An Evolutionary Argument Against Naturalism

by Alvin Plantinga

from part III, chapter 7.IV.B of *Warranted Christian Belief* (2000)

Should awareness of one's evolutionary origins make one skeptical about one's cognitive faculties? In other words, should it lead one to question the reliability of the psychological mechanisms that produce one's beliefs? Darwin was worried that it should, but he did little more than raise the question. Other thinkers since Darwin have developed this idea in more detail, but none with more rigor and sophistication than Alvin Plantinga. Plantinga holds that naturalists, but not theists, should share "Darwin's doubt," not because they believe in evolution—many theists also believe in evolution—but because they believe in *blind* evolution, in evolution that occurs without any prior supernatural planning or concurrent supernatural guidance. According to Plantinga, blind evolution is not likely to lead to reliable cognitive faculties, which means that naturalists who recognize this cannot rationally trust those faculties, and so cannot rationally believe anything at all, including naturalism itself. Plantinga's recommendation to such naturalists would be to reject, not evolution, but naturalism, replacing it with theism.

- Paul Draper ('In Defense of Sensible Naturalism', 2007)

1. Sin and Skepticism

… Thomas Reid, Hume's great contemporary and antagonist, took Hume to be a *skeptic* with respect to external objects, an enduring self, other minds, causality, the past, and so on.[1] As Reid sees him, Hume thinks that there is something *wrong* in believing the things we ordinarily do: it isn't as if Hume simply announces that as a matter of fact we don't really know all we think we know about external objects, causal relations, our own selves. Perhaps that would be bad enough, but there is something much deeper.

We can see what by considering the Hume of the conclusion of Book I of the *Treatise*.[2] Here he isn't coolly announcing, as a mildly interesting fact about us, that fewer of our beliefs constitute knowledge than we ordinarily think. Instead, he finds himself in a sort of existential crisis; he simply doesn't know what to believe. When he follows out what seem to be the promptings and leading of reason, he winds up time after time in a black coal pit, not knowing which way to turn:

> Where am I, or what? From what causes do I derive my existence, and to what condition shall I return? Whose favour shall I court, and whose anger must I dread? What beings surround me? and on whom have I any influence, or who have any influence on me? I am confounded with all these questions, and begin to fancy myself in the most deplorable condition imaginable, inviron'd with the deepest darkness, and utterly depriv'd of the use of every member and faculty. (p. 269)

1 Although Reid's view has been the majority opinion with respect to Humean exegesis, there has always been a minority opinion according to which Hume really wasn't a skeptic at all. This striking divergence is testimony to the fact that Hume is a black enigma: a certain surface clarity masks a deep underlying murkiness that makes confident interpretation impossible.

2 *Treatise of Human Nature*, ed. L. A. Selby-Bigge (Oxford: Clarendon Press, 1951; first published in 1739), pp. 263ff. Subsequent page references to the *Treatise* are to this edition.

Of course this is Hume in his study, sometime before he emerges for that famous game of backgammon. Nature herself, fortunately, dispels these clouds of despair: she "cures me of this philosophical melancholy and delirium, either by relaxing this bent of mind, or by some avocation, and lively impression of my senses, which obliterate all these chimeras. I dine, I play a game of backgammon, I converse, and am merry with my friends" (p. 269).

Still, the enlightened person, Hume thinks, holds the consolations of Nature at arm's length. She knows she can't help acquiescing in the common illusion, but she maintains her skepticism of "the general maxims of the world" and adopts a certain ironic distance, a wary double-mindedness: "I may, nay I must yield to the current of nature, in submitting to my senses and understanding; and in this blind submission I shew most perfectly my sceptical disposition and principles" (p. 269). This is the irony of the human condition: those who are enlightened can see that what nature inevitably leads us to believe is false, or arbitrary, or at best extremely dubious; they also see, however, that even the best of us simply don't have it in them to successfully resist her blandishments. We can't help believing those "general maxims," or if we can, it is only for brief periods of time and in artificial situations. No one can think Humean thoughts about, say, induction, when under attack by a shark or when clinging precariously to a rock face high above the valley floor. …

But the fact is Hume … isn't confessing a frailty or shortcoming, hoping for a cure; he is arguing, as he sees it, from a position of strength or at least insight; the rest of us who unthinkingly accede to the promptings of nature are the ones who suffer from intellectual shortcoming. More than that, we are irrational, in the Humean view, in that reason, carefully preserved from the corrupting influence of everyday attitudes, enjoins this skepticism upon us. To fail to accept it is to fail to follow reason, to go against its teachings, and in that sense to fall into irrationality.

Now Thomas Reid takes issue with Hume (at any rate Hume as he sees him) at just this point. He sees Hume as standing with Descartes in thinking that the deliverances of perception, memory, induction, sympathy, testimony, and any other faculty we might have must be validated before the bar of reason and consciousness. That is, none of these faculties can reasonably be trusted until it has been shown to be reliable by an argument that meets two conditions. First, the argument in question must start from premises that are either self-evident (like elementary truths of arithmetic), or else deliverances of consciousness: such propositions about my own mind as that I seem to see a horse, or am appeared to redly, or believe that the Orkney Islands are north of Aberdeen. Second, the argument must be such that each of its steps is self-evidently valid.

Now Descartes thought that in fact the other sources of belief *could* be legitimated by reason and consciousness. He thought first to establish the reliability of reason itself by giving a reasoned (rational) proof that we have been created by a benevolent God who is non-deceptive (and here we fall into that distressing Cartesian circle), but God would be a deceiver if the world weren't very much like our perceptual faculties reveal it to be. As Reid sees it, Descartes is mistaken at several points; the point of present interest, however, is Descartes's confidence that the reliability of those other sources *can* be

established by reason. It took the work of modern philosophy from Descartes to Hume, so Reid thinks, to show that this is in fact a chimera, a will-o'-the-wisp; it simply can't be done. …

Now one reaction would be to see this condition as interesting and perhaps even mildly regrettable, but of no real importance: these other sources of belief are perfectly acceptable, whether or not we can find arguments of the above sort for their reliability. Reid's Hume, however, takes quite a different tack; he takes it to be a sign of foolishness or error or dupery (in any event, part of the deplorable human condition) to accept the testimony of any source whose veracity hasn't been (or, worse, can't be) established by way of consciousness and reason. He therefore concludes that the *rational* course is to reject these beliefs (given that we can't show in the way in question that their sources are reliable), even if because of nature's imperious edicts we can't actually follow that austere prescription.

This strikes Reid as a piece of consummate arbitrariness:

> The sceptic asks me, Why do you believe the existence of the external object which you perceive? This belief, sir, is none of my manufacture; it came from the mint of Nature; it bears her image and superscription; and, if it is not right, the fault is not mine: I ever took it upon trust, and without suspicion. Reason, says the sceptic, is the only judge of truth, and you ought to throw off every opinion and every belief that is not grounded on reason. Why, sir, should I believe the faculty of reason more than that of perception?—they came both out of the same shop, and were made by the same artist; and if he puts one piece of false ware into my hands, what should hinder him from putting another?[3]

I believe that Reid is substantially right here; the Humean skeptic *is* arbitrary.[4] But this is not the place for a discussion of this point: what I want to argue instead is that Hume has a *different* reason for his skepticism, a reason shared by anyone who concurs with him in agnosticism about our origin and the origin of our cognitive faculties. Suppose, for one reason or another, you give up this idea that we have been created by a benevolent deity. Perhaps with Hume you adopt instead a thoroughgoing agnosticism: there is simply no way to know whether there is any being at all like God, no way to know whether there is a divine being who created the world, no way, indeed, to know anything about the ultimate origin of the world or of the ultimate origin of ourselves and our cognitive faculties. "Our experience," he says, "so imperfect in itself and so limited both in extent and duration, can afford us no probable conjecture concerning the whole of things."[5] *Perhaps* the world owes its existence to intelligent design: just as likely, though (at least as far as we can tell), it owes it to animal or even vegetative generation (perhaps comets are seeds and our world has arisen from one); and there are

3 *Thomas Reid's Inquiry and Essays*, ed. Keith Lehrer and Ronald E. Beanblossom (Indianapolis: Bobbs-Merrill, 1975), pp. 84–85.

4 But perhaps not *entirely* arbitrary; see *Warrant: The Current Debate*, pp. 100ff.

5 *Dialogues Concerning Natural Religion*, ed. Richard Popkin (Indianapolis: Hackett Publishing, 1980), p. 45.

a thousand other possibilities, some of them canvassed with grace and style in the *Dialogues concerning Natural Religion.* Hume's[6] conclusion there, it seems, is that

> In such questions as the present [cosmogony, the origin of the universe], a hundred contradictory views may preserve a kind of imperfect analogy, and invention has here full scope to exert itself. Without any great effort of thought, I believe that I could, in an instant, propose other systems of cosmogony which would have some faint appearance of truth: though it is a thousand, a million to one if either yours or any one of mine be the true system. (*Dialogues*, p. 49)

He adds a bit later that on this topic, "A total suspense of judgment is here our only reasonable resource" (p. 53). Hume so understood has no idea at all how the world got here, how rational creatures such as we ourselves have arisen, and what the origin and provenance of our rational or belief-producing faculties might be.

Now turn to the question whether our cognitive faculties are reliable and do, in fact, produce for the most part true belief. Given Hume's complete agnosticism about the origins of his cognitive faculties, something like his deeply agnostic attitude to that question is no more than sensible. For suppose Hume asks himself how likely it is that our cognitive faculties are reliable, given his views (or rather lack of views) about the origin and provenance of ourselves and those faculties. What is the probability that our faculties produce the considerable preponderance of true belief over false required by reliability, given his views of their origin and purpose (if any)? I should think he would have to say that this probability is either low or inscrutable—impossible to determine. From his point of view, there are innumerable scenarios, innumerable ways in which we and our cognitive faculties could have come into being: perhaps we have been created by God, but perhaps we and the world are the result of some kind of vegetative principle, or a result of copulation on the part of animals we have no knowledge of, or the result of Russell's accidental collocation of atoms, or of . . . . On many of these scenarios, our cognitive faculties wouldn't be reliable (although they might contribute to fitness or survival); perhaps on others they would be reliable; on balance, one just wouldn't know what to think about this probability.

We can see this more fully as follows. Let R be the proposition that our cognitive faculties are reliable: now what is the likelihood of R? As Reid points out, we all instinctively believe or assume that our cognitive faculties are indeed reliable; but what is the probability of that assumption, given the relevant facts? Well, what are the relevant facts? First, they would be facts about those faculties: the probability of R given (relative to) the population of China would not be relevant. And presumably the relevant facts would be facts about how these faculties originated; whether they were designed; if so, by whom and with what end in view; what constraints governed their development; and what their purpose and function is, if, indeed, they have a purpose and function. Were they, as Reid thought, created in us by a being who intends that they function reliably to give us knowledge about our environment, ourselves, and God himself—all the knowledge needed for us to attain shalom, to be the sort of beings God intended us to be? On that scenario, the purpose of our cognitive faculties would be (in

6 Or at any rate Philo's; I make no pretense to settle the question of who speaks for Hume in the *Dialogues*, something Hume artfully conceals.

part, at least) to supply us with true beliefs on those topics, and (given that they are functioning properly) there would be a high probability of their doing just that.

Did they, by contrast, arise by way of some chance mechanism, something like the mindless swerve of atoms in the Democritian void? What is the likelihood, on *that* possibility, that our cognitive faculties are reliable? Well, you might think it pretty low. More likely, you may think that you simply can't say what that probability is: perhaps it is high (though presumably not very high), perhaps it is low; you simply can't tell.[7] There will be many more such scenarios, says Hume, some involving vegetative origin, some copulative origin, some still other kinds of origin; with respect to them, too, the probability that our cognitive faculties are reliable is simply inscrutable. So first, Hume thinks his grasp of the whole set of relevant scenarios is at best infirm; second, with respect to many of these scenarios, those possible origins, the probability of R is inscrutable; and finally, the probability with respect to any of these scenarios that it is in fact the truth of the matter is also, as far as Hume is concerned, quite inscrutable.

But that means that the probability of R, given Hume's agnosticism, is also inscrutable for Hume. Let F be the relevant facts about their origin, purpose, and provenance: my claim is that, for Hume, P(R/F) (the probability of R on F) is inscrutable. He simply doesn't know what it is and has no opinion about its value, although presumably it wouldn't be very high. Another way to put it: the probability of R, given Hume's agnosticism, is inscrutable.

And that gives Hume a reason to be agnostic with respect to R as well; it gives him a reason to doubt that R is, in fact, true. For our cognitive faculties, our belief-producing mechanisms, are a bit like measuring instruments (more exactly, measuring instruments under an interpretation). Our faculties produce beliefs; for each belief, there is the content of that belief, the proposition believed, a proposition that is true if and only if the belief is true. Now a state of a measuring instrument (relative to a scheme of interpretation) can also be said (in an analogically extended sense) to have content. For definiteness, consider a thermometer and suppose its pointer is resting on the number 70. Given the natural scheme of interpretation, this state can be said to have the content that the ambient temperature is 70˚ F. And of course a thermometer is *reliable* only if the propositions it delivers in this way are for the most part true, or nearly true.

Imagine, then, that you embark on a voyage of space exploration and land on a planet revolving about a distant sun. This planet has a favorable atmosphere, but you know little more about it. You crack the hatch, step out, and immediately find something that looks a lot like a radio; it periodically emits strings of sounds that, oddly enough, form sentences in English. The sentences emitted by this instrument express propositions only about topics of which you have no knowledge: what the weather is like in Beijing at the moment, whether Caesar had eggs on toast on the morning he crossed the Rubicon, whether the first human being to cross the Bering Strait and set foot on North America was left-handed, and the like. A bit unduly impressed with your find, you

7 We aren't thinking here of Bayesian personal probability but of some kind of objective probability, the sort of probability Hume has in mind when he says that "it is a thousand, a million to one if either yours or any one of mine be the true system."

initially form the opinion that this quasi radio speaks the truth: that is, the propositions expressed (in English) by those sentences are true. But then you recall that you have no idea at all as to what the purpose of this apparent instrument is, whether it *has* a purpose, or how it came to be. You see that the probability of its being reliable, given what you know about it, is for you inscrutable. Then (in the absence of investigation) you have a *defeater* for your initial belief that the thing does, in fact, speak the truth, a reason to reject that belief, a reason to give it up, to be agnostic with respect to it. Relative to your beliefs about the origin, purpose, and provenance of this apparent instrument, the probability that it is a reliable source of information is low or (more likely) inscrutable. And that gives you a defeater for your original and hasty belief that the thing really does speak the truth. If you don't have or get further information about its reliability, the reasonable course is agnosticism about that proposition.

The same goes, I think, in the case of Humean views (or non-views) about our origins and the origin and purpose, if any, of our cognitive faculties. Suppose I join Hume in that agnosticism. Then P(R/F) is for me inscrutable (as for Hume); I have no idea what the probability of my faculties being reliable is, given the relevant facts about their origin and purpose. But then I have a defeater for my original belief or assumption that my faculties are in fact reliable. If I have or can get no further information about their reliability, the reasonable course for me is agnosticism with respect to R, giving it up, failing to believe it. It isn't that rationality requires that I believe its *denial*, but it does require that I not believe *it*.

Suppose, therefore, that I *am* agnostic with respect to R: I believe neither it nor its denial. And now consider any belief *B* I have: that belief, of course, will be a deliverance of my cognitive faculties. However, I don't believe that my cognitive faculties are reliable—not because I've never thought about the question, but because I *have* thought about it and seen that P(R/F) is inscrutable for me. Well, what does rationality require with respect to this belief *B*? The clear answer seems to be that I have a defeater for this belief too, a reason to withhold it, to be agnostic with respect to it. Perhaps it isn't possible, given my nature, that I *be* agnostic with respect to it, at least much of the time; as Hume says, nature may not permit this. Still, this agnosticism is what reason requires, just as Hume suggests (though for different reasons). And we can take one further step with Hume. Because *B* is just *any* belief I hold—because I have a defeater for just any belief I hold—I also have a defeater for my belief that I *have* a defeater for *B*. This universal, all-purpose defeater provided by my agnosticism is also a defeater for *itself*, a self-defeating defeater.[8] And hence this complex, confusing, multilayered, reflexive skepticism Hume describes, a skepticism in which I am skeptical of my beliefs and also of my doubts, and of the beliefs that lead to those doubts, and of my doubts with respect to those doubts, and the beliefs leading to *them*. Thus the true skeptic will be skeptical all the way down; he "will be diffident of his philosophical doubts, as well as his philosophical conviction."

8 Of course this raises problems: if I have a defeater-defeater (a defeater for my defeater for R), then don't I thereby *lose* my defeater for R? Am I back where I was before I acquired the defeater for R? No; for my defeater-defeater is also a defeater for R. For explanation and detail, see part IV, section E, "The Dreaded Loop," from my "Naturalism Defeated," presently unpublished.

Here we can imagine the following response: "Hey, hang on a minute! You said Hume and any similarly situated agnostic has a defeater for R, a belief to which he is inclined by nature—and you added that the rational course for them therefore is to give up belief in R—*provided they have no other information* about the reliability of their faculties. But what about that strong natural inclination to believe that our faculties are in fact reliable? Doesn't *that* count as 'other information'?" According to Reid (who might object to being pressed into service in defense of Hume), this belief in the reliability of our faculties is a *first principle*:

> Another first principle is—*That the natural faculties, by which we distinguish truth from error, are not fallacious*. (275)

He goes on:

> If any truth can be said to be prior to all others in the order of nature, this seems to have the best claim; because, in every instance of assent, whether upon intuitive, demonstrative, or probable evidence, the truth of our faculties is taken for granted. (277)

Surely there is truth here: this conviction is one normal human beings ordinarily have, and, as Reid gleefully points outs, even skeptics also seem to assume, in the course of ordinary daily living, to be sure, but most poignantly when proposing their skeptical arguments, that their faculties are functioning reliably. Very few skeptics, in offering their skeptical arguments, preface the argument by saying something like, "Well, here is an argument for general skepticism with respect to our cognitive faculties; of course I realize that the premises of this argument are themselves produced by cognitive faculties whose reliability the conclusion impugns, and of whose truth I am therefore extremely doubtful."

But our question is whether this belief can sensibly be pressed into service as information that can defeat the defeater provided for R by Hume's agnosticism about the origin and provenance of ourselves and our faculties. As Reid clearly sees, it cannot. If the general reliability of our cognitive faculties is under question, we can't hope to answer the question whether they *are* reliable by pointing out that these faculties themselves deliver the belief that they are, in fact, reliable. "If a man's honesty were called into question," says Reid, "it would be ridiculous to refer it to the man's own word, whether he be honest or not" (276). Concede that it is part of our nature to assume R; concede further that it is part of our nature to take R in the *basic* way, so that this conviction is not given or achieved by argument and evidence but comes with our mother's milk; concede still further, if you like, that this belief is produced by our cognitive faculties functioning properly. None of this, clearly enough, can serve to defeat the defeater for R provided by Hume's agnosticism. That is because any doubt about our cognitive faculties generally is a doubt about the specific faculty that produces this conviction; therefore we can't allay such a doubt by appealing to the deliverances of that faculty.[9]

9 The same goes, naturally enough, for the suggestion that we try to determine by scientific means whether our cognitive faculties are reliable; any such attempt could proceed only by reliance on the very faculties whose reliability is at issue.

2. Naturalism and Lack of Knowledge

Agnosticism with respect to our origins is one way to reject the theistic belief that we human beings have been created in the image of God: as we have seen, agnosticism with respect to origins destroys knowledge. There is another way to reject the belief in question: by accepting a belief incompatible with it, for example, philosophical or metaphysical naturalism. As Bas van Fraassen notes, it isn't easy to say precisely what naturalism *is*;[10] for present purposes, suppose we take it to be the view that there is no such person as God, nor anyone or anything at all like him (it isn't that you believe, for example, that there are one or more finite gods). Paradigm cases of naturalism would be the views of Daniel Dennett in *Darwin's Dangerous Idea*[11] or Bertrand Russell in "A Free Man's Worship": you think that "man is the product of causes which had no prevision of the end they were achieving, that his origin, his growth, his hopes and fears, his loves and his beliefs, are but the outcome of accidental collocations of atoms."[12] (Perhaps you even go so far as to add, with Richard Dawkins, that the very idea that there is such a person as God is really a kind of cognitive virus, an epistemic sickness or disease, distorting the cognitive stance of what would otherwise be reasonable and rational human beings.[13]) Unlike Hume, therefore, you are not agnostic as to whether there is such a person as God or any being at all like him; you think there is not.

There is likely to be a further difference between you and Hume. Having rejected theism, Hume had no comparable story to put in its place: he was left with no idea as to how humanity arose, under what conditions our cognitive faculties came to be, and so on. The contemporary naturalist, however, is in a different condition; for naturalism now sports a shared myth or story about ourselves and our origins, a set of shared beliefs about who we are, where we come from, and how we got here. The story is familiar; I shall be brief. We human beings have arrived on the scene after millions, indeed, billions of years of organic evolution. In the beginning, there was just inorganic matter; somehow, and by way of processes of which we currently have no grasp, life, despite its enormous and daunting complexity at even the simplest level, arose from nonliving matter, and arose just by way of the regularities studied in physics and chemistry. Once life arose, random genetic mutation and natural selection, those great twin engines of evolution, swung into action.[14] These genetic mutations are multiply random: they weren't intended by anyone, of course, but also were not directed by any sort of natural teleology and do not arise at the behest of the design plan of the organism. They are "not in a response to the needs of the organism" (Ernst Mayr); they

10 See his "Science, Materialism, and False Consciousness," in *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, ed. Jonathan Kvanvig (New York: Rowman and Littlefield, 1996).

11 New York: Simon and Schuster, 1995.

12 In *Why I Am Not a Christian* (New York: Simon and Schuster, 1957), p. 107.

13 "Viruses of the Mind," in *Dennett and His Critics: Demystifying Mind*, ed. Bo Dahlbom (Oxford: Blackwell, 1993), pp. 13ff. As evidence for the virulence and tenacity of this virus, Dawkins cites the fact that it took Sir Anthony Kenny (as learned and sapient a person as we can easily find), a very long time to fight his way clear of it. Others may wonder whether the virus is all Dawkins says it is, given that Dawkins himself apparently escaped it long ago.

14 Various other mechanisms (e.g., genetic drift and neutral evolution) have been proposed, but these two remain the favorites.

just unaccountably appear. Occasionally, some of them yield an adaptive advantage; their possessors come to predominate in the population, and they are passed on to the next and subsequent generations. In this way, all the enormous variety of flora and fauna we behold came into being.

Including ourselves and our cognitive systems. These systems and the underlying mechanisms have also been selected for, directly or indirectly, in the course of evolution. Consider, for example, the mammalian brain in all its enormous complexity. It could have been directly selected for in the following sense: at each stage in its development, the new stage (by virtue of the structures and behaviors it helped bestow) contributed to fitness and conferred an evolutionary advantage, giving its possessors a better chance of surviving and reproducing. Alternatively, at certain stages new structures (or new modifications of old structures) arose, not because they were themselves selected for, but because they were genetically associated with something else that *was* selected for (pliotropy). Either way these structures were not selected for their penchant for producing true beliefs in us; instead, they conferred an adaptive advantage or were genetically associated with something that conferred such an advantage. And the ultimate purpose or function, if any, of these belief-producing mechanisms will not be the production of true beliefs, but *survival*—of the gene, genotype, individual, species, whatever.

If you are a naturalist and also believe these things, then you are what I shall call an *ordinary* naturalist.[15] In chapter 12 of *Warrant and Proper Function* (WPF), I argued that an ordinary naturalist is like Hume in that she has a defeater for any belief she holds—including, ironically enough, ordinary naturalism itself, so that ordinary naturalism is self-defeating.[16] I shall not repeat that argument; instead, I will take this opportunity to make some corrections, simplifications, and additions. …

In essence, the main argument is for the conclusions that P(R/N&E&C) – which I'll abbreviate as P(R/N) – is either low or inscrutable; in either case, so I argued, one who accepts N – and also grasps the argument for a low or inscrutable value of P(R/N) – has a *defeater* for R. This induces a defeater, for him, for any belief produced by his cognitive faculties, including N itself; hence, ordinary naturalism is self-defeating. Now I argued that P(R/N) is low or inscrutable by noting first that natural selection isn't interested in *true belief* but in *adaptive behavior* (taken broadly), so that everything turns on the relation between belief and behavior. I then presented five mutually exclusive and jointly exhaustive possibilities for the relation between belief and behavior, arguing with respect to each possibility P_i that $P(R/N\&P_i)$ is low or inscrutable, yielding the result that P(R/N) is low or inscrutable. …

---

[15] Daniel Dennett's book *Darwin's Dangerous Idea* is a paradigm of ordinary naturalism as well as naturalism simpliciter; the same goes for Richard Dawkins's *The Blind Watchmaker* (London: W. W. Norton, 1986). For animadversions on *Darwin's Dangerous Idea* (and on Darwin's dangerous idea), see my "Dennett's Dangerous Idea," *Books and Culture* (May-June 1996); for a powerful animadversion on the first but not the second, see Jerry Fodor's "Deconstructing Dennett's Darwin," in *Mind and Language* 11, no. 3 (September 1996), pp. 246–62.

16 See James Beilby, ed. *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism* (forthcoming) for fascinating objections to and critical comments on this argument, along with my reply.

Can we mount an argument from the evolutionary origins of the processes, whatever they are, that produce these beliefs to the reliability of those processes? Could we argue, for example, that these beliefs of ours are connected with behavior in such a way that false belief would produce maladaptive behavior, behavior which would tend to reduce the probability of the believers' surviving and reproducing?[17] No. False belief doesn't by any means guarantee maladaptive action. Perhaps a primitive tribe thinks that everything is really alive, or is a witch or a demon of some sort; and perhaps all or nearly all of their beliefs are of the form *this witch is F* or *that demon is G*: *this witch is good to eat*, or *that demon is likely to eat me if I give it a chance.* If they ascribe the right properties to the right witches, their beliefs could be adaptive while nonetheless (assuming that in fact there aren't any witches) false.[18] Also, of course, there is the fact that behavior, if it is partly produced by belief, is also partly produced by desire: it is belief and desire, along with other things, that together produce behavior. But then clearly there could be many different systems of belief and desire that yield the same bit of adaptive behavior, and in many of those systems the belief components are largely false; there are many possible belief-desire systems that yield the whole course of my behavior, where in each system most of the beliefs are false. The fact that my behavior (or that of my ancestors) has been adaptive, therefore, is at best a third-rate reason for thinking my beliefs mostly true and my cognitive faculties reliable—and that is true even given the commonsense view of the relation of belief to behavior. So we can't sensibly argue from the fact that our behavior (or that of our ancestors) has been adaptive, to the conclusion that our beliefs are mostly true and our cognitive faculties reliable. …

In either case, however, doesn't the ordinary naturalist—at any rate, one who sees that P(R/N) is low or inscrutable—have a defeater for R, and for the proposition that his own cognitive faculties are reliable? I say he does. To see how, we must note some analogies with clear cases. … Return to that voyage of space exploration and the radio-like device that emitted sounds that constitute English sentences, sentences that express propositions of whose truth value you are ignorant. At first, you were inclined to believe these propositions, if only because of shock and astonishment. After a bit of cool reflection, however, you realize that you know nothing at all about the purpose, if any, of this instrument, or who or what constructed it. The probability that this device is reliable, given what you know about it, is low or inscrutable; and this gives you a defeater for your initial belief that the instrument indeed speaks the truth. Consider

[17] Thus Quine: "There is some encouragement in Darwin. If people's innate spacing of qualities is a gene-linked trait, then the spacing that has made for the most successful inductions will have tended to predominate through natural selection. Creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind" ("Natural Kinds," in *Ontological Relativity and Other Essays* [New York: Columbia University Press, 1969], p. 126).

[18] Objection: in any event, these tribespeople would be ascribing the right properties to the right things, so that their beliefs are, in some loose sense, accurate, even if strictly speaking false. Reply: by further gerrymandering, we can easily find schemes under which their beliefs would lead to adaptive behavior (thus being functionally equivalent with respect to behavior to the true scheme) but are not accurate even in this loose sense. There are schemes of this sort, in fact, in which the properties ascribed are logically incapable of exemplification. They think everything is a witch; perhaps, then, their analogue of property ascriptions involves ascribing certain sorts of *witches* (rather than properties). (One of these witches, for example, is such that, as *we* would put it, if a thing *has it*, then that thing is red.) Then their beliefs will not be accurate in the above sense and will indeed be necessarily false.

another analogy. You start thinking seriously about the possibility that you are a brain in a vat, being subjected to experiment by Alpha Centaurian cognitive scientists in such a way that your cognitive faculties are not, in fact, reliable. For one reason or another, you come to think this probability is greater than .5; then you have a defeater for your belief that your cognitive faculties are reliable. Suppose instead that you think this is a genuine possibility, but you can't make any estimate at all of its likelihood, so that you can't make any estimate at all of the probability that your faculties are reliable: as far as you can tell, the probability could be anywhere between and 1. Then too you have a defeater for your natural belief that your cognitive faculties are reliable.

The same goes for the naturalist who realizes that P(R/N) is low or inscrutable. With respect to those factors crucially important for coming to a sensible view of the reliability of his belief-producing mechanisms—how they were formed and what their purpose is, if any—he must concede that the probability that those faculties are reliable is at best inscrutable. Unless he has some other information,[19] the right attitude would be to withhold R. But then something like Hume's attitude toward my beliefs would be the appropriate one. I recognize that I can't help forming most of the beliefs I do form; for example, it isn't within my power, just now, to withhold the belief that there are trees and grass outside my window. However, because I now do not believe that my cognitive faculties are reliable (I withhold that proposition), I also realize that these beliefs produced by my cognitive faculties are no more likely to be true than false: I therefore assume a certain skeptical distance with respect to them. And, because my doubts about my beliefs themselves depend on my beliefs, I also assume a certain skeptical distance with respect to these doubts, and with respect to the beliefs prompting those doubts, and with respect to the beliefs prompting the doubts about those doubts. . . . The ordinary naturalist, therefore, should join Hume in this same skeptical, ironic attitude toward his beliefs. This holds, of course, for N itself; for this reason, we might say that N is self-defeating, in that if it is accepted in the ordinary way, it provides a defeater for itself, a defeater that can't be defeated.[20] …

By way of conclusion: … If I reject theism in favor of ordinary naturalism, and also see that P(R/N) is low or inscrutable, then I will have a defeater for any belief I hold. If so, I will not, if forming beliefs rationally, hold any belief firmly enough to constitute knowledge. The same goes if I am merely agnostic as between theism and ordinary naturalism. And the same goes if I am agnostic about my origin and the origin of my cognitive faculties. So rejection of theistic belief doesn't automatically produce skepticism: many who don't believe in God know much. But that is only because they don't accurately think through the consequences of this rejection. Once they do, they will lose their knowledge; here, therefore, is another of those cases where, by learning more, one comes to know less.

---

[19] And how could he have or get other information? Any such information would consist in beliefs that were a product of his cognitive faculties, but he has a defeater for the reliability of those faculties and hence for any belief produced by them.

[20] See chapter 12 of WPF, and "Naturalism Defeated." The defeater can't be defeated because any defeater would arise from the very faculties or belief-producing processes in question. For example, the defeater might take the form of an *argument*, perhaps for the conclusion that those belief-producing processes are reliable after all. But then I would have the same defeater for each of the premises of this argument, as well as for my belief that if the premises are true, then so is the conclusion.

… But isn't there a problem, here, for the naturalist? At any rate for the naturalist who thinks that we and our cognitive capacities have arrived upon the scene after some billions of years of evolution (by way of natural selection and other blind processes working on some such source of genetic variation as random genetic mutation)? The problem begins in the recognition, from this point of view, that the ultimate purpose or function of our cognitive faculties, if they have one, is not to produce true beliefs, but to promote reproductive fitness.[21] What our minds are *for* (if anything) is not the production of true beliefs, but the production of adaptive behavior. That our species has survived and evolved at most guarantees that our behavior is adaptive; it does not guarantee or even suggest that our belief-producing processes are reliable, or that our beliefs are for the most part true. That is because our behavior could be adaptive, but our beliefs mainly false. Darwin himself apparently worried about this question: "With me," says Darwin,

> the horrid doubt always arises whether the convictions of man's mind, which has been developed from the mind of the lower animals, are of any value or at all trustworthy. Would any one trust in the convictions of a monkey's mind, if there are any convictions in such a mind?[22]

[21] As evolutionary psychologist David Sloan Wilson puts it, "the well-adapted mind is ultimately an organ of survival and reproduction" (*Darwin's Cathedral* [Chicago: University of Chicago Press, 2002], p. 228).

[22] Letter to William Graham, Down, July 3rd, 1881. In *The Life and Letters of Charles Darwin Including an Autobiographical Chapter*, ed. Francis Darwin (London: John Murray, Albermarle Street, 1887), Volume 1, pp. 315-16. Evan Fales has suggested that Darwin is thinking, here, not of belief generally, but of religious and philosophical convictions and theoretical beliefs. If he is right, Darwin's doubt would not extend to everyday beliefs to the effect, e.g., that bread is nourishing but mud is not, but to religious and philosophical beliefs—such as naturalism.