹¹ with respect to R and $N\&E$ and

⁸ “Comments on Plantinga’s Two-Volume Work on Warrant”, *Philosophy and Phenomenological Research*, Vol. LV, no. 2, (1995), p. 407. Compare Timothy O’Connor “An Evolutionary Argument Against Naturalism?” *Canadian Journal of Philosophy*, Vol. 24, no. 4, (1994), p. 535: “. . . why can’t she [the naturalist] say that her beliefs on these matters are not limited to $N\&E$ alone, but include O as well, where O is simply a general proposition to the effect that the initial conditions of the development of organic life and the sum total of evolutionary processes (including ones as yet unknown or only dimly understood) were and are such as to render $P(R/N\&E\&C\&O)$ rather high?”

⁹ And surely there are. I assume in the usual way that the thermometer I’ve just purchased is reliable; then I receive notice that it is member of a batch of thermometers 7 out of 10 of which are defective. If I have no other relevant information, I have a probabilistic defeater for my original belief in the reliability of the thermometer.

¹⁰ Pointed out by Richard Otte in “Conditional Probabilities in Plantinga’s Argument”, *Naturalism Defeated?* p. 133.

¹¹ Not defeater-deflectors. The latter would require that one first have a defeater D for R , and then acquire another belief that defeats D . A defeater-deflector, on the other hand, prevents D from being a defeater in the first place.

$P(R/N\&E)$ is low or inscrutable? Well, of course any defeater-deflector D will have to be such that $P(R/N\&E\&D)$ is not low or inscrutable; but which beliefs or propositions D are *admissible*?

This is not a trivial question, as one says when one doesn't really know the answer. But even if we can't easily come up with a rigorous statement of necessary and sufficient conditions for admissibility, we can still see some obvious necessary conditions. R itself is not admissible, and the same goes for any belief equivalent in the broadly logical sense to R (for example, $R \vee 2 + 1 = 4$) as well as any belief that together with N , or E , or their disjunction entails R (for example, $N \supset R$). Other examples of inadmissible beliefs would be *9/10 Americans are reliable and I am an American*, and *if naturalism is true, then 9/10 Americans are reliable*. Further, suppose S believes a proposition q that no rational person in her circumstances would believe—an explicit contradiction, for example, or the denial of an obvious truth of arithmetic: these too aren't admissible. The conclusion of EAAN is that he who accepts $N\&E$ and sees that $P(R/N\&E)$ is low, harbors a certain irrationality in his noetic structure; if that proposition q were admissible, freedom from the irrationality specified by EAAN would merely be purchased at the cost of irrationality elsewhere.

Returning to RM: is it admissible? Can the partisan of $N\&E$ properly add RM to $N\&E$ in order to ward off the impending defeater? RM may seem implausible (I believe it *is* implausible); but is that sufficient for inadmissibility? Well, even if RM is implausible, it is hardly such that there aren't any circumstances in which a rational person could accept it. Considered beliefs about the nature of belief itself can, presumably, be properly added, and RM is one of these.

B. What is $P(R/N\&E\&RM)$?

So RM is admissible. But of course that is not enough for it to be a defeater-deflector with respect to R ; in addition, $P(R/N\&E\&RM)$ must be high. But is it? Is it high enough to deflect the defeater potentially proffered by a low value for $P(R/N\&E)$? I think not; RM, despite its precluding SE, doesn't help. Here too natural selection, in selecting for more adaptive belief-producing processes, won't ordinarily select for more reliable belief-producing processes. Perhaps we can most easily see this as follows. We ordinarily think that creatures who have beliefs, also have complex neural circuitry. Simple prokaryotic creatures, bacteria, for example, probably don't have beliefs (even though they may have structures that function as indicators¹² by covarying with features

¹² See below, pp. 00016.

of the environment); more complex creatures, such as the higher mammals, probably do. So at the ‘bottom’ of the evolutionary scale (apologies for such a politically incorrect term) there aren’t beliefs; at the ‘top’ there are. As we go up the scale, therefore, at some point we start getting actual belief content, something we can properly call a belief, something that is true or false. So suppose we go up the evolutionary scale, starting with creatures that don’t have beliefs at all, until we arrive at the first creatures that do in fact display belief. Of course there will be vagueness: at first there will be the merest glimmer of belief content, something that doesn’t clearly warrant the title ‘belief’ but also doesn’t clearly fail to warrant it. Vagueness won’t matter for present purposes; just consider one of the first occasions on which some creature has what is clearly a belief. For definiteness, imagine that one of the first places where we find actual belief content is in an early member of *C. elegans*.¹³ This small but charismatic beast, we suppose, harbors a neural structure that displays an NP property *P* that constitutes content; *P* is an NP property that is (identical with) the property of having such and such content (perhaps the content Hungry).

Now we may assume that having *P* is adaptive in that it helps to cause adaptive behavior.¹⁴ But (given no more than N&E&RM), we have no reason at all to suppose that this content, the proposition *q* such that *P* is the property *having q as content*, is *true*; it might equally well be false. True, SE is false; the property *having q as content* does indeed enter the causal chain leading to behavior; but it doesn’t matter, as far as adaptiveness goes, whether this first bit of content is true. What matters is only that the NP property in question cause adaptive behavior; whether the content it constitutes is also *true* is simply irrelevant. It can do its job of causing adaptive behavior just as well if it is false as if it is true. It might be true, and it might be false; it doesn’t matter.¹⁵

Given just N&E&RM, therefore, it seems as likely that the first bit of content be false as that it be true. *P* is indeed adaptive; but it is adaptive by virtue of its causing adaptive behavior, not by virtue of its having true content. There is no reason to suppose that first bit of

¹³ Famous for its neural circuitry’s having been completely mapped.

¹⁴ The property itself, naturally enough, doesn’t cause anything; the relevant cause will be the structure that has the property. Following current practice I ignore this distinction in what follows.

¹⁵ Of course I don’t mean that the content actually constituted by *P* is a proposition that might be false; it’s possible, I suppose, that the first bit of content is $2 + 1 = 3$, in which case the first bit of content could not have been false. Here we’re talking *de dicto*, not *de re*; possibly, some true proposition is the first bit of content; equally possibly, some false proposition is.

content is a true proposition. N&E&RM gives us no connection between the truth value of the content and the adaptiveness of the behavior it causes. There would have to be something special about the situation—something beyond N&E&RM—if the property's being adaptive made it more likely than not that the content it constitutes is true. *P*, the property of having this particular content in the relevant kind of situation, was selected for—not because that content was true, but because the behavior *P* causes (caused) in that situation was adaptive. *P* was selected for because in that sort of situation it caused adaptive behavior; but the adaptivity of that behavior doesn't depend on the truth of the content *P* comprises.

Look at the matter from a slightly different perspective. On reductive materialism, any neural structure that is a belief must have at least one property that is both an NP property and also the property of having such and such content. So there will have to be a necessary connection between the neurophysiological aspects of this property, on the one hand, and the proposition constituting that content on the other. More generally, there will have to be a function taking certain kinds of neurophysiology to certain propositions, the ones that constitute the content of the structure displaying the physiology in question. But why suppose these propositions, the values of that function, are *true*? The property *having q as content* is adaptive (we may suppose), but adaptive by virtue of the behavior it causes, not by virtue of its relation to that proposition *q*. So once more there is no reason to think this first bit of content is true rather than false.

Now clearly what holds for that first bit of content will hold for subsequent bits as well. Consider any subsequently exemplified content property *P**; *P** will have been selected for, once again, not because the associated content is *true*, but because *P** causes adaptive behavior in those circumstances. And, just as in the case of that first content property, *P** (given RM) can cause adaptive behavior in those circumstances if false just as well as if true. But then it is not likely that natural selection, in modifying structures that constitute belief (or perhaps the structures that cause beliefs) in the direction of greater adaptiveness, will also modify them in the direction of greater reliability. And what holds for *C. elegans*, naturally enough, will hold for other species as well. We can assume that the content properties displayed by beliefs are adaptive; but it doesn't follow that the content in question is likely to be true. Natural selection, in modifying the content properties of beliefs in the direction of greater adaptiveness, will probably not be modifying belief-producing processes in the direction of greater reliability.

So think again of that hypothetical population and consider a belief *B* with its content *C*; what, given that having *B* (or the belief producing

process that gave rise to it) is adaptive, is the probability that *C* is true? Well, since truth of content doesn't make a difference to the adaptivity of the belief, the belief could be true, but could equally likely be false. We'd therefore have to estimate the probability that it is true as about .5. But then, once more, the probability of this creature's faculties being reliable will be low indeed. Therefore, specified to that population, $P(R/N\&E\&RM)$ will be low—and if low with respect to them, also low with respect to us. Reductive materialism offers a way past semantic epiphenomenalism; it is an admissible addition to N&E; but it doesn't help with the probability of *R* and hence is not a defeater-deflector.

II. Nonreductive Materialism

If *RM* won't do the trick, what about nonreductive materialism ('*NRM*'), perhaps the most popular position in this area? $P(R/N\&E\&RM)$ is low; what about $P(R/N\&E\&NRM)$? According to *NRM*, content properties, like other mental properties, are neither *NP* properties nor Boolean constructions thereon, but a new and different sort of property. Properties of this kind are instantiated by neural structures or events, and in particular by neural structures exhibiting a high degree of complexity. When a neural structure displays a set of *NP* properties of the right degree of complexity and of the right kind, a new property, a mental property, gets instantiated. We might call this new property 'emergent'. It is not, of course, that a new *property* emerges or comes into existence; properties (*pace* existentialism¹⁶) presumably exist necessarily. It is rather that a new sort of property comes to be *exemplified*.

Furthermore, according to the usual varieties of *NRM*, mental properties *supervene* on *NP* properties (together, perhaps, with certain environmental properties, if we wish to accommodate the thought that meaning ain't in the head). There are various kinds of supervenience; perhaps the most relevant, in this context, would be *strong* supervenience. Following Jaegwon Kim¹⁷ let's say that properties of

¹⁶ Existentialism is the view that singular propositions (propositions which, like *Sam is happy*, are directly about individuals), and quidditative properties (properties which, like *being Sam*, directly involve an individual) exist only if the relevant individuals do, and are therefore contingent existents if the relevant individuals exist contingently.

¹⁷ *Mind in a Physical World* (Cambridge: MIT Press, 1999), p. 9.

sort *A* *strongly supervene* on properties of sort *B* just if necessarily, if *P* is a property of type *A*, there is some property *P** of type *B* such that necessarily, if an object exemplifies *P** then it exemplifies *P*.¹⁸ (We may construe the necessity involved as broadly logical or metaphysical necessity, or as causal or nomological necessity, if there is such a variety of necessity.) Accordingly, suppose properties of the sort *has such and such content* supervene on NP properties. Then for the property *having as content the proposition that Proust is more subtle than L'Amour* (call it 'C') there is an NP property *P**¹⁹ such that necessarily, a neurological structure *S* exemplifies C if and only if *S* exemplifies *P**.

Will NRM serve as a defeater-deflector? Turn first to (1). According to NRM, content properties are not NP properties or indeed any kind of physical property at all, although they strongly supervene on such properties. Think again about the first exemplifications of content (perhaps in *C. elegans*). The relevant content property will simply follow (with nomological or broadly logical necessity) from a certain complex NP property—a property, we may assume, that is adaptive. But if the new content property involves *false* content, that won't in the least compromise the adaptivity of the NP property. This property is indeed adaptive; but that is no reason to think the supervening content is true. This new property will be implied with causal or metaphysical necessity by the relevant NP property which, we may assume, is adaptive; but that doesn't give us the ghost of a reason for assuming that the content thus accruing to the structure is *true*. Here natural selection is obliged to take potluck; it selects for adaptive NP properties, but must then accept the content properties, true or false as the case may be, that supervene on them. NRM doesn't specify or imply any connection between content and adaptivity, and indeed no natural connection comes to mind. Consider a population with NP properties on which content supervenes. Having these NP properties is adaptive. Imagine natural selection modifying them in the direction of greater adaptiveness: a different complement of NP properties will arise, as will, presumably, different content by virtue of a different law (or a different instantiation of the same law) connecting NP properties with content properties. But there is no more reason to think this second law will

¹⁸ This definition is consistent with a given property of type *A* being entailed by several different properties of type *B*: thus the subvenient property *P** for pain could be quite different in reptiles and mammals. We can replace the included necessary conditional by the corresponding necessary biconditional if propertyhood is closed under disjunction, and if, for any given property *P* of type *A*, there is a disjunction of the properties *P** of type *B* that entail *P*.

¹⁹ Assuming that propertyhood is closed under Boolean construction.

yield true content than there was to think the first would. Natural selection can modify the NP properties in the direction of greater fitness, but that doesn't mean or make probable that the consequent modification of the supervening content properties is towards truth. Given N&E and NRM, natural selection will not ordinarily modify belief-producing processes of mechanisms in the direction of greater reliability in modifying them towards greater adaptivity.

But then we can deal with (2) very quickly. Take any belief *B* on the part of a member of that hypothetical population: what is the probability, given N&E&NRM, that *B* is true? We can assume that the NP properties on which the content of *B* supervenes are adaptive; they cause adaptive behavior. But as we have seen, there isn't the slightest reason to think that the law connecting those NP properties with the content of *B* will yield or lead to true content. Therefore the probability that *B* is true, once more, will have to be estimated as in the neighborhood of .5. But then it is unlikely that these creatures have reliable belief-producing processes. Therefore NRM, like RM, is an admissible addition to N&E, but it isn't a defeater-deflector.

III. Three More Candidates

Of course there are many more candidates for the post of defeater-deflection, and obviously I can't deal with them all. Instead, I'll consider three candidates drawn, like reductive and nonreductive materialism, from current philosophy of mind. Clearly there are many more; equally clearly, each has several variations. I can therefore deal with no more than a small sample of candidates for defeater-deflection, hoping that what I say will at least offer hints as to how to approach those I don't consider here.

A. Indicator Semantics

There are several varieties of indicator semantics, most notably, perhaps, those of Fred Dretske and Jerry Fodor. We can begin by asking a question: given materialism, how could natural selection be expected to modify belief production in the direction of reliability? How would belief content, behavior, and environment have to be related? First, there would have to be some kind of regular relation between belief content and behavior, a relation such that (roughly, and for the most part) when that content is true, then the behavior it causes is adaptive or anyway not seriously maladaptive, but when it is false, maladaptive, or anyway not adaptive. This would require, second, some kind of regular relation between belief content and environment. Whenever a predator shows up, for example, perhaps a belief with a certain content

shows up too, and perhaps that belief is a part-cause of appropriate behavior—fleeing, for example. But again, the old problem rears its ugly head; it doesn't matter, so far, whether this content is true or false. A certain kind of belief content regularly arises upon the appearance of the predator; this belief (content) causes appropriate action; but it can do that whether it is true or false.

This is where indication or concomitant variation enters the picture.²⁰ Deer tracks in my backyard indicate that deer have run through it; smoke indicates fire; the height of the mercury column indicates the ambient temperature; buds on the trees indicate the coming of spring. When one event indicates or is a natural sign of another, there is ordinarily some sort of causal connection, or at least concomitant variation, between them, by virtue of which the first is reliably correlated with the second. Measles cause red spots and fever, which is why those symptoms indicate measles; there is a causal connection between the height of the mercury column and the temperature, so that the former indicates the latter (and the latter the former).

The nervous systems of organisms contain such indicators. A famous example: when a frog sees a fly zooming by, the frog's brain (so it is thought) displays a certain pattern of neural firing; we could (and sometimes do) call such patterns 'fly detectors'. Another well-known example: some anaerobic marine bacteria have magnetosomes, tiny internal magnets. These function like compass needles, indicating magnetic north. The direction to magnetic north (in the northern hemisphere) has a downward component; hence these bacteria, which can't flourish in oxygen-rich surface water, move towards the more oxygen-free water at the bottom of the ocean.²¹ Of course there are also indicators in human bodies. There are structures that respond in a regular way to blood temperature; they are part of a complex feedback system that maintains a more or less constant blood temperature by (e.g.) inducing either shivering or sweating. There are structures that monitor the amount of sugar in the blood and its sodium content, structures that respond in a regular way to a pattern of light striking the retina, to the amount of food in your stomach, and so on. Presumably there are structures in the brain that are correlated with features of the environment; it is widely assumed that when you see a tree from a certain

²⁰ See Fred Dretske's *Explaining Behavior* (Cambridge, MA: MIT Press, 1988), pp. 54ff. See also William Ramsey's *Representation Reconsidered* (Cambridge, MA: Cambridge University Press, 2007).

²¹ Dretske, op. cit., p. 63 and "Misrepresentation" in *Mental Representation*, p. 163. All is apparently not well with this neat little story: see "South-seeking magnetotactic bacteria in the Northern Hemisphere", by Sheri Simmons, in *Science* 311: 371–374, (Jan. 20, 2006).

distance and angle, there is a distinctive pattern of neural firing in your brain that is correlated with and caused by it.

Now we can, if we like, speak of ‘content’ here; it’s a free country. We can say that the mercury column, on a given occasion (more exactly, the state of affairs consisting in its rising to a certain level on that occasion), has a certain content: the state of affairs correlated with its rising to that level, i.e., the ambient temperature’s being n degrees. (And of course we can say equally well that the ambient temperature’s being n degrees has as content the mercury column’s rising to that level.) We could say, if we like, that those structures in the body that indicate blood pressure or temperature or saline concentration have a content on a given occasion: whatever it is that the structure indicates on that occasion. We could say, if we like, that the neural structure that is correlated with my looking at a tree has a content: its content, we could say, is what it then indicates. We can also, if we like, speak of *information* in these cases: the structure that registers my blood temperature, we can say, carries the information that my blood temperature is thus and so. Of course this sort of content or information doesn’t as such constitute or require *belief* or belief content. Neither the thermostat nor any of its components believes that the room temperature is thus and so; when the sap rises in Vermont maples, neither the maples nor the sap believes that winter is about to end.

Still, might it not be that *some* indications are in fact beliefs? Couldn’t we promote indications, at least some of them, to the status of beliefs? This is the course taken by those who adopt indicator semantics. I’ll concentrate on Dretske’s version; it is developed with real clarity and care. The basic initial idea is that some indicators also become or subvene beliefs. Not, of course, just any old belief—it’s not that the frog’s bug detector becomes or subvenes the belief that, say, Louis L’Amour is the reincarnation of Marcel Proust. No, the basic idea is that the supervening belief content *just is* the indicator content. So if the frog’s bug detector is a belief—if its indicator content also gets to be also belief content—the belief content will be the state of affairs the indicator indicates. I’ll briefly explain and examine Dretske’s account, pointing out some difficulties with it. Then I’ll ask whether this account—call it Dretske Semantics—is an admissible addition to N&E&(RMvNRM) and whether, if it is, it is a defeater-deflector.

Dretske begins with the notion of indication, correlation (perhaps causal) between events of one kind and events of another, and he explains belief in terms of indication plus two additional ideas. First, *function*. Beliefs are representations, and representations essentially involve functions: “The fundamental idea [of representation] is that a system, S, represents a property F, if and only if S has the function of

indicating (providing information about) the F of a certain domain of objects.”²² So not all cases of indication are cases of representation: the fuel gauge in my automobile indicates the force on the bolts holding the tank to the frame, the amount of air in the tank, the air pressure, the altitude, the temperature, the potential across a certain circuit, and many other things; its *function*, however, is to indicate the amount of gasoline in the tank. What it represents, then, is the amount of fuel in the tank; it does not represent those other properties and quantities, fascinating as they may be.

But, just as not every case of indication involves representation, so, according to Dretske, not every case of representation is a case of belief (or “proto-belief”, as he tends to put it). He cites the noctuid moth, which, upon detecting the bursts of high frequency sound emitted by the bat’s sonar, executes evasive maneuvers. Here we have representation; it is the function of those neural structures registering that sound to indicate the presence of bats, to carry the information that bats are present. But these structures, says Dretske, are not beliefs and do not have belief content. Where *C* is a structure representing something or other (and now we come to the second additional idea), belief content is present *only if C causes some motor output or movement M, and the explanation of C’s causing M is C’s carrying the information that it does*. That is not so in the case of those structures in the noctuid moth:

. . . the explanation of why *this C* is causing *this M*, why the moth is now executing evasive maneuvers—has nothing to do with what *this C* indicates about this moth’s surroundings. The explanation lies in the moth’s genes (*Explaining Behavior*, p. 92).

Where, then, *do* we get belief? Where there is *learning*, says Dretske (here learning, on pain of circularity, does not entail or presuppose belief). Consider a bird that learns to peck at a red spot because it is rewarded when it does. At first the bird pecks aimlessly, now at the red spot, now at the black spot, now at the bars of its cage. But then we reward it when it pecks at the red spot. Soon the bird will peck only or mainly at the red spot; it has learned something. What has happened here? Well, the bird had a red spot detector to start with; by virtue of learning, that structure came to cause the bird to peck at the red spot. And the structure in question causes the motor output in question *because* that structure indicates a red spot, carries the information that the figure in front of the bird is a red spot. Here, therefore, we do have

²² *Naturalizing the Mind* (Cambridge, MA: MIT Press, reprint ed., 1997), p. 2.

a case of belief content, and the bird can be said to believe (or proto-believe) that there is a red spot in front of it.

Dretske's complete account of belief can apparently be put as follows:

(D) x is a belief (has belief content) if and only if (1) x is a state of an indicating element E in a representational system (e.g., the event consisting in the system's being 'on') (2) x 's function is to indicate something F , (3) x is in the mode or state it is in when it indicates F (4) x causes some movement M , and (5) the explanation of x 's causing M is that it indicates F .

This is a complex and sophisticated account; Dretske develops it with style and power. Naturally there are problems. First, there is a serious problem with *necessary* beliefs. I believe that $7 + 5 = 12$; nothing, however, indicates that state of affairs, or carries the information that $7 + 5 = 12$. (Indeed, $7 + 5$'s *equaling 12* isn't information; it doesn't reduce the probabilities with respect to anything.) An indicator covaries with what it indicates; when it occurs, what it indicates also occurs (or probably occurs). $7 + 5$'s *equaling 12*, however, always obtains; hence nothing covaries with it; hence nothing indicates it. The problem isn't just that on this account, there can't be distinct but logically equivalent beliefs (so that, e.g., there is only one true necessary belief); that would be bad enough. The problem is rather that there can't be any necessary beliefs at all. On this account no one believes any truths of mathematics, or truths of logic, or any other necessary truths.

There are other difficulties; I won't go into them here.²³ Our present question isn't whether Dretske Semantics (henceforth 'DS') is acceptable; we are asking instead whether it constitutes an admissible addition to N&E and (if so) whether it is a defeater-deflector. Now perhaps there's no reason to doubt that $P(R/N\&E\&DS)$ will be high. (Indeed, this probability may be *too* high: it is hard to see how, given DS, there can be *false* belief.²⁴) Our question, therefore, is whether DS is an admissible addition to N&E&(RMvNRM). The question is whether the

²³ For further difficulties, see my paper "Against Materialism" in *Persons: Human and Divine*, ed. Peter van Inwagen and Dean Zimmerman (Oxford: Clarendon Press, 2007).

²⁴ For problems with Dretske's most recent attempt to deal with this worry, see Barry Loewer, "From Information to Intensionality" in *Mental Representation*, ed. Stephen Stich and Ted Warfield (Cambridge: Blackwell, 1994) pp. 174–79. For animaladversions on an earlier attempt by Dretske, see Jerry Fodor, *A Theory of Content and Other Essays* (Cambridge: MIT Press, 1990), p. 61ff.

devotee of N&E can properly add it to N&E in the interest of fending off the looming defeater. Of course DS is not the bland assertion of R, or that all beliefs are true, or that the process that produces beliefs produces only true belief; those would be inadmissible. It is instead the stipulation that a *given* process *P*, described in some detail, the process described in DS, is in fact the process that produces beliefs; and the fact is that process will produce only or mostly true beliefs. Is *that* belief admissible? That's a wholly non-trivial question, and a proper discussion would take us too far afield, as well as threaten to be inconclusive. For present purposes, therefore, let's concede that it is.

The devotee of N&E isn't home free, however; she can't evade the impending defeater by accepting DS. That is because on DS, sadly enough, there can't be any such belief as naturalism. As you recall, naturalism entails the proposition that theism is false: there is no immaterial, omnipotent, omniscient, perfectly good person. When the naturalist says there is no such thing, presumably the quantifier is to be taken transtemporally: there isn't any such being now, to be sure, but also there neither has been or will be any such being. (And anyway a being who used to exist but then went out of existence wouldn't be at all like God; *God is dead* fails on more than one account.) But then the state of affairs in question holds at every time and place, in which case nothing carries the information that it holds. Therefore no neural structure will carry that information; therefore no neural structure will be an indicator of naturalism; therefore (on DS) there is no such belief as naturalism, and hence no naturalists. Given DS, one who thinks she is a naturalist is mistaken.

But then DS isn't admissible; the partisan of N&E can't sensibly believe N&E&(RMvNRM)&DS. For surely, if she reflects on her beliefs at all, she will see that she believes N. No doubt there are beliefs (associated with wide content, perhaps) such that one can't really tell whether one has them; presumably N isn't one of those. If she is at all reflective, however she can also easily see that DS implies that no one can believe N. But then she would be in the position of believing both that no one can believe N, and that she believes N—and of course she would or could easily know that she believes both these things. Not a pretty sight. True, it requires just a bit of reflection to see these things; but (as I pointed out earlier) the same goes for her acquiring a defeater for R (in seeing that $P(R/N\&E)$ is low) in the first place; one way to avoid acquiring a defeater for R is simply not to form any beliefs about $P(R/N\&E)$. It is only the (reasonably) reflective naturalist who gets a defeater for R. (In the same way, suppose the probability of theism on our total evidence including evil is low, and suppose that's proposed as a defeater for theism: one way to evade the proposed

defeater is to form no beliefs about the probability of theism on total evidence.) Accepting DS isn't really a way to avoid the possession of irrational beliefs; the irrationality just pops up elsewhere.

B. Functionalism

Again, there are several varieties—analytic functionalism,²⁵ psychofunctionalism, machine state functionalism, conceptual role semantics with its own varieties, and still others. According to what we may call 'generic functionalism', what makes a given process (presumably a neurophysiological structure or process) a given mental state (token) is its playing a certain role, perhaps mediating in some way between sensory input and behavioral output, and other mental states. So beliefs, presumably, are neural structures, or token mental states: there is, e.g., my belief that all men are mortal. There is also the property *being the belief that all men are mortal*. Now a neural structure is the belief that all men are mortal just if it has as content the proposition *all men are mortal*. And what brings it about that a given (token) neural state or structure *N* is in fact a belief, and in fact *that* belief (has that content), is the fact that *N* plays a certain complex causal role.

Given just this much, that is, given generic functionalism, what constraints are placed on the content accruing to beliefs? Need they even have anything to do with the condition or environment of the believer? Imagine a believer—for definiteness, a frog, say—that has just one belief. This belief will be a neural structure *N* and its content this belief has is determined by its causal role—its relation to sensory input or stimulation, and behavior output and other mental states (and possibly certain properties of the environment, to accommodate wide content)—though not to other beliefs, there being no other beliefs. Suppose, therefore, that *N* causes fleeing, and is caused by a certain sensory stimulation: a certain pattern of retinal stimulation ordinarily itself caused by the presence of a predator (together with eyes being pointed in the right direction, etc.—fill it out any appropriate way). And add, if you like, that *N*, or the mechanism that produces it, is adaptive; it helps the creature survive and reproduce.

But this much doesn't seem to place any constraints at all on the content enjoyed by *N*. Granted, we might ordinarily assume that the content of *N* will have to do with the surroundings or environment of

²⁵ Analytic functionalism, at least as proposed by David Lewis, is really a set of directions for giving a definition of mental terms. Presumably adding these directions to N&E&(RMvNRM) will not affect P(R/N&E&(RMvNRM)). The results of following these directions, if the directions are correct, will be sentences expressing necessary truths; hence adding those truths to N&E&(RMvNRM) will also have no effect on P(R/N&E&(RMvNRM)).

the creature in question, but that doesn't follow from the description we've got of N and its causal role, nor, as far as I can see, is it made probable by that description. For all we know (given this much) its content might be the proposition that $2 + 1 = 3$ —or any other proposition, for that matter. Of course there will probably be indicators—neural structures correlated with features of the environment such as the presence of a predator: that particular pattern of retinal stimulation might be one of them. But indication is not belief.)

Now suppose another belief gets added: the creature in question acquires another belief N^* . N^* , of course, is also a neural structure; and again its content is determined by the causal role it plays, this time involving also its relation to N . Suppose N^* is caused by a certain pattern of neural stimulation, one normally accompanied by the predator's gaining on the fleeing creature; and suppose N^* together with N causes an increase in the velocity of the flight. N^* is a belief, so of course it has a content. But again, there seems to be no constraint, so far, on what content accrues to N^* —what proposition p gets associated with N^* as its content. p could be the proposition that $3 + 1 = 4$, or any other proposition, including the proposition that $2 + 1 = 5$ (even though that proposition is inconsistent with the content of the first belief N). Again, we can think of the content as supervening on other properties, properties involving the relation of N^* to sensory input, other mental states, behavioral output and N ; but we don't have any constraint on what N^* 's content will be.

We can imagine the same thing, not in the history of evolution, but in the history of an individual human being. The same pattern: the first belief will be a neural structure with a content determined by the causal role that neural structure plays with respect to sensory stimulation and behavioral output and other (nondoxastic) mental states. This seems to place no constraints on what gets to be the content of that first belief. The content of the next belief is determined by the causal role of *that* neural structure, which again places no restrictions on the content of the belief, and so on. From this perspective, it looks as if generic functionalism places no restrictions on content.

We might think that beliefs don't or can't arise one at a time; what you have instead is beliefs arising in clusters. But it is hard to see how this makes any difference. Again, take the first cluster of beliefs: each of its members will be a neural structure; the content of each will be determined by its causal relations to sensory input, behavior output, nondoxastic mental states, and the other beliefs. But what content gets determined for these neural structures? As far as one can see, given just this much, it could be anything; for example a set of propositions from elementary arithmetic could be the content. Functionalism tells us that

there is a function (pardon the pun) from circumstances and causal roles to content assigned; but it doesn't provide any constraints for that function. And wouldn't the same go for any set of beliefs a given person has at a given time? The complex property that involves each member of the set having just the content it does, supervenes on a complex property that involves the causal roles of those beliefs, (i.e., those neural structures); but, at least as far as we can see, this doesn't place any restrictions on what that complex content property might be, i.e., on what content gets assigned to those neural structures.

A natural response, on the part of functionalists, would be to claim that content is in fact determined, at least in part, by indication, as in indicator semantics. A gazelle perceives a stalking cheetah; part of what's involved in this perception is a certain neurological state (we could call it a 'representation') that is causally correlated with the presence of cheetahs or other predators; and this representation has a certain content, perhaps the content "Cheetah there!" This story can be elaborated in various ways, but the main point would be that content is thus connected with representation. Indication, we may say, is central to belief content; it is indication that gives content to neural states; a neural state acquires the content it has by virtue of what it indicates.

We can think of this suggestion as an addition to generic functionalism, a specification of the particular kind of functional role that constitutes or determines content. For first, indication, of course, is not automatically belief; there are many indicators in human bodies that don't issue in belief and don't have belief content, even though we may say that they have indicator content and even representational content. One possibility, therefore, is that the above suggestion holds for indicator content; but nothing follows about belief content. Of course it is open to the functionalist to declare that the suggestion also holds for belief content: at least some indicators get promoted to belief.

A crucial problem, however, still raises its ugly head, and it is the same problem that afflicts the thought that Dretske Semantics can serve as a defeater-deflector. That problem, recall, is that indicators are supposed to carry information. An indicator—of the presence of a predator, for example—must be correlated with the presence of predators; it must be present when predators are present and absent when they are not present. But then it looks as if there will be no indicators of necessary states of affairs, and also none of necessarily non-obtaining states of affairs. Hence there will be no indicators of theism, taken as including the proposition that God exists necessarily, and there will also be no indicators of naturalism. Naturalism is not the sort of state of affairs that can obtain at one time but fail to obtain at another; either it always obtains or it never obtains. Hence the reflective

naturalist cannot sensibly accept this version of functionalism; it entails that there is no such belief as naturalism, but the reflective naturalist presumably knows that she does in fact believe naturalism. Hence functionalism—generic functionalism and indicator functionalism anyway—is not a defeater-deflector.

C. Teleosemantics

Finally, I'd like to consider so-called teleosemantics, still another theory of content. Here the main spokesperson is Ruth Millikan:

For a system to use an inner item as a representation, I propose, is for the following two conditions to be met. First unless the representation accords, *so* (by a certain rule), with a represented, the consumer's normal use of, or response to, the representation will not be able to fulfill all of the consumer's proper functions in so responding—not, at least, in accordance with a normal explanation. (Of course it might still fulfill these functions by freak accident, but not in the normal way) Putting this more formally, that the representation and the represented accord with one another, *so*, is a normal condition for proper functioning of the consumer device as it reacts to the representation.²⁶

The content of the representation (“the represented”) is the condition in the world that must obtain for the consumers to fulfill their proper function(s) in making the normal response(s) to the representation. Consider a couple of examples used as paradigms by Millikan: beaver tail slaps and bee dances. A beaver smartly slaps the water with its tail; beavers within earshot dive beneath the surface and perhaps swim to the underwater entrance of the lodge. Here we have a representational system. The producer is the tail-slapping beaver; the token representation is the tail slap; the consumers are the beavers that dive. Diving is the *normal* response, that is, the one that (by virtue of its adaptive quality) has been selected for in the evolutionary history of the species. There is also a condition—that there is danger nearby—which is necessary if the consumer is to fulfill its proper functions in making the normal response. (If this condition isn't met, diving will be a waste of time and energy that can be better invested in activities likely to maximize expected progeny.) This condition—the condition that must be met if the consumer is to be functioning properly in producing the normal

²⁶ “Biosemantics,” in *Mental Representation*, op. cit., p. 247. This, says Millikan, is the first condition; the second isn't relevant to the current concern.

response—is the *content* of the token representation (the tail slap). Similarly for bee dances: the producer is the dancing bee, the token representation is the particular dance; the consumers are those bees that witness the dance; the normal response is to fly off in the direction indicated by the dance; and the content of the representation, the condition necessary for the response to fulfill its proper function, is the presence of nectar at the location specified.

A problem arises for each example: there are many other conditions necessary to the proper function of the consumer, including, for example, the presence of water, air, gravity, and so on. Millikan deals with this problem by noting that these more general conditions are necessary for the proper function of very many normal responses; the content of the representation is the condition necessary just for *this* response's constituting proper function on the part of the consumers. This has its difficulties; for present purposes we can ignore them.

In the beaver and bee cases the producer is one organism and the consumer another; a large class of representational systems, however, involve organs or systems within a given organism. Thus, for example, in human beings there is a system involving the detection of blood temperature: if the temperature is too low, this system sends a signal that results in shivering and similar appropriate responses; if too high, the result is sweating, seeking shade, and the like. Perhaps another such system would involve as producer some cognitive system *S1* that produces a token representation in the presence of predators. The consumer would be some other cognitive system *S2* that causes a normal response, in this case perhaps fleeing. The content of the token representation would be the environmental condition that is required for the proper function of *S2* in producing fleeing; presumably this condition would be the presence of predators or other danger (fleeing in the absence of this condition would waste resources). Such a system would (or could) involve a belief, something like the belief that predators are present.

Of course many questions arise, and many have been asked; our question is: how does this account apply to the topic at hand? Our question is: does teleosemantics (which for present purposes I shall identify with Millikan's version) provide a defeater-deflector for the naturalist? Once again, this question splits into two: is teleosemantics (T) admissible? And is $P(R/N \& E \& (RM \vee NRM) \& T)$ reasonably high? Turning to the first, note that Millikan's ideas on semantics are part of an attempt to 'naturalize' meaning. It's not very clear just what it is to naturalize meaning, but in her case it requires, among other things, that her basic idea must apply both to beaver slaps and human beliefs, both to bee dances and such beliefs as theism, naturalism, nominalism

and supralapsarianism. That is a pretty tall order; it isn't at all clear how it is to be met or even if it can be met; and if her suggestions don't very naturally apply to beliefs of the sort just mentioned, that probably isn't much of a point against her account. Still, in order to consider whether T is admissible, we must consider how such beliefs fare under that account.

We can see that T is not admissible; we get the same result here as with Dretske Semantics: if T were true, there wouldn't be any such belief as naturalism. The reasons for so thinking are twofold. First, consider Millikan's basic idea and apply it to naturalism as a belief—the belief (N) that there is no such person as God or anything like God. There will have to be the usual elements: a representational system consisting in (1) a producer, (2) the token representation(s) used by one or more consumers, (3) the normal response on the part of the consumer(s), and (4) the state of affairs *S* that is a necessary condition of the consumer's fulfilling its proper function in making the (or a) normal response; the content of the token representation (the token belief) will have to be *S*. The problem is that it's extremely difficult to find or even conceive of any such system. Suppose there is such a system. (1) and (2) present no particular difficulty: we can agree (for present purposes, anyway) that there is such a belief as N, that in one who accepts N there is a token representation *r* whose content is the state of affairs N* consisting in there being no such person as God or anything like God, and that there is something (the producer) that produces *r* in the naturalist. What about the consumer? What would be the consumer for *r*? Presumably it would be either the whole human being (the whole naturalist) or one of her cognitive systems.

Here the problems start: the consuming system must have a normal response, a response that has been selected for by virtue of its adaptive character. Further, the obtaining of N* must be a necessary condition of the consumer's fulfilling its proper function in making that response. If the content of the belief in question is N*, then *there being no such person as God or anything like God* would be that necessary condition.

What could this response possibly be? *Fleeing* or *causing fleeing* is the response on the part of the consumer to the belief that there is a predator present; *there being a predator present* is (we may suppose) an environmental necessary condition of that consumer's functioning properly in causing fleeing. But what is the normal (evolutionarily conferred) response to the belief that naturalism is true? What environmental condition is necessary for that consumer's functioning properly in making that response? What response *r* to naturalism (the belief) is such that a necessary condition of the proper function of whatever

makes r is naturalism's being true? There are many different responses to this belief: despair, relief, indifference, skipping church, or, in the case of the more evangelical naturalists, preaching the truth of naturalism and writing such balanced and subtly nuanced tracts as *The God Delusion* or *God is not Great*. These responses, however, are not normal in Millikan's sense; they haven't been selected for by virtue of their adaptive character. (It is only the occasional member of the Young Atheist's Club whose reproductive prospects will be enhanced by proclaiming naturalism.) And even if there were a normal response to belief in naturalism, the truth of naturalism would not be a necessary condition of the consumer's functioning properly in making that response. So it looks as though, if teleosemantics were true, there wouldn't be any such belief as naturalism.

This appearance is confirmed by a second circumstance. According to teleosemantics, if a representation has content, it must stand in a relevant relation to that condition whose holding is its content. ("First unless the representation accords, *so* (by a certain rule), with a represented, the consumer's normal use of, or response to, the representation will not be able to fulfill all of the consumer's proper functions in so responding . . .") What relation? The natural thing to think, here, is that the representation must carry information about that condition. So, for example, a representation that has as content *predators here* will have to carry information about predators being present if it is to have that content; the probability of there being predators present will have to be greater, given the production of that representation, than it otherwise would have been. According to Millikan, her account must "ride piggy-back" on one or another of the other semantical accounts, and the Dretske account looks most promising. As we saw in considering Dretske's account, however, no neural structure is an indicator of naturalism; hence it follows that no neural structure will carry naturalism as information. If naturalism is true, then for any representation, the (objective) probability of naturalism won't be higher given the occurrence of that representation than it otherwise would have been. But then no neural structure carries the information that naturalism is true (carries naturalism as information); hence (assuming that naturalism, the belief, would be a neural structure) there is no such belief as naturalism. Presumably the same will hold for any other account on which Millikan's could plausibly ride piggy-back. Given teleosemantics, therefore, it is hard to see how there could be a representation whose content was the state of affairs consisting in naturalism. If teleosemantics and naturalism were true, there would be no such thing as the belief that naturalism is true. It follows that teleosemantics, like Dretske's indicator semantics, is not admissible as a defeater-deflector.

IV. Concluding Peroration

The conclusion to be drawn, so it seems to me, is that none of the proposed defeater-deflectors with respect to R and N&E—reductive materialism, nonreductive materialism, functionalism, indicator semantics, teleosemantics—none of these is a successful defeater-deflector. Of course there are other candidates lurking in the neighborhood, too many to consider here. But perhaps we can see how to deal with these others by reflecting on my responses to the ones I considered.²⁷

²⁷ My thanks to Alex Arnold, Andrew Bailey, Paul Boghossian, Brian Boeninger, Marian David, Eric Hagedorn, Don Howard, Marcin Iwanicki, Matthew Kotzen, Anders Kraal, Matthew Lee, Thomas Nagel, Ric Otte, Dan McKaughan, J. Brian Pitts, Josh Rasmussen, Elliott Sober, Sharon Street, Michael Strevens, Luke Van Horn, Fritz Warfield, Roger White, and Rene van Woudenberg; my thanks also to the others whom I have inadvertently overlooked.