

Knowledge of God

Alvin Plantinga
and
Michael Tooley

 **Blackwell**
Publishing

Contents

Acknowledgements	ix
1 Against Naturalism <i>Alvin Plantinga</i>	1
2 Does God Exist? <i>Michael Tooley</i>	70
3 Reply to Tooley's Opening Statement <i>Alvin Plantinga</i>	151
4 Reply to Plantinga's Opening Statement <i>Michael Tooley</i>	184
5 Can Robots Think? Reply to Tooley's Second Statement <i>Alvin Plantinga</i>	218
6 Closing Statement and Response to Plantinga's Comments <i>Michael Tooley</i>	233
Bibliography	249
Index	257

1

Against Naturalism

Alvin Plantinga

I Theism

Our topic is theistic belief, i.e., belief in God, and, more particularly, the epistemology of theistic belief. Now the main part of my initial contribution will be an epistemological attack on one of the two principal alternatives to theism: philosophical naturalism. First, however, I must say something to characterize theism. Following that, I'll argue that theism has a significant epistemic virtue: if it is true, it is (very likely) warranted; this is a virtue naturalism most emphatically lacks. Then I'll deliver a three-part indictment against philosophical naturalism. I'll argue that (1) if naturalism were true, there would be no such thing as *proper function*, and therefore also no such thing as malfunction or dysfunction. Hence there would be no such thing as health or sickness, sanity or madness; further, and in this epistemological context crucial, there would be no such thing as knowledge. That's bad enough, but there's worse to follow: I'll argue (2) that the naturalist is committed to the sort of deep and debilitating skepticism according to which he can't trust his cognitive faculties to furnish him with mainly true beliefs; he has a *defeater* for whatever he believes, including naturalism itself. And (3) I'll argue that naturalism, insofar as it implies materialism about human beings, has no room for the essential features of our mental life, including in particular *belief*.

A. Theistic belief: what is it?

According to classical theistic belief—classical Muslim and Jewish as well as Christian belief—first of all there is God, the chief being of the universe, who has neither beginning nor end. Most important, God is *personal*. That is,

God is the kind of being who is conscious and enjoys some kind of awareness of his surroundings (in God's case, that would be everything). Second (though not second in importance), a person has loves and hates, wishes and desires; she approves of some things and disapproves of others; she wants things to be a certain way. We might put this by saying that persons have *affections*. A person, third, is a being who has *beliefs* and, if fortunate, knowledge. We human beings, for example, believe a host of things. At the moment I believe that I am typing on my computer, that I just had breakfast, that outside it is bright and sunny, that I recently went rock climbing, that I live in Indiana, and so on. I also have a host of beliefs about things more distant from myself: that Beijing is larger than Chicago, that scientists seem to believe that quantum mechanics is highly confirmed, that there once was a war between the Athenians and the Spartans, that even the simplest forms of life are enormously complex, that there is such a person as God, and a thousand other things.

Persons, therefore, have beliefs and affections. Further, a person is a being who has *aims* and *intentions*; a person aims to bring it about that things should be a certain way, intends to act so that things will be the way he wants them to be. Thus I intend to bring it about that my part of this book is written, and written by me. (Put less pedantically, I intend to write my part of this book.) Finally, persons can often act to fulfill their intentions; they can bring it about that things are a certain way; they can cause things to happen. To be more technical (though not more insightful or more clear), we might say that a person is a being who can actualize states of affairs. Persons can often act on the basis of what they believe in order to bring about states of affairs whose actuality they desire.

So a person is conscious, has affections, beliefs, and intentions, and can act. Of course this would be a well-formed, paradigmatic person. Disease or malfunction can deprive a person of one or another of the above characteristics. Due to malfunction a person may lose affect, so that nothing seems either good or bad, desirable or undesirable. A person in a coma is not conscious and cannot act; perhaps a person in a coma also lacks belief or knowledge. The point is that a properly functioning well-formed person will display these properties, not that every person must display them at every time at which she exists.

First, therefore, God is a person. But second, unlike human persons, God is a person without a body.¹ He acts, and acts in the world, as human beings do, but, unlike human beings, not by way of a body. Rather, God acts just by *willing*: he wills that things be a certain way, and they are that way. (God said "Let there

¹The Christian doctrine of the Trinity introduces complications here: the second person of the Trinity had, and indeed *has*, a body. Here I propose to avoid these complications; I'll use the word 'God' as a name of the first person of the Trinity.

be light”; and there was light.) You and I can move our limbs just by willing;² but we can’t just by willing cause it to be the case that Lake Michigan warms up by 10 degrees, or that it’s sunny and pleasant at the top of Mt. Rainier. God isn’t subject to any such limitations; whatever he wills must necessarily come to pass. God is all-powerful (‘omnipotent’). Of course he can’t cause something logically impossible. He can’t bring it about that there is a married bachelor, or that $7 + 5 = 14$. And he also can’t cause a person to do something or other *freely*: if God *causes* me to do something, then I don’t do that thing freely. So God can act, and we can act, but God can act in ways that we can’t.

Something similar goes for knowledge: we human beings know a few things (maybe fewer than we ordinarily think), but there is much beyond our ken. Again, not so for God: given that he is all-knowing (‘omniscient’) as well as all-powerful, he knows everything, whatever can be known. Of course there are disputes in this area. Theists argue about whether God knows the future; they also argue about whether, even if he knows much about the future, he knows what free beings will in fact do. There is also dispute about whether God knows *counterfactuals of freedom*—propositions that specify what free creatures—you and I, for example, if in fact we are free creatures—would do in situations they will never be in. (Propositions like *If Mike were offered \$15,000 for his old van, he would [freely] sell it.*) These are really arguments about what can be known; the basic idea is that God knows whatever can be known, even if it isn’t clear, in every case, just what can be known. Still further, God is perfectly good. We human beings are a mixture of good and bad; there is evil in the best of us and good in the worst. Not so for God: there is no evil in him at all, and nothing bad about him. It is of course this combination of perfect goodness with omnipotence and omniscience that leads to the traditional problem of evil: it isn’t easy to see why there would be so much suffering and evil in a world created by an all-knowing, all-powerful, and perfectly good God.³

Finally, God has created the world—from the largest things it contains to the smallest. He has created all the stars and planets, all the galaxies and black holes, all the quarks and gluons and electrons (assuming that there really are such things). He has created all living things—plants and animals and human beings—either directly, or by employing other beings and processes. From the perspective of classical theists—Jewish, Muslim, and Christian theists—human beings are special. From this perspective, God has created human beings “in his

²Although it is extraordinarily hard to say in detail what goes on when we will to move an arm, and how it is that as a result of that willing, the arm moves.

³For interesting recent work on the problem of evil, see, e.g., *God and the Problem of Evil*, ed. William Rowe (Oxford: Blackwell, 2002), *The Evidential Argument from Evil*, ed. Daniel Howard-Snyder (Bloomington: Indiana Press, 1996), and *The Problem of Evil*, ed. Marilyn and Robert Adams (Oxford: Oxford University Press, 1990).

own image”—i.e., in such a way that in certain crucial respects they resemble him. Perhaps the central focus here is *personhood*: we human beings resemble God in being persons. Like God, we human beings have knowledge and affection; we too form intentions and are able to act on the basis of what we know in order to accomplish ends we seek. Of course there are enormous differences between human persons and God (a point some people tend to overlook, at least in their own case): he is an unlimited person and we are decidedly limited. Nonetheless the properties that make us persons—intellect, will, and affection, to use an old triumvirate—are ones we share with God.

God has created the world, but he also sustains it in existence; without this sustenance, the world would disappear like a candle flame in a high wind. Further, God governs the world in such a way that it displays a certain constancy and regularity. These regularities are everywhere: heavier-than-air objects dropped near the surface of the earth ordinarily fall down rather than up; bread is nourishing but mud is not; there is breathable air near the surface of the earth, though not at 35,000 feet or under water. Unlike rocks, seeds planted in soil sprout and take root; heavy steel beams will hold a lot of weight for a long time; a confined explosion will exert pressure on the walls of its container. It is by virtue of these regularities that human beings can act in the world, can learn about it, and act on what they have learned.

These regularities, of course, are what make science and technology possible. From the theistic point of view, the world God has designed and created is something like a vast machine, although that is perhaps too mechanical a term. (Perhaps it should also be thought of as something like a vast organism, or perhaps some amalgam of machine and organism.) In any event it is a structure of enormous complexity. (Think of the incredible complexity of a living cell, with its own hundreds of substructures in the form of molecular machines.) From a theistic point of view, one task of science is to come to know something about this wonderful structure—to learn about it in the systematic and communal way that is characteristic of science. Theism is thus, as such, not only hospitable to science, but enthusiastic about it. It is because God has created the world with these regularities and structures that it can be apprehended and known (to a significant degree) by creatures such as we are. It is because God has created us human beings in his image that we are able to apprehend and know the world.

A particularly interesting feature of the theistic view of the world, in this context, is that God created the various structures of the world *freely*. First, God wasn't obliged, by his nature or by some antecedent structure, or by anything else, to create anything at all. And given that he does create, he wasn't obliged to create just the things he did create. He has created horses, anacondas, and paramecia; he wasn't obliged to create any of them. And given that he creates the things—horses, for example—that we do in fact find, he

wasn't obliged to create them with just the properties they do in fact have. It's not a necessary truth that horses have the number of teeth they do have, or a stomach that works just the way an equine stomach does work. Further, given that he has created the creatures the world displays, he wasn't obliged to create them in any particular manner; he could have created them all specially, or, as presently seems more likely, by way of some evolutionary process. These things are all contingent; God could have done things differently. We ordinarily think that it is by *reason* that we know necessary truths; we know these things *a priori*, prior to or in some way independent of experience. Our knowledge of contingent truths, on the other hand, is (at least in part) by *experience*. Now the theistic idea is that what laws or regularities the world displays is a contingent matter; the same goes for the sorts of structures and organisms the world contains, and the properties of those structures and organisms. This suggests that science, as a systematic effort to come to knowledge of the world God has created, will have to be significantly *empirical*. From a theistic point of view, we can perhaps see this as the root of the empirical nature of science.⁴

There are stories about early opponents of modern science refusing to count the number of a horse's teeth or look through a telescope to see how many moons Jupiter has. These stories may or may not be true; nevertheless they illustrate a point. If you think you can figure out the number of teeth in a horse's mouth *a priori*, you won't feel obliged to open that horse's mouth and actually count them. If you think you know just by reasoning that Jupiter has no moons, you won't feel compelled to actually take a look through a telescope to see how many there are. (If the result of looking agrees with reason, the looking is unnecessary; if it doesn't, it is misleading.) On the other hand, if you think the world and its structures are contingent—contingent on God's freely choosing to make them one way as opposed to other possibilities—you'll think looking to see is the appropriate way to find out. In this way the empirical nature of science, as well as its basic charter, arise out of a theistic way of looking at the world and fit in well with it.

⁴This thought goes back to the beginning of modern science; thus John Brook and Geoffrey Cantor in *Reconstructing Nature: the Engagement of Science and Religion* (Edinburgh: T&T Clarke, 1998), p. 20:

Marin Mersenne, who was at the nerve-centre of one of the first scientific correspondence networks, objected to Aristotle's claim that the earth must be at the centre of the cosmos. For Mersenne there was no "must" about it. It was wrong to say that the centre was the earth's *natural* place. God had been free to put it where he liked. It was incumbent on us to find where this was.

See also Peter Dear, *Mersenne and the Learning of the Schools* (Ithaca, NY: Cornell University Press, 1988).

B. *Why do people believe theism?*

Most of the world's population endorse and accept some form of theism. Why do they do so? Why do they believe that there is such a person as God, that he is all-powerful, all-knowing, wholly good; and that he has created the world? How do they think they know these things? How do they know there is such a person as God; how do they know he is all-powerful, all-knowing, all-good; how do they know that he has created the world? Well, of course there are the famous theistic proofs, the classical arguments for the existence of God. For example, there are the traditional big three: the cosmological or first cause argument, going back to the ancient world and in particular Aristotle; the so-called ontological argument, first stated by Anselm of Canterbury in the eleventh century; and the argument from apparent design, sometimes also called the teleological argument. Although opinions vary widely as to their cogency, each of these arguments is of great interest and each is under intense contemporary discussion; each also has contemporary exponents.⁵ In addition to the traditional big three, there are a host of other theistic arguments—arguments from the nature of propositions, numbers, properties, from colors and flavors, from counterfactuals, and even a couple of arguments from the nature of evil.⁶ None of these arguments, nor even all of them taken together, I think, can sensibly be called a proof, if a proof is an argument such that it isn't possible to reject it without irrationality. Of course that's not saying much; there aren't arguments of that level of stringency for much of anything in philosophy.

But believers in God haven't traditionally relied upon proofs or arguments for their belief in God; most of the world's believers, I suppose, have barely heard of these theistic arguments. Why, then, *do* they believe? That question seems to presuppose that the natural or usual or expected way to believe in God would be on the basis of proof or argument. But why think a thing like that? Most of what we believe, we don't believe on the basis of proof or argument; so why think one can properly believe in God, or the essentials of the Christian faith, only on the basis of argument? On the other hand, if theists don't believe

⁵ For contemporary defense of the cosmological argument, see, e.g., William Craig, *The Cosmological Kaalam Argument* (London: Macmillan, 1979); for the ontological argument, see my *God, Freedom, and Evil* (Grand Rapids, MI: Eerdmans, 1977), pp. 85ff.; for the argument from design, particularly in its 'fine-tuning' version, see Robin Collins, "A Scientific Argument for the Existence of God: The Fine-Tuning Argument," in *Reason for the Hope Within*, ed. Michael Murray (Grand Rapids, MI: Eerdmans, 1999) and Robin Collins, "Evidence for Fine-Tuning," in *God and Design: the Teleological and Modern Science*, ed. Neil A. Manson (London: Routledge, 2003).

⁶ See my "Two Dozen (or So) Theistic Arguments" on the web at <http://philofreligion.homestead.com/files/Theisticarguments.html>.

in God on the basis of argument, what *is* the basis on which they believe? If there is no basis, wouldn't belief in God be just arbitrary?

To answer, we must ascend (or descend) into epistemology. Let's think briefly about our whole cognitive establishment, our whole set of cognitive faculties, the set of faculties or processes whereby we form beliefs, reject beliefs, and revise and change beliefs. From a natural and pre-philosophical position, these faculties seem designed to enable us to achieve true belief with respect to a wide variety of propositions—about our immediate environment (by perception), about our own interior life (introspection), about our past (memory), about the thoughts and experiences of other persons, about our universe at large, about right and wrong, about the whole realm of *abstracta* (properties, propositions, states of affairs, numbers, and the like), about modality (what's necessary and possible), and about God himself. These faculties work in such a way that under the appropriate circumstances we form the appropriate belief. More exactly, the appropriate belief is *formed in us*. In the typical case we do not *decide* to hold or form the belief in question, but simply find ourselves with it. Upon considering an instance of *modus ponens*, I find myself believing that it is a valid argument. I look into the backyard and have a certain visual experience; I find myself believing that the trees in the backyard are covered with snow. Upon being asked what I had for breakfast, I reflect for a moment and then find myself with the belief that what I had was eggs on toast. In these and other cases I do not *decide* what to believe; I don't total up the evidence (I'm being appeared to redly; on most occasions when thus appeared to I am in the presence of something red; so most probably in this case I am) and make a decision as to what seems best supported; I simply believe. In other sorts of cases I take a more active role in the formation of my beliefs; I look for evidence, or carefully sift and consider the evidence I have, or evaluate arguments, or consult people in the know. On the (or a) natural theistic view of the cognitive enterprise, God has created us with a complex, subtle, highly articulated establishment of faculties enabling us to achieve true beliefs on a wide variety of topics.

But how does belief in God fit into this picture? As follows.⁷ God created human beings originally with something like what John Calvin called a "Sensus Divinitatis"—a sense of divinity, a faculty, a set of cognitive processes whereby we come to know about God. The idea is that the *Sensus Divinitatis* is a faculty analogous, in some ways, to sense perception. By way of this faculty we human beings could know of the presence and properties of God. More important, by way of this faculty we could have the sort of relationship with God that

⁷Here what I'm presenting is a widely accepted Christian view of the epistemology of theistic belief; Jews and Muslims will think about the matter in a similar but somewhat different manner.

we have with other persons; there was to be communication and conversation, closeness, mutual love, and affection. Through the greatest catastrophe ever to befall the human race, however, we human beings somehow fell into sin, a ruinous condition in which we turned our backs upon God and rejected him. This condition is one all human beings share. Sin is a sort of madness of the will, a condition in which we love and hate the wrong things. Instead of loving God above all and our neighbor as ourselves, we tend to love ourselves above all, God and our neighbor coming in at best a very distant second. Indeed, according to the Heidelberg Catechism, we human beings are inclined to hate God and our neighbor; we resent the second as in competition with us, and the first as interfering with our own projects and autonomy.

The Christian story continues, however. God wasn't content to leave us in this appalling state. Instead, he proposed a scheme by which we human beings could be rescued from this sea of sin, and restored to our original condition of fellowship with God. The principal feature of this scheme is the incarnation, suffering, death, and resurrection of the divine Word, the second person of the Trinity, the Son of God. According to the Christian story, Jesus Christ, the second person of the Trinity, became a human being, took on our nature and our flesh. During his life and especially at its end he underwent enormous suffering, suffering of a depth and magnitude of which we have no understanding, suffering including not just crucifixion, but also that of feeling abandoned and forsaken by God the Father himself. He was crucified, and died (and rose from the dead); in this way he assumed the burden of human sin, redeeming us human beings from sin and its consequences. This salvation from sin is available to everyone; all that is required is to accept the proffered gift.

But, of course, God needed a way of informing people of every sort and condition, in all sorts of times and places, of the availability of the gift. No doubt he could have done this in many different ways. According to a classical Christian view I'll adopt here, God chose a three-part process. First, he arranged for the production of the Bible, a library of books or writings, each of which has a human author (or authors). These human authors, however, are inspired by God in a special way, a way in virtue of which God himself is the principal author of these books. The central theme and focus of the library is the gospel, the scheme of salvation God has offered. By virtue of this central theme and focus, the library is itself a book.

Second, there is the presence and activity of the Holy Spirit, the third member of Trinity, whose presence and action was promised by Christ himself before his death and resurrection, and invoked and celebrated in the epistles of the apostle Paul. On the classical Christian view, a person hears of God's scheme of salvation—in a sermon, from her parents, by reading the Bible, from a friend, or in some other way. The invitation of the Holy Spirit then enables her to see that the offer of salvation is a live option—not just for others, but also

for herself. If the person accepts the offer of salvation, the Holy Spirit produces *faith* (the third part of the three-part process) in the mind and heart of the believer—a deep conviction that the main lines of the Christian story are in fact true. This work of the Holy Spirit also repairs the ravages of sin, including the damage to the *Sensus Divinitatis*. This whole process may go on gradually, over a period of years, or it may happen suddenly, with a shattering impact. The resulting belief can be of maximal firmness; it can also be much more tentative and fragile. What is central to the process is this work of the Holy Spirit in producing faith, whereby Christians come to grasp and believe, endorse and rejoice in the main lines of the Christian gospel.

C. Theistic belief and knowledge

According to Mark Twain, “faith is believing what you know ain’t true”; but from a Christian point of view that’s a whopping mistake. On the present way of thinking, faith is instead a way of coming to *know* the main elements of Christian teaching.⁸ One hears the phrase ‘leap of faith,’ which suggests something like a leap in the dark. On the present way of thinking, however, Christian faith is not at all like a leap in the dark. It is not a matter of believing something on the basis of scanty evidence. Faith is not to be contrasted with knowledge; rather, if things go properly, it just *is* a certain kind of knowledge, and knowledge of truths of the greatest importance.

But how can faith be a form of knowledge? In order to understand this, we’ll have to descend even further into the depths of epistemology. The first question to ask, here, is this: what is the difference between knowledge and mere true belief? I believe the Detroit Tigers will win the pennant this year, and that despite the fact that they finished last in their division last year, and during the off-season they dealt away most of their best talent. Now suppose they do in fact win the pennant: would that show that I knew all along that they would? Certainly not: the fact is I didn’t know at all, but merely made a lucky guess. So the question is: what is the difference between knowledge and a lucky guess? What is it that distinguishes knowledge from mere true belief? Suppose we name that quantity, whatever it is, warrant. Warrant is the quantity enough of which distinguishes knowledge from true belief. But what exactly is warrant? That is one of the chief questions of epistemology.

Naturally enough there are several theories of warrant. The most successful, I think, crucially involves the idea of *proper function*, the notion of our cognitive faculties working properly, being subject to no dysfunction or malfunction.

⁸ See my *Warranted Christian Belief* (New York: Oxford University Press, 2000; hereafter ‘WCB’), pp. 256ff.

Any successful theory of warrant, I think, must involve reference to the proper function of our cognitive faculties. I don't have the space here to argue for this thesis properly;⁹ let me just give a couple of examples. First example: one very popular theory of warrant is process reliabilism, according to which a belief has warrant for me if it is produced by cognitive faculties or processes that are reliable.¹⁰ We needn't spend a lot of time, here, trying to say just what it is for a faculty or process to be reliable; the basic idea is that a reliable process produces a suitable preponderance of true beliefs over false beliefs. Of course that's not sufficient; a thermometer stuck on 75° F in a place where it is always 75° F (San Diego?) isn't reliable, even though it always correctly registers the current temperature. What it *would* register if the temperature were different is also relevant. To be reliable, a mechanism like a thermometer must produce mainly true readings, not only in the actual world, but in other nearby possible worlds as well. The same goes for a cognitive process like my vision: it is reliable only if it produces mainly true beliefs, and would have done so even if I had been in different surroundings looking at different things. But being produced by a reliable cognitive process isn't enough for warrant. Suppose my vision is in fact a reliable process, but I've been drinking; I'm now seeing double. Maybe I form the belief that there are two policemen looking disapprovingly at me; that belief doesn't have much warrant for me, even though it is produced by a reliable belief-producing process. So process reliabilism won't be satisfactory without a reference to proper function.

Second example: another theory of warrant is coherentism, according to which a belief has warrant for me if it is coherent with the total set of my beliefs.¹¹ This, too, however, won't work unless it is supplemented by a clause specifying proper function. We can see this by considering the Case of the Epistemically Inflexible Climber. Ric is climbing Guide's Wall, on Storm Point in the Grand Tetons; having just led the classic next to last pitch, he is seated on a comfortable ledge, bringing his partner up. He believes that Cascade Canyon is down to his left, that the cliffs of Mt. Owen are directly in front of him, that there is a hawk gliding in lazy circles 200 feet below him, that he is wearing his new Mythos rock shoes, and so on. His beliefs, we may stipulate, are coherent. Now add that Ric is struck by a wayward burst of high-energy cosmic radiation. This induces a cognitive malfunction; his beliefs become fixed,

⁹ For a fuller account, see my book *Warrant: the Current Debate*, hereafter WCD (New York: Oxford University Press, 1993).

¹⁰ See Alvin Goldman, "What Is Justified Belief?," in *Justification and Knowledge: New Studies in Epistemology*, ed. George Pappas (Dordrecht: D. Reidel, 1979), p. 10.

¹¹ See Laurence Bonjour, *The Structure of Knowledge* (Cambridge, MA: Harvard University Press, 1985). (Bonjour apparently no longer accepts coherentism.)

no longer responsive to changes in experience. No matter what his experience, his beliefs remain the same. At the cost of considerable effort his partner gets him down and, in a desperate last-ditch attempt at therapy, takes him to the opera in nearby Jackson, where the New York Metropolitan Opera on tour is performing *La Traviata*. Ric is appeared to in the same way as everyone else there; he is inundated by wave after wave of golden sound. Sadly enough, the effort at therapy fails: Ric's beliefs remain fixed and wholly unresponsive to his experience; he still believes that he is on the belay ledge at the top of the next to last pitch of Guide's Wall, that Cascade Canyon is down to his left, that there is a hawk sailing in lazy circles 200 feet below him, that he is wearing his new Mythos rock shoes, and so on. Furthermore, since he believes the very same things he believed when seated on the ledge, his beliefs are coherent. But surely they have little or no warrant for him. The reason is cognitive malfunction; his beliefs are not appropriately responsive to his experience. Coherentism, too, then, requires a reference to proper function.

The correct account of warrant, I believe, is as follows.¹² First, a belief is warranted only if it is produced by cognitive faculties that are functioning properly, working the way they are supposed to work, subject to no dysfunction or malfunction. But, second, they must be in an appropriate cognitive environment. Imagine that you take a space voyage to a planet circling a nearby star; as it turns out, a subtle gas is pervading the atmosphere there—one that causes human beings to believe there are elephants present, even when there aren't any within miles. You form the belief that there is an elephant present; this belief has no warrant for you, even though your cognitive faculties are functioning perfectly properly. So we must add that the cognitive environment is appropriate for the faculties in question; from a theistic point of view, that would be an environment similar to the one for which God created our faculties.

But these two aren't enough. To see this, note that people stricken by serious disease often estimate their chances of recovery as greater than they actually are; we may call this process 'the optimistic overrider.' So suppose I fall victim to a life-threatening disease and form the belief that my chances of recovery are excellent. Would this be a matter of cognitive malfunction? Probably not. The fact is someone is more likely to recover if he adopts an optimistic attitude and thinks he'll recover; the cognitive process producing the belief in question, we may suppose, is directed towards recovery and survival. So the belief in question is produced by properly functioning epistemic processes. The purpose or function of this process, however, isn't to produce true belief, but belief with a different virtue: conductivity to survival. But then it doesn't have warrant.

¹²For a fuller account, see my book *Warrant and Proper Function* (New York: Oxford University Press, 1993; hereafter 'WPF'), ch. 2 and *passim*.

So we must add, as a third condition, that the purpose or function of the belief-producing processes or faculties is the production of true belief.

There is still another condition we must add. For suppose our cognitive faculties were designed by a junior deity, one with very little competence or experience.¹³ His design is unsuccessful. His heart is in the right place and he intends that they produce mainly true beliefs, but (due to his incompetence) when these faculties work properly, i.e., the way he designed them to work, they produce mainly false beliefs. So suppose I form a certain belief *B*, and suppose the conditions so far mentioned are all met: this belief is formed by faculties functioning properly according to a design plan aimed at truth in the right kind of environment. But *B* still has little or no warrant; the problem is that the design plan isn't *successfully* aimed at truth. We must therefore add a clause to that effect. The way to put it, then, is that a belief *B* has warrant for a person *S* if and only if *B* is produced by properly functioning faculties in an appropriate environment according to a design plan successfully aimed at truth.¹⁴

Now we can return to the question that precipitated this excursus into the nature of warrant: how can it be that faith—belief in the great things of the gospel, as Jonathan Edwards called them—is *knowledge*? The answer is simplicity itself. According to the above story (p. 9), faith, belief in the central features of the Christian message, is produced in the believer by the activity of the Holy Spirit. This is a belief-producing process; it is not part of our natural and original epistemic endowment, but it is still a belief-producing process. When it operates, clearly enough, it functions properly (works the way it was designed to work) in an appropriate epistemic environment (the one for which it was designed) according to a design plan successfully aimed at truth. It therefore meets the central conditions for warrant; if it is held with sufficient firmness (and assuming that it is true) it will constitute knowledge.¹⁵

So faith is really a special case of knowledge. But isn't this assuming that the Christian story is in fact true? Those who don't think that story true will not be inclined to think that it constitutes knowledge! That is certainly true, but not presently relevant. We have been looking into the question why Christians believe as they do, and how they think they know these things. The above account is one classical Christian answer to that question. This account presupposes the truth of Christian belief, but of course Christians think Christian belief *is* true. But isn't it somehow objectionably circular to answer the

¹³ See David Hume's *Dialogues Concerning Natural Religion*, ed. Richard Popkin (Indianapolis: Hackett Publishing Co., 1980), p. 37.

¹⁴ Actually, these conditions are not quite sufficient; for nuance and qualification and an additional condition, see WPF, ch. 2, and WCB, pp. 156ff.

¹⁵ See WCB, pp. 256–8.

question ‘how do you know p ?’ in a way that assumes that p is in fact true? No. An epistemologist might try to give an account of perceptual knowledge: how is it that our vision, for example, provides us with knowledge of our immediate surroundings? Here the account may proceed in terms of light from objects striking the retina and activating photo-receptor cells, signals being propagated along the optic nerve to the brain, neural activity in the vision centers of the brain, and all the rest. This account, however, will clearly presuppose that perceptual beliefs are in fact true; the very idea that there *is* such a thing as the retina presupposes that certain perceptually obtained beliefs are true. But there is nothing objectionably circular in this process. Similarly, there is nothing objectionably circular in an epistemological account of Christian belief that presupposes the truth of such belief. Of course there would be something viciously circular in using such an account to argue for the *truth* of Christian belief.

Finally, by way of concluding this account of the epistemology of Christian belief, it is worth noting that if Christian belief is true, then very likely it does have warrant—if not in the way proposed here, then in some other similar way. For if it is true, then indeed there is such a person as God, who has created us in his image; we have fallen into sin and require salvation; and the means to such restoral and renewal have been provided in the incarnation, suffering, death, and resurrection of Jesus Christ, the second person of the Trinity. Furthermore, the typical way of appropriating this restoral is by way of faith, which of course involves belief in these things, i.e., belief in the great things of the gospel. If so, however, God would of course intend that we be able to be aware of these truths. And if *that* is so, the natural thing to think is that the cognitive processes that do indeed produce belief in the central elements of the Christian faith are aimed by their designer at producing that belief. But then these beliefs will be produced by cognitive processes functioning properly according to a design plan successfully aimed at truth; they will therefore have warrant.

Well, you say, OK, if Christian theism is true, it is probably also warranted. But so what? Isn't that a pretty piddling conclusion? Why should that be thought important? Two reasons. First, a common objection to Christian belief (at least since the Enlightenment) has been not to its *truth*, but to its *rationality* or *reasonableness*. The objection goes something like this: “I don't know whether Christian belief is *true* (who could know a thing like that?), but I do know that it is irrational, or unreasonable, or such that a sensible person couldn't accept it.” But what is this ‘irrationality’ or ‘unreasonability’ the objector speaks of? In *Warranted Christian Belief* I argued that the only plausible way to construe this objection is in terms of warrant: the objector must be arguing that Christian belief is not in fact *warranted* for those who hold it. This objection, furthermore, is supposed to hold whether or not Christian belief is *true*; the idea is that even if it happens, somehow to be true, it still doesn't have warrant for those who accept it. But that can't be right if, as I claim, Christian belief is true,

if it is warranted. So sensible objections to Christian belief will really have to be to its truth, not to its rationality or sensibleness or warrant; the whole class of *de jure* objections—objections that are supposed to be independent of its truth or falsehood—can't sensibly be made.

But isn't the same true for nearly everything? Isn't most any belief such that if it is true, then (very likely) it is warranted? Certainly not; there are plenty of beliefs for which this isn't true. On page 9 I referred to my belief that the Tigers will win the pennant; that is certainly not such that if it is true, then it is probably warranted. It's a lucky guess: even if it does turn out to be true, I was right, so to speak, just by accident; that means that I don't know that the Tigers will win, and hence the belief does not have warrant for me. The same goes, I'd say, for beliefs formed in the process of inquiry or investigations which are at or near the limit of our powers, as in parts of contemporary physics. Perhaps current theories about what happened during the first 10^{-32} seconds after the Big Bang are correct (and then again perhaps not); but even if they are correct, these theories don't have much warrant for us. That is because in thinking about these things, we are close to the limits of our cognitive powers; it is then very easy for us to make mistakes. More to the present point, this isn't true for naturalism either; it isn't true that if naturalism is true, then very likely it has warrant for us. In fact in section B I'll argue that if naturalism is true, then it *doesn't* have warrant for those who believe it. But that means (for someone who sees the soundness of the argument) that naturalism can't rationally be believed.

II Alternatives to Theism

My focus is on naturalism as a main alternative to theism. We should first note, however, that it isn't the only main alternative. There are at present, and in the Western world, fundamentally three worldviews vying for supremacy: three fundamental perspectives or ways of thinking about what the world is like, what we ourselves are like, what is most important about the world, what our place in it is, and what we must do to live the good life. The first of these perspectives is theism; I've already said a bit about that. In addition to that perspective, however, there are fundamentally two others. Both of them have been with us since the ancient world; but each has received much more powerful expression in modern times. According to the first perspective, *philosophical naturalism*, there is no God, and we human beings are insignificant parts of a giant cosmic machine that proceeds in majestic indifference to us, our hopes and aspirations, our needs and desires, and our sense of right and wrong. This picture goes back to Epicurus, Democritus, and others in the Ancient world and finds magnificent expression in Lucretius' poem, *De Rerum Natura*; it is also extremely popular in the contemporary (Western) world.

According to the second perspective, it is we ourselves—we human beings—who are somehow responsible for the basic structure of the world. *We* somehow bring it about that the world has the structure and nature it displays; it is we who are somehow responsible for the truth of those propositions that *are* true. Call this *creative anti-realism*. This notion, too, goes back to the ancient world, in particular to Protagoras, with his claim that man is the measure of all things, but it has been developed with much greater power and detail in the modern world. Creative anti-realism begins (in the modern world) with the publication of Immanuel Kant's *Critique of Pure Reason*; it is especially popular in continental Europe. Kant's basic idea, at least in that book, is that, in some deep and important way, we human beings confer upon the world its fundamental structure—its spatio-temporal structure, its thing-property structure, and so on. We do this, he says, by our conceptual activity. Exactly how this is supposed to go, according to Kant, is both obscure, and (like nearly everything else about Kant's thought) a subject of great controversy. This much is clear, however: as Kant thought of it, we all do this world formation or structuring *together*, and we all live in the *same* world.

But if you follow Kant in thinking our world is in some mysterious way created or structured by human beings, you may note that human beings do not all seem to live in the *same* world. The world of Jerry Falwell seems quite different from that of Richard Dawkins; which one, then (if either), corresponds to the world as it really is? Here it is an easy step to a characteristic thought of contemporary forms of Creative Anti-Realism: the thought that there simply *isn't* any such thing as an objective way the world is, a way the world is that is the same for all of us. Rather, there is my version of reality, the way I've somehow structured things, and your version, the way *you've* structured things; and many other versions. As Marlowe's Dr. Faustus in effect says, "Man is the measure of all things; I am a man; therefore I am the measure of all things." Call this way of thinking *postmodern anti-realism*.

Now the existence of *truth* is intimately connected with there being a way things really are, a way the world really is. For it is *true* that there are horses (for example) if and only if there being horses is part of the way things are. What lies at the heart of postmodern anti-realism (and is responsible for some of its astonishing excesses) is the idea that there really *isn't* any such thing as the way the world is, and therefore no such thing as truth. That is, there isn't any such thing as truth as we ordinarily think of it. Usually something else is proposed as a replacement for truth—typically, something that somehow depends upon what we (we humans, or our society, or the scientists of our culture circle, or the speakers of our language) do or say or think. For example, there is Richard Rorty's idea that truth is "what our peers will let us get away with saying." This suggestion seems initially unpromising; after all, *your* peers may let *you* get away with saying something my peers won't let *me* get away with saying: are we to

suppose that some things are true for you but not for me? Furthermore, if that's what truth is, wouldn't there be a much easier way of dealing with all the ills flesh is heir to, for example cancer? If we all let each other get away with saying that there just isn't any such thing as cancer, or AIDS, then, on this Rortyesque view, it would be *true* that there isn't any such thing as cancer; and if it were *true* that there isn't any such thing as cancer, then there wouldn't *be* any such thing. So all we have to do to get rid of cancer, or poverty, or war, or other nasty things is to let each other get away with saying there aren't any such things. That seems much easier than the more conventional methods, with their substantial cost in time, energy, and money. Second, if you have done something bad, lie about it—try to get your peers to let you get away with saying that you didn't do it. If you succeed, then it will be true that you haven't done it, in which case you won't have done it. Indeed, as an added bonus, you won't even have lied about it! On the face of it, then, this way of thinking doesn't seem at all plausible. Of course I don't mean to suggest that there isn't a serious and sensible view somewhere in the neighborhood; but (as it stands) this isn't it.

Perhaps you will object that I'm just belaboring a straw man: Rorty couldn't really mean that truth is what our peers will let us get away with saying. You may be right. What Rorty actually says is:

For philosophers like Chisholm and Bergmann, such explanations *must* be attempted if the realism of common sense is to be preserved. The aim of all such explanations is to make truth something more than what Dewey called "warranted assertability": more than what our peers will, *ceteris paribus*, let us get away with saying.¹⁶

(It is clear from the context here (and elsewhere) that Rorty sides with Dewey against Chisholm and Bergmann). Perhaps you will say that this is just a rough and ready conversational version of his real opinion. Rorty scorns the usual analytic philosopher's necessary and sufficient conditions, principles, analyses, attempts at rigor, and argumentation (maybe taking to these the same playful attitude Derrida takes towards that obsessive concern with quotation marks he ascribes to Oxford philosophers¹⁷). Philosophy, he thinks, should be conversational; and this is his conversational way of putting his point; but then, of course, it's not fair to hold him to the letter of what he says. Well, perhaps so. A person certainly has a right to write conversationally even on such an austere subject as philosophy; it's a free country. Still, this does complicate matters.

¹⁶ *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979), pp. 175–6.

¹⁷ *The Post Card: From Socrates to Freud and Beyond*, tr. Alan Bass (Chicago: University of Chicago Press, 1987), p. 98.

If we want to take Rorty's alleged anti-realism seriously, we need a relatively clear and serious way of stating what this view amounts to; that's just what we don't have.

III Naturalism and Its Woes

There is a great deal more to be said about postmodern anti-realism, but this isn't the place to say it;¹⁸ our focus is the other of the two main alternatives to theism, namely naturalism. As you may have noticed, naturalism is all the rage these days; naturalist philosophers spend a great deal of time and energy trying to work out and develop naturalistic accounts of the sorts of problems and topics philosophers ordinarily work on. Thus, for example, they try to develop naturalistic accounts of epistemology; the idea is to develop an epistemology that is purged of any of the elements a proper naturalist would regard with suspicion. They also try to develop naturalistic accounts of personhood, of what it is to *believe* something, of consciousness, of language and meaning, of mathematics, of abstract objects such as universals, of ethics and morality, of religion, and much more. In each case, the idea is to develop an adequate account of the area in question while appealing only to entities—for example, concrete objects and perhaps sets of them—that won't bring a blush to the cheek of even the tenderest naturalist.

But what sorts of entities *are* those—what *is* it for an account to be properly naturalistic? To answer that question, we must first ask another: what is naturalism? Of course the word is used in many different ways. There is naturalism in art and literature, for example, which may have little to do with naturalism in philosophy. There is methodological naturalism in science. In philosophy, there is the sort of naturalism in ethics G. E. Moore objected to when he spoke of the naturalistic fallacy; there is the naturalism of John Dewey and Willard van Orman Quine. What is the basic idea of naturalism, the core notion in terms of which all these others can be understood, perhaps as analogically related to it? This is by no means an easy question; naturalism is not at all easy to characterize.¹⁹ (In this regard it is a little like pornography: as Justice Potter Stewart said, maybe you can't say what it is, but you can tell it when you see it.) Indeed, some who think about naturalism believe that it isn't a *doctrine* at all; it isn't a belief,

¹⁸ For a fuller characterization and criticism of postmodern anti-realism, see chapter 13 of WCB; see also my "The Twin Pillars of Christian Scholarship," in *Seeking Understanding: the Stob Lectures 1986–1998*, ed. Calvin College (Grand Rapids, MI: William B. Eerdmans, 2001), pp. 128–32.

¹⁹ Chapters 2 and 3 of Michael Rea's *World without Design* (New York: Oxford University Press, 2002) contain an excellent discussion of some of the ways of characterizing or defining naturalism.

or a proposition. According to Bas van Fraassen, for example, to be a naturalist is not to *believe* anything special—e.g., that there aren't any fairies, or angels, or gods; to be a naturalist is to adopt a certain *attitude*, an attitude involving among other things an exclusive commitment to science in guiding one's opinions.²⁰ And according to Mike Rea (*World without Design*) naturalism is really, at bottom, a *research program*, a way of conducting inquiry; and what it centrally involves is a commitment or determination, in conducting inquiry, to use only the methods and techniques employed in the empirical sciences.²¹

Without taking a position on this question of the essence or basic idea of naturalism, I'd like to think of it, for present purposes, as fundamentally a way of looking at the world, a high-level belief about the world. There certainly *is* this way of looking at the world, even if naturalism itself is really an attitude or research program; I'll use the term 'philosophical naturalism' to refer to this way of thinking. Here is Bertrand Russell's famous statement of it:

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; that no fire, no heroism, no intensity of thought and feeling, can preserve an individual life beyond the grave, that all the labors of the ages, all the devotion, all the inspiration, all the noonday brightness of human genius, are destined to extinction in the vast death of the solar system, and that the whole temple of man's achievement must inevitably be buried beneath the debris of a universe in ruins—all of these things, if not quite beyond dispute, are yet so nearly certain that no philosophy which rejects them can hope to stand. Only within the scaffolding of these truths, only on the firm foundation of unyielding despair, can the soul's habitation henceforth be safely built.²²

This may sound a bit florid and overheated, but it does serve to give the flavor of the view I mean to attack.²³

Following Quentin Smith, we could characterize the perspective Russell expresses as "the thesis that there exist inanimate or animate bodies, with animate bodies being either intelligent organisms or non-intelligent organisms, but there exists nothing supernatural."²⁴ Of course, as it stands that definition isn't terribly informative; it contains the word 'supernatural,' which presumably

²⁰ *The Empirical Stance* (New Haven: Yale University Press, 2002), pp. 49ff.

²¹ *World without Design* (New York: Oxford University Press, 2002).

²² *Mysticism and Logic* (New York: Barnes & Noble, 1917), pp. 47–8.

²³ A comprehensive and enthusiastic contemporary book-length exposition of naturalism is Daniel Dennett's *Darwin's Dangerous Idea* (New York: Simon and Schuster, 1995). For critical animadversions on this book, see my "Dennett's Dangerous Idea: Darwin, Mind and Meaning," *Books and Culture*, May–June 1996, pp. 16–18, 35.

²⁴ "The Metaphilosophy of Naturalism," *Philo* 4:2 (fall–winter 2001), p. 202.

needs just as much by way of definition as does ‘naturalism’ itself. Perhaps the best way to get at naturalism, taken as a philosophical doctrine, is to contrast it with theism. I outlined the theistic perspective above: there is God, with his special and unique properties, and then there is the world he has created. The basic idea of philosophical naturalism (which from now on I’ll just call ‘naturalism’) is that there is no such person as God, or anything at all like him. So first, a naturalist (as I’m using the term) will be an atheist. But not every atheist is a naturalist. Naturalism is *stronger* than atheism, in the sense that it is possible to be an atheist but not a naturalist, but not possible to be a naturalist but not an atheist. After proposing the above account of naturalism, Quentin Smith goes on:

The example of something supernatural of most interest to contemporary analytic philosophers is an unembodied mind that is the original and/or continuous creator of the universe, and has the omniattributes described in perfect being theology. Other examples of hypothesized supernatural realities that govern or create in some sense the universe are the governing mind posited by the Stoics or the “Absolute I” posited by the early Fichte.

So naturalism includes atheism, and more. If you are a naturalist, you don’t believe in God, but you also don’t believe in the Stoics’ *Mind*, or Fichte’s Absolute I, or Plato’s Idea of the Good, or Aristotle’s Unmoved Mover, or Hegel’s Absolute. This account of naturalism suffers a certain vagueness (nothing at all *similar to* God, but just *how* similar?), but in practice I doubt that there is much of a problem here.

So much for a characterization of naturalism and for the contrast between theism and naturalism. What I want to argue next is that naturalism is an unacceptable belief. As I said above, naturalism should be rejected, and for at least three different reasons. First, naturalism cannot accommodate the idea of *proper function*, for such organisms as plants and animals and human beings. It therefore cannot accommodate the notions of health, sanity, sickness, disease, and the like. Further, as I argued above, warrant, the quality or quantity that makes true belief into knowledge, essentially involves proper function. This means, then, that if naturalism were true, there would be no such thing as knowledge, as well as no such thing as health, sanity, illness, or any other condition that entails these. Second, and more devastating, naturalism leads directly to Humean skepticism, the condition in which you have a defeater for whatever you believe and cannot sensibly trust your cognitive faculties. In this connection I’ll also argue that naturalism is self-defeating, in that if it is true, it is irrational to believe it. Third, and perhaps most devastating, naturalism cannot accommodate *belief*; if naturalism is true, no one believes anything.

A. *Naturalism vs. proper function*

First, then, if naturalism were true (so I shall argue), neither human beings nor their component organs and systems would function properly (or, for that matter, improperly). Fundamentally, *that* is because the notion of proper function really applies only to things that have been designed by conscious, purposeful intelligent agents; the basic notion of proper function is that of working in a way the designer(s) intended. Of course this requires modification and nuance. My refrigerator was designed to keep things cool; it starts malfunctioning, its interior temperature a constant 150° F. I give it to you, and you use it for a warming oven. Is it malfunctioning or not? My grill rusts out and can no longer be used for the purpose for which it was designed (i.e., grilling); my grandchildren paint attractive designs on it, and it is now a very nice planter: is it malfunctioning? In these and other cases qualification and nuance are required.²⁵ But the basic idea is still that proper function requires intelligent design.

It is this that gives trouble for the naturalist bent on explaining the notion of proper function in naturalistic terms. Proper function requires design; but the only plausible designer for us human beings and our systems and organs would be God, or something very much like God. (Conceivably God himself didn't design human beings, but delegated the task to a high-ranking angel.) Of course a naturalist might maintain that we have been designed and brought into existence by extraterrestrial beings of great intellectual accomplishments. Perhaps these extraterrestrials brought us into being by taking a hand in the course of terrestrial evolution, causing the right mutations to arise at the right times, adjusting the environment so that the right organisms survived, and the like. This is a bit farfetched, perhaps, but not clearly impossible. But it won't help the naturalist. For the same sorts of questions will arise about those talented extraterrestrials: presumably the notion of proper function will apply to *them*, but *they* weren't designed. (Or, if they were, the question will arise with respect to their designers, or to the designers of their designers, . . .)

Note that this is not a problem for the theist. She believes that human beings have been designed and created by God, and created in the image of God. As she sees it, God could have created us human beings in many different ways. Perhaps he directly and immediately created a first human pair or group of humans; or perhaps he orchestrated the course of evolution so that we came to be; or perhaps he directly modified an earlier primate form of life in such a way that the result was us human beings; or perhaps our bodies have evolved, but, as is part of Catholic doctrine, God directly creates a new human soul or self every time a human being comes into existence; or perhaps . . . Clearly there are many different possibilities here. What they all have in common is

²⁵ See WPF, pp. 21–31.

that God designs us human beings, so that it makes sense to say that we and our systems and organs can function in the way God designed us to function. When a system or organ functions the way God intended, then it functions properly; when it functions in a way incompatible with the way God intended, then it malfunctions.

I say the naturalist can't accommodate the notion of proper function; but you may not be inclined simply to take my word for that. Can't a naturalist, just as well as anyone else, see that a bird's wing is damaged and incapable of proper function? Can't a naturalist, just as much as anyone else, see that someone who thinks he's Jesus Christ (and isn't) is suffering from a cognitive disorder? Well yes, of course. The point is not that naturalists can't sometimes see that a part of an organism isn't functioning properly; the point is that the naturalist can't give an account of proper function that is compatible with his naturalism. Naturalism can't accommodate proper function. Let's look into this a bit further; just what is this 'accommodating'? One way to argue that naturalism *can* accommodate proper function would be to give an *analysis* of proper function in terms of properties that are naturalistically acceptable. What properties are *those*? Naturalistically acceptable properties are properties that could be instantiated, (even) if naturalism is true. Such properties as *weighing 200 lbs*, *living in Boston*, and *liking strawberries* are naturalistically acceptable; such properties as *being designed by God*, or *created by God*, or *approved or commanded by God* are not naturalistically acceptable. That is because they could be instantiated only if there is such a person as God; and according to naturalism there is no such person.

So one way to argue that proper function can be accommodated by naturalism is to give an analysis of proper function in terms of properties that are naturalistically acceptable. To give an analysis of a concept or property (or relation), furthermore, is at least to give necessary and sufficient conditions for it.²⁶ Suppose *P* is the property in question: to give an analysis of *P* is first of all to suggest some other (possibly complex) property *Q*, such that it is necessary in the broadly logical sense that a thing has that property *P* if it has the property *Q*, and necessary that it has *P* only if it has *Q*. That is, the *analyzans* (the analysis) must be necessary and sufficient for the *analyzandum* (the property or relation to be analyzed). Consider, for example, the traditional analysis of knowledge as justified true belief:

A person *S* knows a proposition *P* if and only if *S* believes *P*, *P* is true, and *S* is justified in believing *P*.

²⁶ Actually more than this is required, but the more isn't relevant to our present concerns.

This analysis may or, more likely, may not²⁷ be correct; the point is that it proposes necessary and sufficient conditions for the property *knowing P*. It also gives necessary and sufficient conditions for the relation that holds between a person *S* and a proposition *P* when the former knows the latter. The idea is that if a person knows *P*, then she must believe *P*, *P* must be true, and she must be justified in believing *P*; conversely, if she believes *P*, *P* is true, and she is justified in believing *P*, then she must know *P*. If the analysis is successful, it isn't so much as possible that the *analyzandum* hold but the *analyzans* fail to hold; and it is equally impossible that the *analyzans* hold but the *analyzandum* fail to hold.

Returning to our present concern, proper function can be accommodated by naturalism if and only if there are necessary and sufficient conditions for proper function in terms of naturalistically acceptable properties—properties that could be instantiated even if naturalism were true. Often, however, when naturalists talk about proper function they give an account not of proper function itself, but of some other notion in the neighborhood, one that perhaps, as they think, can nicely *replace* our notion of proper function. (Perhaps they do this partly because it is clear to them that one can't give a naturalistic account of proper function itself.) Most of these accounts—both those of proper function itself, and those of other nearby notions—invoke evolution, in particular natural selection.

Karen Neander, for example, gives the following account of 'proper function':

It is the proper function of an item *X* of an organism *O* to do that which items of *X*'s type did to contribute to the inclusive fitness of *O*'s ancestors, and which caused the genotype, of which *X* is the phenotypic expression, to be selected by natural selection.²⁸

So consider a heart: it is the proper function of your heart to do what previous hearts did to contribute to the inclusive fitness of your ancestors: i.e., circulate your blood (and circulate it in a certain way); and when it does that, it is functioning properly. It is not part of its function to make that thumpa-thumpa sound; your ancestors' hearts presumably made that sound, but that was just a byproduct of their function and did not itself contribute to your ancestors' fitness.²⁹ Now, despite her use of the term 'proper function,' Neander doesn't propose her

²⁷ See my *Warrant: the Current Debate* (New York: Oxford University Press, 1993; hereafter 'WCD'), chs 1–3.

²⁸ "Functions as Selected Effects: the Conceptual Analyst's Defense", *Philosophy of Science* 58 (1991): 174.

²⁹ Well, some people speculate that this sound *does* contribute to fitness: human infants may be quieted or comforted upon being held by mothers (or fathers) whose hearts make that sound. If this is a problem for you, move to a different example of the distinction between function and byproduct.

account as an analysis of the notion of proper function, i.e., the notion we all have and use in ordinary life. What she says is that the concept *she* is analyzing is a *scientific* concept, and one that may significantly differ from the ordinary one. This is an important point, because it is the everyday ordinary concept of proper function that is involved in the notions of warrant, sanity, health, and the like—not some *other* concept in the neighborhood, no matter how scientifically respectable that other concept may be. It is *proper function* (not some other concept) that is entailed by the notion of warrant; and it is proper function, I say, that can't be given a naturalistically acceptable analysis. The fact that we can construct *other* concepts out of naturalistically acceptable elements is perhaps of interest in some contexts, but doesn't cut any ice in *this* context.

So Neander isn't offering an analysis of the ordinary everyday concept of proper function. Still, it could serendipitously happen that the analysis she proposes really is a good analysis of that everyday concept, even if she doesn't intend it as such. But it isn't. It is instructive to see why not, because we can thereby see that no analysis involving natural selection or evolution can work as an account of the ordinary notion of proper function. As we recall, where *proper function* is the *analyzandum*, the *analyzans* must be a property *Q* such that necessarily, whatever has the property of proper function also has *Q*, and, necessarily, whatever has *Q* also has the property of proper function. If the analysis is successful, it won't be possible (in the broadly logical sense) that there be an object that falls under the *analyzandum* but not the *analyzans*, or falls under the latter but not the former.

This condition isn't met by any analysis of proper function that invokes the notion of evolution, i.e., any analysis where the *analyzans* includes the property of having been produced by some process of evolution. Any such analysis will be too strong; it isn't necessary that all biological organs or systems capable of proper function be produced by such a process. That is because even if it is a truth, it is not a *necessary* truth that organisms have come to be by way of evolution. Evolution is a dandy idea (Daniel Dennett exuberantly declares it the best idea human beings have come up with³⁰); the idea that we come to be by such a process has currently achieved the status of orthodoxy;³¹ still, the proposition that we have come to be by some such process is at most *contingently* true. It is *possible*, in the broadly logical sense, that the view is flatly false. It is possible, for example, that each of the main forms of life was created by

³⁰ *Darwin's Dangerous Idea*, p. 21.

³¹ According to the 1979 edition of the *New Encyclopedia Britannica*, "evolution is accepted by all biologists and natural selection is recognized as its cause. . . . Objections . . . have come from theological and, for a time, from political standpoints" (Vol. 7, article on Evolution).

God (or by some other powerful and knowledgeable being) *ex nihilo*, or by instantaneous modification of previous life forms, or in some other way incompatible with mechanisms proposed by contemporary evolutionary theory. If that is possible, however, no correct account of proper function can *presuppose* the truth of contemporary evolutionary theory. In particular, then, the account Neander presents, while it may serve *other* purposes, won't serve as an analysis of proper function. For clearly it is logically possible that a thing *X*—a heart, for example—display a proper function even if it is not the case that it is doing “that which items of *X*'s type did to contribute to the inclusive fitness of *O*'s ancestors”; indeed, it is logically possible that *O* doesn't even *have* any ancestors. Whether or not God directly and immediately created Adam and Eve, clearly he *could* have—and if he had, they would have had no ancestors. Still, their hearts would have had proper functions: the very functions performed by yours and mine.

In my book *Warrant and Proper Function* I examined a number of naturalistic accounts of proper function, arguing in each case that they fail. I won't repeat what I said there; what I'll do instead is briefly mention the most important and widely cited kinds of accounts and explain how they fail; then I'll look more closely at a couple of accounts I didn't examine in *Warrant and Proper Function*. One of the most widely discussed and endorsed accounts of proper function is that offered by Ruth Millikan in her book *Language, Thought, and Other Biological Categories*.³² That account is difficult and complex and hard to get really clear about;³³ fortunately she has since offered a simplified version:

Putting things very roughly, for an item *A* to have function *F* as a “proper function” it is necessary (and close to sufficient) that one of these two conditions should hold. (1) *A* originated as a “reproduction” (to give one example, as a copy, or a copy of a copy) of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed *F* in the past, and *A* exists because (causally historically because) of this or these performances. (2) *A* originated as the product of some prior device that, given its circumstances, had performance of *F* as a proper function and that, under these circumstances normally causes *F* to be performed by means of producing an item like *A*. Items that fall under condition (2) have “derived proper functions”, functions derived from the devices that produce them.³⁴

So consider, once more, a human heart; it would presumably fall under condition (1) above. Your heart has the function of pumping blood because it

³² Cambridge, MA: MIT Press, 1984, p. 17.

³³ See WPF, pp. 201–2.

³⁴ “In Defense of Proper Functions,” *Philosophy of Science* 56 (1989): 288–302.

originated as a reproduction or copy of a prior heart which also pumped blood, and your heart exists because that prior item performed that function.

Like Neander, Millikan proposes her account as a “theoretical definition” of a “technical term,” not as an analysis of proper function itself; and far be it from me to say otherwise.³⁵ Again, however, it is of interest to see whether her account does in fact provide the materials for an adequate or accurate analysis of our ordinary concept or understanding of proper function. And clearly it doesn’t. First, there is the same problem as with Neander: Millikan’s account entails that anything that functions properly has ancestors. Now even if it were in fact true that everything that functions properly had ancestors, it is certainly *possible* that something (a telephone, Adam’s heart) be the first of its kind and still function properly. The condition proposed, therefore, taken as an analysis, is too strong; it isn’t *necessary*.

It’s not sufficient either, as the following shows.³⁶ Imagine that a Hitler-like madman gains control: as part of his Nietzschean plan to play God, he orders his minions to induce a genetic mutation in selected non-Aryan victims. Those born with this mutation can’t see at all well (their visual field is a uniform shade of light green with little more than a few shadowy shapes projected on it). When they open their eyes and use them, furthermore, the result is constant pain. As a result, they are unable to listen to music, or read (or write) poetry or literature; they can’t do mathematics or philosophy or evolutionary biology; they can’t enjoy humor, play, adventure, friendship, or any of the other things that make for human flourishing. Their lives are nasty, poor, brutish, and short. By way of amusing themselves, this Hitler and his henchmen also begin a program of weeding out the non-Aryan non-mutants before they reach reproductive maturity. Contrary to their intention, however, the mutation spreads; it gets out of control; after a few generations the bulk of the world’s population, including many of the Aryans themselves, display it; the number of non-mutants dwindles.

But now take some *n*th-generation non-Aryan mutant *m* and consider his visual system *A* and its way of working *F*. In accordance with Millikan’s definition, *A* originated as a reproduction or copy of some prior item, i.e., the visual

³⁵This has proven a hard point to communicate. In WPF (p. 201) I said the very same thing and quoted Millikan’s disclaimer: “Proper function is intended as a technical term. It is of interest because it can be used to unravel certain problems, not because it does or doesn’t accord with common notions such as ‘purpose’ or the ordinary notion of ‘function’.” John Post, however, comments as follows: “Plantinga . . . badly misreads [Millikan] as attempting an analysis, then tries to counterexample accordingly” (“Critical notice of Ruth Millikan’s *White Queen Psychology and Other Essays for Alice*,” *Philosophy and Phenomenological Research* 58, p. 233).

³⁶Here I return to an example offered in WPF.

system of his ancestors. Those earlier visual systems worked in way F , the same way m 's visual system works, and they worked that way due to the possession of the properties reproduced. Further, m 's visual system A exists in part because his ancestors' visual systems worked that way; that way of functioning conferred a survival advantage, in that this Hitler, his thugs, and their successors were selectively eliminating those who didn't display it, allowing those who did to live. So working in way F , for m 's visual system, meets Millikan's conditions for functioning properly. But wouldn't it be wrong (not to mention crazy) to say that m 's visual system is functioning properly? Or that its function is to produce both pain and a visual field that is uniformly green? Or that the resistance medical technicians who desperately try to repair the damage are interfering with the proper function of the visual system? So Millikan's conditions are neither necessary nor sufficient. Let me repeat: she wasn't *trying* to give necessary and sufficient conditions for the notion of proper function, but for some other notion she thinks will be useful in solving certain problems; so this isn't a problem for her project. But it does show that her account doesn't contain the resources for an accurate account of the notion of proper function.

One prominent account I didn't consider in *Warrant and Proper Function* is that of Larry Wright,³⁷ who, according to Michael Levin, held that "an effect F of S is a function of S just in case S exists or persists because it F 's; i.e., a thing's functions are those of its effects that explain it."³⁸ A heart does many things; its functions are those things it does that explain its existence or persistence: pumping in such a way as to circulate the blood would be an example. Unlike Neander and Millikan, Wright does propose this as an analysis of function, i.e., our concept of function. It's easy to see, however, that this condition is neither necessary nor sufficient. It's not sufficient: a thing might persist because of some feature that wasn't its function or one of its functions, as in the Hitler case I described above. There, a given visual system might continue to exist because it (mal)functions in the way intended by the Nazis; but of course that way of working is not among its functions. The definition has another unhappy feature: it restricts the functions of a thing to the *actual effects* of a thing. But the function of a smoke detector is to detect smoke and sound the alarm; that is its function, even if it never actually detects any smoke or gives an alarm and (therefore) never has smoke detection or sounding the alarm as an actual effect. A diseased heart still has the function of pumping blood in a certain fashion, even if the pumping of blood in that fashion is (because of its diseased condition) not among its effects.

³⁷In "Functions," *The Philosophical Review* 82 (1973): 139–68.

³⁸"Plantinga on Functions and the Theory of Evolution," *Australasian Journal of Philosophy* 75 (1997): 86. Page references hereafter given in the text.

Levin notes some *other* difficulties with Wright's analysis and, in the course of an interesting and useful discussion of functions, attempts to repair it as follows:

F is a function of *S* if and only if "*S* is explained by its leading to *F* and the efficient cause *S'* of *S* is explained by its leading to *S*." (p. 89)

This is a *strengthening* of Wright's analysis: Levin adds the second clause of the *analyzans* in order to deal with the difficulties he mentions in Wright's analysis as it stands (Levin, pp. 88–9). By way of example, pumping blood is a function of a human heart in that the existence of the heart is explained by its 'leading to' the pumping of blood (there are such things as hearts because they pump blood); and the existence of the efficient cause of the heart—whatever mechanism it is in human beings that causes the existence of hearts—is explained by its 'leading to' or in this case causing hearts.

I propose to argue that this attempt is as unsuccessful as the rest. Note first that it is a little vague: when does a thing *S* *lead to* something *S'*? What is this 'leading to'? Causing? Being part of a cause? Making probable or making more probable? Being an element in a causal chain resulting in? Being followed by? Can it be that something *A* leads to something *B*, and *B* also leads to *A*, as with drinking and depression? Note second that *explanation* is also a slippery customer. What explains what is relative to context and interests. What explains the way the visual system of that mutant works? The cruelty of the Nazis, or the technological excellence of their minions, or the nature of the gene involved in the mutation, or the fact that one of *m*'s parents had the gene, or that the gene is dominant, or In one context an event or state of affairs *A* can be explained by a state of affairs *B*, and in another *B* can be explained by *A*. What explains the fact that the porch is shady? The position of that big shade tree. What explains the position of that shade tree? It was planted there so that it can shade the porch. This fact about explanations introduces a certain flexibility (not to mention flaccidity) into Levin's analysis. Further, it's not immediately obvious how Levin's analysis applies to individual organs or systems, like a heart. Is the existence or persistence of my heart to be explained by its 'leading to' the pumping of my blood in the relevant way? But what if it is defective, and doesn't pump blood in that way, perhaps only beating at the rate of twenty-five beats per minute? Wouldn't it still be its function to pump blood in that relevant way? In response to this problem Levin says we should think instead about *other* hearts: "my heart exists (in part) because of the blood-pumping of the hearts of my ancestors" (p. 87).

Even with this nuance it is not hard to see that Levin's condition too is neither necessary nor sufficient. First, it isn't necessary: it is possible that *F* is a function of *S* even if Levin's condition isn't met. Consider again the fact that God could

have created Adam (or Eve) directly; if he had, the function of Adam's heart would have been just what the function of our hearts is, namely to circulate the blood in a certain way. But (the second clause of) Levin's condition isn't met in this case: it is not the case that, under these conditions, the efficient cause of Adam (namely God) is explained by his 'leading to' Adam's heart. God isn't explained by anything at all, and in particular isn't explained by his being the cause of Adam and Eve.

It is equally easy to see, I think, that the condition isn't sufficient. Consider again that Hitler scenario. Take a given mutant m and his visual system S , which works in that unfortunate way. The existence of S is explained by its working in that way: working in that miserable way kept m (or m 's ancestors) from being killed by the Nazis. The efficient cause of S —whatever system it is, in human beings, that causes the existence of visual systems—furthermore, is explained by its leading to S . In this case, then, the proposed necessary and sufficient condition is met; but it is not the function of m 's visual system to cause pain and display only a uniform green visual field with a few shadowy figures projected on it.

Now here Levin protests. Speaking of my example, he says that the non-Aryans wouldn't be able to get around and reproduce with a visual system as defective as all that, unless the Nazis let them: "They would be unable to find their way around by themselves in a world full of hostile Nazis, and their constant discomfort would presumably dull their sex drive. There would *be* no non-Aryans after the first generation . . ." (p. 91). Under these conditions, says Levin, "non-Aryans would have to be *grown* by Nazis, much as exotic plants and animals are grown by human breeders now" (p. 91). What this means, he says, is that the mutant's visual systems *really are* working properly, because these traits are purposely bred in by the Nazis. In the same way, we might breed a line of dogs with very small and very dull teeth suitable only for eating oatmeal and jello; such dental systems would be working properly, fulfilling their function, even if they couldn't fulfill the functions fulfilled by their ancestors' teeth a few generations back.

What do I have to say for myself? The first thing to say is that Levin is apparently trying to hijack my example. It's my example, after all, and I get to decide on its details. The fact is that in my example the mutants manage to get around very well, partly because of the help of non-mutant non-Aryans and also anti-Nazi Aryans. Their sex drives, furthermore, are not dulled; in fact their only consolation, in their miserable condition, is sex (during which they keep their eyes tightly closed). It is therefore not the case that in my example the mutants are being "grown" by the Nazis. Still, we needn't argue about that. That is because, as you recall, in my example the mutation gets out of control: it spreads to many of the *Aryans*. So consider some *Aryan* mutant m^* : the way his visual system works also meets Levin's two-part condition for being its function.

The second condition is obviously met. The first condition is also met: why does this visual system, with its peculiar way of working, exist? There will be several explanations (for example, that one of his ancestors mated with a non-Aryan mutant); but among them is the fact ancestors of this visual system functioned in that way, thus escaping the wrath of the Nazis.

There are plenty of simpler examples. Why does my car continue to exist? Because it spun out and landed in the ditch just moments before it would have been destroyed by that onrushing passenger train I didn't see. *Spinning out and landing in the ditch* seems to meet Levin's conditions for being the function of my car. The first condition is obviously satisfied: the (an) explanation of my car's (still) existing is that it spun out and landed in the ditch. But so, clearly, is the second: consider whatever mechanism *m* it is (in a GM plant somewhere) that is the efficient cause of my car's existing (whereby my car came to exist): an explanation of *m* is that it 'leads to' the existence of cars, including mine. Should we conclude that it is the or a function of my car to spin out and land in the ditch? Hardly. I am about to drive my old car to the junkyard, where it would have been destroyed; unfortunately, it won't start. *Not starting* meets Levin's conditions. The second condition is clearly fulfilled; but so is the first: an explanation of my old car's (still) existing is its failure to start. But failing to start (one hopes) is hardly a function of my car. Next week my garage will run a contest to see whose car takes the most oil; the winner gets a three-day trip to Philadelphia. (Second prize, as you undoubtedly know, is a weeklong trip to Philadelphia.) I would have destroyed my oil-guzzling junker a month ago, except that it occurred to me I might win that contest. So my junker exists now because it takes a lot of oil; that is hardly one of its functions. Levin's proposed necessary and sufficient condition is neither necessary nor sufficient.

As far as I know, no one has been able to come up with a naturalistic analysis of proper function that is anywhere nearly adequate or accurate, and by now the project is beginning to look unhopeful. The fundamental reason, I suggest, is that this notion, the notion of function or proper function, essentially involves the aims and intentions of one or more conscious and intelligent designers. The notion of proper function really implies the idea of design by conscious, intentional, and intelligent designers. But that means that the organs and parts of plants, animals, and human beings can function properly (or improperly) only if they are designed and caused to be by one or more conscious, intelligent agents. Of course that is no problem for theism. According to theism—Jewish, Muslim, and Christian theism anyway—God has designed and created human beings and other creatures. But there *is* a serious problem here for naturalists. Naturalists, of course, can't think of human beings as being designed and created by a being like God, they can't think of human beings or their systems or organs as functioning properly (or, for that matter improperly). This means that naturalism has no place for proper function and other allied properties

such as health, sickness, sanity, insanity, and the like: if naturalism were true, nothing would display any of these properties. Nor would anyone know anything, if I am correct in thinking that the analysis of warrant crucially involves proper function. This is my first complaint about naturalism.

B. *Naturalism, skepticism, self-defeat*

It is unfair to Descartes to call his appeal to God's credibility frivolous. Indeed, only if we assume a God who is morally our like can "truth" and the search for truth be at all something meaningful and promising of success. This God left aside, the question is permitted whether being deceived is not one of the conditions of life.

—Friedrich Nietzsche³⁹

Much of what Nietzsche says doesn't inspire confidence, but here he may be on to something. For suppose you are a naturalist: what I'll argue is that you have a good and sufficient reason for doubting that your beliefs are mostly true. More exactly, you have a good reason for doubting that your cognitive faculties—your perception, memory, rational intuition, and the like—are reliable, provide you with mostly true beliefs. I'll argue that the probability that your beliefs are reliable, given what you believe about how they come to be, is low. But if that is so, then (so I'll argue) you have a *defeater* for the natural belief that your cognitive faculties are in fact reliable. This defeater is one that can't itself be defeated; therefore you have an undefeated defeater for that belief. But in that case the rational thing to do is to reject that belief, to give it up. Rationality doesn't require that you believe that your faculties are not in fact reliable; but it does require that you not believe them to be reliable. Further, if you have a defeater for the belief that your faculties are reliable, then you also have a defeater for each of the beliefs produced by those faculties; you therefore have a defeater for each of your beliefs. That means that you have a defeater for your belief in naturalism itself; hence naturalism is self-defeating.

Still further: if you have a defeater for each of your beliefs, then you are enmeshed in a particularly virulent sort of skepticism. Let me explain. One kind of skepticism—a less virulent kind—would be the position that as a matter of fact we don't really *know* much, or don't know what we think we know, or even don't really know anything. Here the emphasis is on the word 'know' or 'really know.' Now it is perfectly possible to think you don't know a given thing you believe, and still be in no particular distress about that fact. You can still

³⁹ Nietzsche: *Writings from the Late Notebooks* (Cambridge Texts in the History of Philosophy), ed. Rüdiger Bittner, tr. Kate Sturge (Cambridge: Cambridge University Press, 2003), Notebook 36, June–July 1885, p. 26.

think that the belief in question is perfectly sensible and appropriate. Skeptics tell me that I don't really know that I am not dreaming, or a brain in a vat, or a victim of a deceitful Cartesian evil demon. Well, perhaps they're right; perhaps we don't know the things we think we know. But that need not cause much distress. Perhaps the standards for knowledge are very high: perhaps I know only what is self-evident, or self-evidently follows from what is self-evident. Then I don't really know that there's been a past, or that there are other people, or that there are trees in my backyard. That needn't cause me much epistemic pain, however. It still certainly seems right and sensible to believe those things; even if those beliefs don't constitute knowledge, they are perfectly sensible, and are the right ones to hold.

But there is also a much more profound kind of skepticism; this is the sort expressed by David Hume in the early sections of his *A Treatise of Human Nature*. When he follows out what seem to be the promptings and leading of reason, when he does his intellectual best, he winds up time after time in an intellectual black hole, 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, environ'd with the deepest darkness, and utterly depriv'd of the use of every member and faculty.⁴⁰

That is a much more serious kind of skepticism. Here it's not merely that you don't know the things we ordinarily think we do know; you don't know what to believe about anything. You don't even believe that your cognitive faculties are reliable. You also don't believe that they are not reliable; you simply don't know what to think—about anything.

I propose to argue first that naturalism, construed as including materialism, implies this peculiarly virulent form of skepticism. I'll go on to argue briefly that naturalism implies skepticism even if not construed as including materialism.

1. Materialistic naturalism

Most naturalists accept *materialism* with respect to human beings: the claim that human beings are material objects. On this view human beings have no immaterial parts—no immaterial soul, or mind, or self, for example. From this perspective it is not the case that a human person is an immaterial substance or thing that is connected with or joined to (has?) a material body; nor is it

⁴⁰ *Treatise*, Selby Bigge edition (Oxford: Clarendon Press, 1888), p. 269.

the case that a human being *has* an immaterial soul or mind. Instead, so the materialist thinks, a person *just is* her body, or perhaps some part of her body,⁴¹ so that talk about ‘my body’ is misleading. I *am* my body (or perhaps my brain, or some part of it, or some other part of my body). Nearly all naturalists would agree. They give at least three sorts of reasons for materialism. First, naturalists often argue that dualism (the thought that a human being is an immaterial self or substance intimately related to a human body) is incoherent or subject to crushing philosophical difficulties; hence, so they say, we are rationally compelled to be materialists. You can find a typical set of such objections to dualism in Daniel Dennett’s book *Consciousness Explained*.⁴² Most of these objections (including Dennett’s versions) are astonishingly weak;⁴³ no one not already convinced of materialism would find them persuasive. Still, they are often trotted out as showing that we are obliged to be materialists.

A second and better reason is this: many naturalists think it is just part of naturalism as such to have no truck with immaterial souls or selves or minds. It may not be completely easy to see or say precisely what naturalism is, but, so goes the thought, at any rate it excludes things like immaterial selves or souls. Naturalism is the idea that there is no such person as God, or anything like him; immaterial selves would be too much like God, who, after all, is himself an immaterial self. This reason is really quite persuasive, but it isn’t conclusive. That is because of the vagueness of the concept of naturalism. If naturalism is true, there isn’t anything *like* God in the world; but just how much similarity to God is tolerable, from a naturalistic perspective? After all, everything resembles God in *some* respect (if only in being something or other); how much similarity to God can a reasonably sensitive naturalist manage to put up with? Plato’s idea of the good and Aristotle’s Unmoved Mover (who is also immaterial) clearly won’t pass muster, but what about immaterial soul substances? Can a proper naturalist countenance such a thing? That’s not entirely easy to say, and I will leave naturalists to decide this issue for themselves.

⁴¹ A somewhat different materialist view is that a human person is a material object distinct from but *constituted by* her body, or by the same matter that constitutes her body: see, e.g., Lynne Rudder Baker, *Persons and Bodies* (Cambridge: Cambridge University Press, 2000). My argument, I believe, will also hold for materialist views of this sort.

⁴² Boston: Little, Brown and Co., 1991. Some who don’t admire the book have complained that a better title would be *Consciousness Explained Aaway*. Dennett’s book illustrates, I think, the problem for one who accepts materialism but also (like the rest of us) can’t help thinking that there is such a thing as consciousness.

⁴³ See, e.g., Charles Taliaferro, “Incorporeality,” in *A Companion to Philosophy of Religion*, ed. Philip L. Quinn and Charles Taliaferro (Oxford: Blackwell, 1997), pp. 271ff., who does a nice job of exposing some of these weaknesses.

A third reason is as follows. Naturalists will ordinarily endorse Darwinian evolution; but how could an immaterial soul or self have come to exist by way of the processes that evolutionary science posits? Thus Richard Dawkins: “Catholic Morality demands the presence of a great gulf between *Homo Sapiens* and the rest of the animal kingdom. Such a gulf is fundamentally anti-evolutionary. The sudden injection of an immortal soul in the timeline is an anti-evolutionary intrusion into the domain of science.”⁴⁴ According to contemporary evolutionary theory, new forms of life arise (for the most part) by way of natural selection working on some form of genetic variation—the usual candidate is random genetic mutation. Most mutations of this sort are lethal; but a few are advantageous in the struggle for survival. Those lucky organisms that sport them have a reproductive advantage over those that do not, and eventually the new feature comes to dominate the population; then the process can start over. But how could an *immaterial self* or soul evolve this way? What sort of genetic mutation would result in an immaterial soul? Could there be a section of DNA that codes not for the production of proteins of a certain sort, but for an immaterial self? That seems unlikely. These reasons clearly aren’t conclusive, but most naturalists find them (or perhaps other arguments for materialism) at least reasonably compelling.

Now what sort of thing will a belief *be*, from this materialist perspective? Suppose you are a materialist, and also think, as we ordinarily do, that there are such things as beliefs. For example, you believe that Proust is more subtle than Louis L’Amour. What kind of a thing is this belief? Well, from a materialist perspective, it looks as if it would have to be something like a long-standing event or structure in your brain or nervous system. Presumably this event will involve many neurons connected to each other in various ways. There are plenty of neurons to go around: a normal human brain contains some 100 billion neurons. These neurons, furthermore, are connected with other neurons at synapses; a single neuron can be involved in many synapses. The total number of possible brain states, then, is absolutely enormous, much greater than the number of electrons in the universe. Under certain conditions, a neuron fires, i.e., produces an electrical impulse; by virtue of its connection with other neurons, this impulse can be transmitted (with appropriate modification from other neurons) down the cables of neurons that constitute effector nerves to muscles or glands, causing, e.g., muscular contraction and thus behavior.

So (from the materialist’s point of view) a belief will be a neuronal event or structure of this sort, with input from other parts of the nervous system and output to still other parts. But if this is the sort of thing beliefs are, if they are neuronal events or structures, they will have two quite different sorts

⁴⁴“When religion Steps on Science’s Turf,” *Free Inquiry Magazine* 18:2, pp. 18–19.

of properties. On the one hand there will be *electro-chemical* or *neurophysiological* properties (NP properties, for short). Among these would be such properties as that of involving n neurons and n^* connections between neurons, properties that specify which neurons are connected with which others, what the rates of fire in the various parts of the event are, how these rates of fire change in response to changes in input, and so on. But if the event in question is really a *belief*, then in addition to those NP properties it will have another property as well: it will have to have a *content*.⁴⁵ It will have to be the belief that p , for some proposition p . If it's the belief that Proust is a more subtle writer than Louis L'Amour, then its content is the proposition *Proust is more subtle than Louis L'Amour*. If it is instead the belief that Cleveland is a beautiful city, then its content is the proposition *Cleveland is a beautiful city*. My belief that naturalism is all the rage these days has as content the proposition *Naturalism is all the rage these days*. (That same proposition is the content of the German speaker's belief that naturalism is all the rage these days, even though he expresses this belief by uttering the German sentence 'Der Naturalismus ist diese Tage ganz gross in Mode'; beliefs, unlike sentences, do not come in different languages.) It is in virtue of having a content, of course, that a belief is true or false: it is true if the proposition which is its content is true, and false otherwise. My belief that all men are mortal is true because the proposition which constitutes its content is true, but Hitler's belief that the Third Reich would last a thousand years was false, because the proposition that constitutes its content is (was) false.

Given materialism, therefore, beliefs would be long-standing neural events. As such, they would have content, but also NP properties. Now how is it that we human beings have come to have beliefs, and how is it that we have come to have beliefs with the content those beliefs do in fact have? Naturalists (and of course not only naturalists) ordinarily believe that human beings have come to be by way of evolution; they have evolved according to the mechanisms specified in contemporary evolutionary theory. (The prime candidates are natural selection operating on some source of genetic variability such as random genetic mutation.) We have something of an idea as to the history of those NP properties: structures with these properties have come to exist by small increments, each increment such that it has proved to be useful in the struggle for survival.⁴⁶ But what about the *content* of belief? If a belief is a neuronal event, where does its content come from? How does it get to be associated in that way with a given proposition?

⁴⁵ It is of course extremely difficult to see how a material structure or event could have content in the way a belief does; below (p. 54). I'll argue that in fact such a structure *can't* have content, so that materialism cannot accommodate belief.

⁴⁶ Here we can ignore plectropy and spandrels.

Materialists offer two (or possibly three) main theories here. According to the first, content *supervenes upon* NP properties; according to the second, content *is reducible to* NP properties.

Note that if content properties are reducible to NP properties in the sense of ‘reducible’ suggested below, then they also supervene upon them. Note also that for present purposes I ignore so-called ‘wide content.’ If we were to take wide content into account, we’d say that content supervenes, not just on NP properties, but on NP properties together with certain properties of the environment. The same would go, *mutatis mutandis*, for the suggestion that content is reducible to or identical with NP properties. In the interest of simplicity, I ignore wide content; nothing in my argument below hinges on this omission.

Suppose we think about the second theory first. Consider the property of having as content the proposition *Naturalism is all the rage these days*, and call this property *C*. On the present suggestion, *C* just is a certain combination of NP properties. It might be a disjunction of such properties: where P_1 to P_n are NP properties, *C*, the property of having the content in question, might be something like (where ‘v’ represents ‘or’)

$$P_1 \vee P_3 \vee P_8 \vee \dots \vee P_n$$

More likely, it would be something more complicated: perhaps a disjunction of conjunctions, something like (where ‘&’ represents ‘and’)

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

We could put this by saying that any content property is a Boolean combination of NP properties, that is, a combination constructed from NP properties by disjunction, conjunction, and negation. And to say that content properties are reducible to NP properties is just to say that every content property is some Boolean combination of NP properties. In fact, if we think that any Boolean combination of NP properties is itself an NP property, we could say that content properties just are NP properties—a special sort of NP property, to be sure, but still NP properties. So, on this theory, content properties—e.g., the property of having *Naturalism is all the rage these days* as content—are, or are reducible to, NP properties.

That’s one of the two materialistic proposals; the other is that a content property isn’t an NP property, or a Boolean combination of NP properties, but rather *supervenes on* NP properties. What does that mean; what is this ‘supervenience’? The basic idea is that a set of properties *S* supervenes on a set of properties *S** just if any pair of objects which agree on the *S** properties must also agree on the *S* properties. For example, beauty (of a picture, a face) supervenes on molecular constitution; any two pictures (or faces) with the same molecular

constitution will be beautiful to the same degree. Content properties supervene on NP properties, then, if and only if any two objects or structures with the same NP properties must have the same content properties. You couldn't have a pair of structures—neuronal events, say—that had the same NP properties but different contents.⁴⁷ Content is a *function* of NP properties.

We can put this officially as follows:

(S) Necessarily, any structures that have the same NP properties have the same content.

This is a *weak* form of supervenience; a stronger one could be put as

(S⁺) For any possible worlds W and W^* and any structures S and S^* , if S has the same NP properties in W as S^* has in W^* , then S has the same content in W as S^* has in W^* .

If we think of supervenience as involving nomic rather than broadly logical necessity, then in (S⁺) we'll quantify just over nomically possible worlds, not possible worlds *simpliciter*. Those who think that content properties supervene on NP properties for the most part think, I believe, that the former supervene on the latter in the stronger sense (S⁺) (and hence also, of course, in the weaker sense (S)). For present purposes, however, it doesn't matter which sense we employ.

But what about that "necessarily"? Here this supervenience suggestion divides into two branches. On the first branch, the necessity in question is broadly logical necessity, the sort of necessity enjoyed by the truths of logic and mathematics, but by many other propositions as well. (For example, such propositions as *No prime minister is a prime number*, *No people are numbers*, *Bachelors are unmarried*, *Dogs are animals*, and *There aren't any things that do not exist*.) Necessary propositions are true in every possible world. If a proposition p is necessary, then every way things could have been is such that if things had been that way, p would have been true. And now the current suggestion is that the proposition *Any structures that have the same NP properties also have the same content* is necessary in that same sense.

According to the other branch of the supervenience theory, the necessity in question isn't broadly logical necessity, but something more obscure—something we could call 'causal' or 'natural' or 'nomic' necessity. The idea is that some propositions aren't necessary in that broadly logical sense, but still enjoy a certain sort of necessity. Consider Newton's Law of Gravitation, for example, according to which any two physical objects attract each other with a force proportional to the sum of their masses and inversely proportional to the square

⁴⁷ So the second possibility is really a special case of the first: if content properties are reducible to NP properties, then clearly structures with the same NP properties will have the same content properties.

of the distance between them. This proposition doesn't seem to be necessary in the broadly logical sense. It could have been false; there are possible worlds in which it is false.⁴⁸ Still, the proposition seems to be necessary in some way; it is certainly not just a coincidence that objects behave in this way. It seems that a pair of objects *must* behave in this way; it isn't that they just *happen* to. So there seem to be two kinds of necessity; hence we can speak of two kinds of supervenience, weak and strong. According to strong supervenience, it is necessary in the broadly logical sense that any two structures that have the same NP properties also have the same content; according to weak supervenience, that proposition has the kind of necessity had by the laws of nature.

Return now to the question that led us into reduction and supervenience: how does it happen that those neural structures, the ones that constitute belief, have *content*? Where does it come from and how do they get it? The basic idea is something like this. As we go up the evolutionary scale, we find neural structures with greater and greater complexity. Near one end of the scale, for example, we find *C. elegans*, a small but charismatic worm with a nervous system composed of only a few neurons. (The nervous system of *C. elegans* has been completely mapped.) At the other end of the scale there are human beings, whose brains contain many billions of neurons connected in complex and multifarious ways, so that the number of different possible brain states is absolutely enormous. And now the idea is that as you rise in the evolutionary scale, as you go through more and more complex neural structures, at a certain point content shows up. At a certain level of complexity, these neural structures start to display content. Perhaps this starts gradually and early on (possibly *C. elegans* displays just the merest glimmer of consciousness and the merest glimmer of content), or perhaps later and more abruptly; that doesn't matter. What does matter is that at a certain level of complexity of neural structures, content appears. This is true whether content properties are reducible to NP properties or supervene on them.

So (given materialism) some neural structures at a given level of complexity acquire content; they thus become beliefs. And the question I want to ask is this: what is the likelihood, *given naturalism*, that the content that thus arises is in fact *true*? In particular, what is the likelihood, given *N*, that the content associated with *our* neural structures is true? More generally, what is the likelihood, given naturalism, that our cognitive faculties are reliable, thereby producing mostly true beliefs?

We all commonsensically assume that our cognitive faculties are for the most part reliable, at least over a large area of their functioning. I remember where

⁴⁸Indeed, according to contemporary physics it is false in the actual world, although it is a good approximation to the truth.

I was last night, that I've just had cold cereal for breakfast, that my elder son's name is not Archibald, that a year ago I lived in the same house I live in now, and much else besides. I can see that the light is on in my study, that the flower garden is overgrown with weeds, and that my neighbor put on weight over the winter. I know a few truths of mathematics and logic, mostly pretty simple, no doubt, but still . . . The natural thing to assume, and what we all do assume (at least before we are corrupted by philosophy (or neuroscience)), is that when our cognitive faculties aren't subject to malfunction, then, for the most part and over a wide area of everyday life, the beliefs they produce in us are true. We assume that our cognitive faculties are reliable. But what I want to argue is that the naturalist has a powerful reason against this initial presumption and should give it up.

By way of entering this argument, suppose we conduct a thought experiment. Consider a hypothetical species that is cognitively a lot like us: members of this species hold beliefs, make inferences, change beliefs, and the like. And let us suppose naturalism holds for them; they exist in a world in which there is no such person as God or anything like God. Our question, then, is this: what is the probability that their cognitive faculties are reliable? Consider any particular belief on the part of one of these hypothetical creatures. That belief, of course, is a neural structure of a given sort, and one sufficiently complex to generate content. We may add, if we like, that this structure occurs or takes place in response to something in the environment; perhaps it is a certain pattern of firing of neurons in the optical portion of the brain, and perhaps this pattern arises in response to the appearance of a predator in the middle distance. And a certain proposition has somehow come to be associated with this structure, so that the structure acquires belief content and is a belief.

Now what is the probability (given naturalism) that this proposition is true? Well, what we know about the belief in question is that it is a neurological structure that has certain NP properties, properties the possession of which is logically or causally sufficient for the possession of that particular content. We are assuming also that this structure arises in response to the presence of that predator, and we can also assume, if we like, that this structure is a reliable indicator of that kind of predator. This structure, we may suppose, arises when and only when there is a predator in the mid-distance. Even so, the content generated by this structure, on this occasion, need have nothing to do with that predator, or with anything else in the environment. Indication is one thing; belief content is something else altogether, and we know of no reason why the one should be related to the other. Content simply arises upon the appearance of neural structures of sufficient complexity; there is no reason why that content need be related to what the structures indicate, if anything. The proposition constituting that content need not be so much as *about* that predator.

So what, then, is the likelihood that this proposition, this content, is true? Given this much, shouldn't we suppose that the proposition in question is as likely to be false as true? Shouldn't we suppose that the proposition in question has a probability of roughly one/half of being true? Shouldn't we estimate its probability, on the condition in question, as in the neighborhood of .5? That would be the sensible course. Neither seems more probable than the other; hence we estimate the probability of its being true as .5.

The probability we are thinking of here is objective,⁴⁹ not the personalist's subjective probability, and also not epistemic probability. (Of course there will be a connection between objective and epistemic probability, perhaps a connection in the neighborhood of Miller's Principle; presumably epistemic probability will in some way follow known objective probability.)⁵⁰ But then, in suggesting the first attitude above, am I not relying upon the notorious Principle of Indifference? We are trying to estimate the probability that the content in question is true, given that it is generated by adaptive neural structures; I say that given this condition, for all we can see, it is as likely to be false as to be true; so we should judge that probability to be .5. Isn't that to endorse some version of the Principle of Indifference? And hasn't that principle been discredited?⁵¹ Not really. The Bertrand paradoxes show that certain incautious statements of PI come to grief—just as Goodman's grue/bleen paradoxes show that incautious statements of a principle governing the projection of predicates or properties comes to grief. But, of course, the fact is we project properties all the time, and do so perfectly sensibly. In the same way, I think, we often employ a principle of indifference in ordinary reasoning, and do so quite properly. We also use it in science—for example, in statistical mechanics.⁵² Of course, problems arise where there are equally natural or plausible ways of analyzing a situation into the relevant possibilities.

But suppose, for some reason, we take a somewhat different attitude to this probability: how could we possibly know, we ask, what this probability is? For all we can tell, it is very high; but also, for all we can tell, it is very low. We really can't form any opinion at all as to what it is; this probability is inscrutable for us.

⁴⁹ See WPF, ch. 9. It's worth noting that the argument can also be conducted in terms of epistemic probability, although I don't have space here to show how.

⁵⁰ See WPF p. 163.

⁵¹ See, e.g., Bas van Fraassen's *Laws and Symmetry* (Oxford: Clarendon Press, 1989), pp. 293ff.

⁵² “. . . an astonishing number of extremely complex problems in probability theory have been solved, and usefully so, by calculation based entirely on the assumption of equiprobable alternatives.” Roy Weatherford, *Philosophical Foundations of Probability Theory* (Routledge and Kegan Paul, 1983), p. 35. See also Robin Collins' "A Defense of the Probabilistic Principle of Indifference" (lecture to History and Philosophy of Science Colloquium, Univ. of Notre Dame, Oct. 8, 1998, presently unpublished).

This, too, seems a sensible option. My argument, fortunately, will work just as well given the premise that the relevant probability is inscrutable.

But aren't we forgetting something important? These hypothetical creatures have arisen, presumably, by way of evolution. They have come to be by way of something like natural selection working on some process of genetic variation—perhaps random genetic mutation. Presumably, then, it has proven adaptively useful for creatures of that sort to display that neural structure in the circumstances in which this creature finds itself. This structure's arising in those circumstances has (or had) survival value; it contributes to the reproductive fitness of the creature in question, presumably by helping cause the right sort of behavior (fleeing, or wary watchfulness, maybe). Whatever exactly the appropriate action is, the neuronal event in question is useful because it is a cause (part-cause) of that behavior. And doesn't that mean that it's likely that the content associated with this structure is in fact a true proposition?

It is crucially important to see that the answer to this question is NO. This neuronal event or structure has NP properties such as sending electrical signals to other parts of the nervous system as well as to muscles and/or glands. By virtue of these NP properties, it causes adaptive behavior such as fleeing. This neuronal structure also displays NP properties that are sufficient, causally or logically, for the presence of content. As a result of having that neuronal event with that particular constellation of NP properties, the creature in which this event is to be found also believes a certain proposition. But what reason is there to think that proposition *true*? Granted, the structure in question helps cause adaptive behavior. But that doesn't so much as slyly suggest that the content that gets associated with the structure is true. As far as its causing the right kind of behavior is concerned, it simply doesn't matter whether the content, that associated proposition, is true or false. At this point, as far as the truth or falsehood of the content that arises, natural selection just has to take potluck. (Not that it minds—it's interested, so to speak, just in adaptive behavior, not in true belief.) Natural selection selects for structures that have adaptive NP properties; as it happens, these structures are of sufficient complexity to generate content; but there isn't even the faintest reason to think that content true. Given naturalism, it would be sheer coincidence, an enormous cosmic serendipity, if the content that is associated with adaptively useful NP properties should also turn out to be all or mostly true content. Naturalists who think content supervenes on NP properties (and that would be most naturalists) tend to assume automatically (at least when it comes to us human beings) that the content in question *would be* true; but why think that? This assumption is at best a piece of charming but ingenuous piety. Given naturalism, the belief in question is as likely to be false as to be true.

So, with respect to the relevant facts about the origin and provenance of this particular belief on the part of this hypothetical creature, the probability of its being true—i.e., the probability that the content of the neural structure in question should be a true proposition—would have to be estimated as about .5. The associated content in question could, of course, be true; but it could also, and with equal likelihood, be false.

What, then, is the probability that the cognitive faculties of these creatures will be *reliable*? A reliable belief-producing faculty will produce a considerable preponderance of true belief over false belief. We ordinarily think our cognitive faculties are more reliable in some circumstances than in others: we are good at such things as remembering what we had for breakfast or perceiving whether there are any trees in the backyard; we are less good at determining (without artificial aids) whether a mountain goat we see at 500 yards has horns. We are also less reliable when working at the limits of our faculties, as in trying to determine what happened in the first 10^{-33} seconds after the Big Bang. (Given all the disagreements, perhaps we are also less reliable when it comes to philosophy.) But any reasonable degree of reliability, as we ordinarily think of it, requires producing a substantial preponderance of true beliefs. A thermometer that didn't produce more true than false readings (within the appropriate limits of error) would not be reliable.

As we saw above, it's not enough that it produces more true than false readings, or even that it produces only true readings. A reliable thermometer must produce a preponderance of true readings not just in fact, but also in the appropriately close possible worlds. Just how much of a preponderance? Well, of course it won't be possible to come up with a precise figure here; but surely a thermometer that doesn't produce true readings in more than, say, $\frac{3}{4}$ of the appropriate circumstances can't be accounted reliable.

And the same sort of thing goes for the reliability of cognitive faculties; they, too, are reliable, and reliable in a certain area, only if they produce a preponderance of true beliefs over false. Going back to those hypothetical creatures, what we've seen is that the probability, on the relevant condition, that any given belief of theirs should be true is in the neighborhood of $1/2$. This means that the probability that their faculties produce the preponderance of true beliefs over false required by reliability is very small indeed. If I have 1000 independent⁵³ beliefs, for example, the probability (under these conditions) that three quarters or more of these beliefs are true (certainly a modest enough

⁵³'Independent': it could be that a pair of neural structures with content were such that if either occurred, so would the other; then the beliefs in question would not be independent. Similarly when the content of one neural structure entails the content of another: there, too, the beliefs in question won't be independent.

requirement for reliability) will be less than 10^{-58} .⁵⁴ And even if I am running a modest epistemic establishment of only 100 beliefs, the probability that $\frac{3}{4}$ of them are true, given that the probability of any one's being true is $\frac{1}{2}$, is very low, something like .000001. So the chances that this creature's true beliefs substantially outnumber its false beliefs (even in a particular area) are small. The conclusion to be drawn is that it is very unlikely that the cognitive faculties of those creatures are reliable.

So far what we've seen is that, given naturalism and the supervenience of content upon NP properties, it is unlikely that the cognitive faculties of these creatures are reliable; this is true even if we add that the content of their beliefs is generated by structures with NP properties that are fitness-enhancing, adaptively useful.

That's how things stand if content *supervenes* upon NP properties. But what about the other option, reductionism? What if content properties (for example, the property of having as content the proposition *Naturalism is all the rage these days*) just *are* NP properties, or complex clusters of NP properties? In this case we get the very same results. To see why, consider, again, a given belief on the part of a given member of that hypothetical group of creatures. That belief, of course, is a neuronal event, a congeries of neurons connected in complex ways and firing away in the fashion neurons are wont to do. This neuronal event displays a lot of NP properties. Again, we may suppose that it is adaptively useful for a creature of the kind in question to harbor neuronal structures of the sort in question in the circumstances in question. The event's having the NP properties it does have is fitness-enhancing in that by virtue of having these properties, the organism is caused to perform adaptively useful action—fleeing, for example. But some subset of these NP properties together constitute its having a certain content, constitute its being associated, in that way, with some proposition. What is the probability that this content is true? What is the probability that the associated proposition is a true proposition? The answer is the same as in the case we've already considered. The content doesn't have to be true, of course, for the neuronal structure to cause the appropriate kind of behavior. It just happens that this particular arrangement of adaptive NP properties also constitutes having content. But again: it would be a piece of enormous serendipity if this content, this proposition, were *true*; it could just as well be false. So the probability that this content is true would have to be rated at about $\frac{1}{2}$, just as in the case of supervenience. If this is true for each of the independent beliefs of the organism in question, the probability (on naturalism) that the cognitive faculties of these creatures are reliable would have to be rated as low. The conclusion to be drawn so far, then, is that given naturalism, it is unlikely that these creatures have reliable cognitive faculties.

⁵⁴My thanks to Paul Zwier, who performed the calculation.

Now the next step in the argument is to note that of course what goes for these hypothetical creatures also goes for us. Suppose naturalism (construed as including materialism) is in fact true with respect to us human beings: there is no such person as God or anything like God. Then the probability that our cognitive faculties are reliable is low, just as in the case of those hypothetical creatures. For us, too, the main possibilities would have to be supervenience (logical or causal) and reduction or identity. In our case, too, if we focus on any particular belief—say, the belief that naturalism is all the rage these days—on the part of a particular believer, we see that this belief (given materialism) will have to be a neuronal event of some kind. This event will be of sufficient complexity to generate content (by supervenience or reduction); somehow a proposition gets associated with it as its content. We may suppose, if we wish, that it is adaptively useful for creatures like us to harbor structures of that kind in the circumstances in which the believer finds herself. It would be the merest coincidence, however, if the content generated by the structure in question should be *true* content, if the proposition which is the content of the belief in question should turn out to be a *true* proposition. That means that the probability of this belief's being true would have to be judged to be in the neighborhood of 1/2, not much more likely to be true than to be false. But then it will be exceedingly improbable that the whole set of this believer's beliefs should display the preponderance of true belief over false required by the reliability of her cognitive faculties. So our case is like that of those hypothetical creatures; in our case, too, the probability that our cognitive faculties are reliable, given naturalism, is low. Let ' $P(\dots/ _)$ ' stand for 'the probability of ... on $_$,' let ' R ' stand for the proposition that our cognitive faculties are reliable, and ' N ' stand for naturalism (construed as including materialism). We can then put this point briefly as ' $P(R/N)$ is low.' (If we like, we can include in ' N ' the proposition that our cognitive faculties have come to be by way of the processes proposed in current evolutionary theory.)

But now let's take one more step: a person who accepts naturalism and recognizes that $P(R/N)$ is low, thereby acquires a *defeater* for R . A defeater⁵⁵ for a belief B I hold—at any rate this kind of defeater—is another belief B

⁵⁵ Of course there are several kinds of defeaters; here it isn't necessary to canvass these kinds. The kind of defeater presently relevant would be a *rationality* defeater, and an *undercutting* rationality defeater. In addition to rationality defeaters, there are also *warrant* defeaters; these, too, come in several kinds. For more on defeaters, see Michael Bergmann, "Deontology and Defeat," *Philosophy and Phenomenological Research* 60 (2000): 87–102, and "Internalism, Externalism and the No-Defeater Condition," *Synthese* 110 (1997): 399–417, and see my "Reply to Beilby's Cohorts," in *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*, ed. J. Beilby (Ithaca, NY: Cornell University Press, 2002), pp. 205–11.

I come to hold which is such that given that I hold B^* , I can no longer rationally hold B . For example, I look into a field and see what I take to be a sheep. You come along, identify yourself as the owner of the field, and tell me that there aren't any sheep in that field and that what I see is really a dog that's indistinguishable from a sheep at this distance. Then I give up the belief that what I see is a sheep. Another example: on the basis of what the guidebook says I form the belief that the University of Aberdeen was established in 1695. You, the university's public relations director, tell me the embarrassing truth: this guidebook is notorious for giving the wrong date for the foundation of the University. (Actually it was established in 1595.) My new belief that the University was established in 1595 is a defeater for my old belief. In the same way, if I accept naturalism and see that $P(R/N)$ is low, then I have a defeater for R ; I can no longer rationally believe that my cognitive faculties are reliable.

The problem isn't that I don't have enough *evidence* for R , to believe it rationally. The fact is I don't *need* evidence for R . That's a good thing, because it doesn't seem possible to acquire evidence for it, at least if I have any doubts about it. For suppose I think up some argument for R , and on the basis of this argument come to believe that R is indeed true. Clearly this is not a sensible procedure; to become convinced of R on the basis of that argument, I must, of course, believe the premises of the argument, and also believe that if those premises are true, then so is the conclusion. But if I do that, I am already assuming R to be true, at least for the faculties or belief-producing processes that produce in me belief in the premises of the argument and belief that if the premises are true, so is the conclusion. As the great Scottish philosopher Thomas Reid says,

If a man's honesty were called into question, it would be ridiculous to refer to the man's own word, whether he be honest or not. The same absurdity there is in attempting to prove, by any kind of reasoning, probable or demonstrative, that our reason is not fallacious, since the very point in question is, whether reasoning may be trusted.⁵⁶

My accepting any argument for R , or any evidence for it, would clearly presuppose my believing R ; any such procedure would therefore be viciously circular.

More important, however, is the following. We all naturally assume R , and assume it from our earliest days as cognitive agents. Now rationality is best explained in terms of proper function: a belief is rational, in a given set of circumstances, just if a rational person, one whose cognitive faculties are functioning

⁵⁶ *Essays on the Intellectual Powers of Man* in *Thomas Reid's Inquiry and Essays*, ed. Ronald Beanblossom and Keith Lehrer (Indianapolis: Hackett Publishing Co., 1983), p. 276.

properly, could hold that belief in those circumstances.⁵⁷ But then clearly it is perfectly rational to assume, without evidence, that your cognitive faculties are functioning reliably. We rational agents do this all the time, and do not thereby display cognitive malfunction. You might wind up in a care facility for believing that you are Napoleon, but not for believing that your cognitive faculties are functioning reliably. It is therefore perfectly rational to believe R , and to believe it in the basic way, i.e., not on the basis of propositional evidence.

But that doesn't mean that it is not possible to acquire a defeater for R ; even if a belief is properly basic it is still possible to acquire a defeater for it. In the above example about the sheep in the field, my original belief, we may suppose, was basic, and properly so; I still acquired a defeater for it. Here is another famous example to show the same thing. You and I are driving through southern Wisconsin; I see what looks like a fine barn and form the belief *Now that's a fine barn!* Furthermore, I hold that belief in the basic way; I don't accept it on the basis of evidence from other propositions I believe. You then tell me that the whole area is full of barn façades (indistinguishable, from the highway, from real barns) erected by the local inhabitants in an effort to make themselves look more prosperous than they really are. If I believe you, I then have a defeater for my belief that what I saw was a fine barn, even though I was rational in holding the defeated belief in the basic way. It is therefore perfectly possible to acquire a defeater for a belief B even when it is rational to hold B in the basic way. This is what happens when I believe naturalism, and come to see that $P(R/N)$ is low: I acquire a defeater for R . I can then no longer rationally accept R ; I must be agnostic about it, or believe its denial.

Perhaps we can see more clearly here by considering an analogy. Imagine a drug—call it XX —that destroys your cognitive reliability. Ninety-five percent of those who ingest XX become cognitively unreliable within two hours of ingesting it; they then believe mostly false propositions. Suppose further that I now believe both that I've ingested XX a couple of hours ago and that $P(R/\text{I've ingested } XX \text{ a couple of hours ago})$ is low; taken together, these two beliefs give me a defeater for my initial belief that my cognitive faculties are reliable. Furthermore, I can't appeal to any of my other beliefs to show or argue that my cognitive faculties are still reliable. For example, I can't appeal to my belief that my cognitive faculties have always been reliable in the past or seem to me to be reliable now; any such other belief is also now suspect or compromised, just as R is. Any such other belief B is a product of my cognitive faculties; but then in recognizing this and having a defeater for R , I also have a defeater for B .

Of course not just any belief with respect to which R is unlikely is a defeater for R . It is not the case that for just any belief A I have and belief B I acquire,

⁵⁷ See my WCD, pp. 133–7.

if $P(A/B)$ is low, then B is a defeater, for me, for A . I'm looking (from up close) at a sheep in the field and form the belief that (A) there is a sheep in the field; you come along and tell me that (B) at least 85% of the time there are no sheep there. I take your word for B , and $P(A/B)$ is low; still, B isn't a defeater, for me, for A . I learn that 2483 is prime. Given just that information it is unlikely that there are exactly three books on my desk; I don't thereby acquire a defeater for my belief that there are exactly three books on my desk. Can we state more general conditions under which a belief B will be a defeater, for S , for a belief A ? Following and adapting a suggestion of Michael Rea's,⁵⁸ we might try:

(D) B is a defeater for A , for S , if (but not only if) (1) S sees that $P(A/B)$ is low, and (2) there is no experience E S has or proposition P (distinct from A) S believes such that the epistemic probability of A on $B \& E$ or $B \& P$ is high.

The application of (D) to the above cases of defeat is obvious.

But what about the case in question, where the beliefs are:

$P(R/N)$ is low & N ,

on the one hand, and R , on the other? Does the former constitute a defeater for the latter, according to (D)? Are there beliefs or experiences X such that the epistemic probability of R on

$P(R/N)$ is low & $N \& X$

is high? Say that a belief X of S is a *defeater-deflector* for R and $P(R/N)$ is low & N if the epistemic probability of R on $P(R/N)$ is low & $N \& X$ is high. Are there defeater-deflectors for R and $P(R/N)$ is low & N ? Well it certainly looks as if there are. What about R itself? That's presumably something the naturalist believes. The epistemic probability of R on

$P(R/N)$ is low & $N \& R$

is certainly high. But of course R itself isn't a proper candidate for being a defeater-deflector here. If a belief A could *itself* be a defeater-deflector for a putative defeater of A , no belief could ever be defeated.⁵⁹ Which beliefs are such that

⁵⁸ See his *World without Design: the Ontological Consequences of Naturalism*, p. 205. Rea puts his principle in terms of epistemic probability, not objective probability, and adds to the antecedent a third clause: "(3) A is believed by S not on the basis of evidence."

⁵⁹ See *Naturalism Defeated?*, p. 224.

they can properly function as defeater-deflectors? This is the *Conditionalization Problem*.⁶⁰ It isn't easy to give a complete answer, but we can say at least the following:⁶¹ first, neither R itself nor any proposition equivalent to it (e.g., $(R \vee (2 + 1 = 4)) \& \neg(2 + 1 = 4)$) is a defeater-deflector here. Second, conjunctions of R with other propositions P the naturalist believes— $(2 + 1 = 3) \& R$ —will not be defeater-deflectors, unless P itself is; more generally, propositions P that entail R will not be defeater-deflectors, unless a result of deleting R from P ⁶² is a defeater-deflector. Finally, no proposition P that is evidentially dependent upon R for S —such that S believes P on the evidential basis of R —is a defeater-deflector for R . Thus *either R or naturalism is true* is evidentially dependent, for me, upon R , as is *either R or Friesland is larger than the US*, and *there is some true proposition P such that $P(R/N \& P)$ is high*. Given this account of defeater-deflection, principle (D) seems at the least plausible.

Two final matters. First, perhaps you believe the thing to think about $P(R/N)$ is not that it is low, but that it is *inscrutable*. How, you ask, can we possibly tell what that probability would be? Return to page 38 and the question of the probability that a belief is true, conditional on its supervening on or being reducible to adaptive NP properties. There I said that this probability should be thought of as in the neighborhood of $\frac{1}{2}$ (in which case it would be unlikely *in excelsis* that the creature's true beliefs should exceed its false with a preponderance sufficient for its cognitive faculties' being reliable). But maybe the right answer is that we just can't tell what that probability is: it's inscrutable.

There may be something to this objection. But all the argument as stated really requires is that the probability in question not be very high; that it isn't very high seems clear enough. Suppose, however, that this probability really is completely inscrutable: we haven't the faintest idea what it is. As far as we can tell, it could be as high as 1; it could also be zero; and it could be anything in between. We still get the same result. If this probability is inscrutable, then so will be $P(R/N)$; but $N \& P(R/N)$ is *inscrutable* is a defeater for R , just as is $N \& P(R/N)$ is *low*. Consider an analogy. You learn that your cousin Sam, whose cognitive faculties you have always assumed to be reliable, has ingested XX (above, p. 45). You know that *some* proportion of those who ingest XX become wholly unreliable; but you don't know what that proportion is; as far as you are concerned, $P(\text{Sam's faculties are reliable}/\text{Sam has ingested } XX)$ is inscrutable. It could be as low as zero; it could be as high as 1; and it could be anything in between. Under these conditions you have a defeater for your assumption that

⁶⁰ See *Naturalism Defeated?*, pp. 220–5.

⁶¹ and here I follow *Naturalism Defeated?* pp. 224–5.

⁶² Where P entails R , a result of deleting R from P will be any proposition Q such that Q is logically independent of R and such that P is logically equivalent to the conjunction of R with Q .

Sam's cognitive faculties are reliable. You would also have a defeater for R if you believed you had ingested XX and that $P(R/I've\ ingested\ XX)$ is inscrutable. So what the argument really requires is only that $P(R/N)$ be low or inscrutable.⁶³

Finally, there is one more wrinkle, or perhaps fly in the ointment.⁶⁴ Consider someone who is cognitively normal, and who comes to believe that she has ingested XX , that reliability-destroying drug mentioned above. This person may very well continue to assume that her cognitive faculties are functioning properly. She may very well carry on her cognitive life in the usual way, even if she becomes convinced she's contracted mad cow disease, a disease, as she believes, that renders its victims cognitively unreliable. And of course the same goes (in spades) if she believes N and sees that $P(R/N)$ is low. But (and this is the crucial point), in so doing, might she not be functioning perfectly properly, without so much as a hint of dysfunction or malfunction? The answer certainly seems to be Yes. If so, however, then given my account of defeat (in terms of proper function), she doesn't have a defeater for R in the belief that she has ingested XX or has contracted mad cow disease, and my argument fails.

Here I can only gesture at the response.⁶⁵ The first thing to see is that one who really rejects R is in a state of cognitive disaster. And some modules of our cognitive design plan are aimed not at the production of true beliefs, but at the production of other worthwhile conditions, including avoidance of disaster. For example, if you fall victim to a usually fatal disease, you may somehow think your chances are much better than is indicated by the statistics you know; this is the so-called 'optimistic override.' Your faculties may be functioning perfectly properly in producing this belief; this particular bit of the cognitive design plan is aimed, not at producing true beliefs about the possible course of your disease, but beliefs that will maximize your chances of recovery. Still, in some sense those statistics really do give you a defeater for your belief that in all likelihood you will recover. What they give you is a *Humean Defeater*. You have a Humean defeater for a belief B in a given situation if (1) the production of B is governed by a bit of the design plan that is aimed not at the production of true belief, but at some other state of affairs (such as recovery from disease or the avoidance of cognitive disaster), and (2) if only truth aimed processes were at work in this situation, you would have an ordinary rationality defeater for B . One who believes she's taken XX has a Humean defeater for R , as does someone who thinks she has mad cow disease. My claim is that the naturalist who sees that $P(R/N)$ is low has a Humean defeater for R .

⁶³The first clause of (D) (above p. 46) should thus be amended to '(1) S sees that $P(A/B)$ is low or inscrutable.'

⁶⁴As William Talbott pointed out to me.

⁶⁵For a full version of the response, see *Naturalism Defeated?*, pp. 205–11.

I therefore have a defeater for R . But if I consider R and do not believe it, then I have a defeater for any belief I take to be a product of my cognitive faculties. Naturally enough, that would be *all* of my beliefs; all of my beliefs are products of my cognitive faculties. The result so far, then, is that if I believe N (construed as including materialism) and I also see that the probability of R with respect to N is low, then I have a defeater for each of my beliefs. Since N itself is one of my beliefs, I also have a defeater for it; N , therefore, is self-defeating.

Further, if you believe N and see that $P(R/N)$ is low, you will be enmeshed in that particularly virulent sort of skepticism mentioned above (p. 30). It may be that you can't really reject R in the heat and press of day-to-day activities: for example, when you are playing poker with your friends, or building a house, or climbing a cliff. You can't think Humean thoughts about, say, induction when clinging unroped (you're free-soloing) to a rock face 500 feet up the East Buttress of El Capitan. (You won't find yourself saying, "Well, of course I can't help believing that if my foot slips I'll hurtle down to the ground and smash into those rocks, but [fleeting, sardonic, self-deprecatory smile] I also know that I have a defeater for this belief and hence shouldn't take it seriously.") But in the calm and reflective atmosphere of your study, you see that this is in fact the case. Of course you also see that the very reflections that lead you to this position are also no more acceptable than their denials; you have a universal defeater for whatever it is you find yourself believing. This is that really crushing skepticism, and it is this skepticism to which the naturalist is committed.

2. Dualistic naturalism

Now the vast majority of naturalists, I think, are materialists about human beings, and I've been conducting my argument (that naturalism implies skepticism) under the assumption that to be a naturalist is to be a materialist. However, there have been and are at least a few naturalists who are not materialists; at any rate there have been at least a few non-materialists who are in the near vicinity of naturalism.⁶⁶ Perhaps these philosophers are moved by the powerful arguments against materialism—for example, the apparent impossibility, as I'll argue below, that a congeries of neurons or any other material processes could be *about* something, or that human beings should be conscious or hold beliefs, if they were in fact material objects. Alternatively, the naturalist might be moved by the thought that it seems possible (in the broadly logical sense) for him to exist, even if neither his body nor any part of his body⁶⁷ existed. By virtue of these or other considerations, a naturalist (or near naturalist) might reject materialism about human beings; he might suppose that a human being

⁶⁶ Among them would be Bertrand Russell, C. D. Broad, possibly G. E. Moore.

⁶⁷ Nor any material object coincident with his body, if there are any such things.

is really an immaterial self possessing a body: he might be a dualist. So even though most naturalists are materialists, naturalism doesn't obviously entail or imply materialism. In what follows I want to consider, briefly, whether the naturalist can evade the above argument (that naturalism implies debilitating skepticism and is self-defeating) by rejecting materialism about human beings in favor of dualism.

The dualistic naturalist will add, of course, that an immaterial self of the kind he endorses will stand in close relation to a material body, the body of the person in question. This relation can be as tight as you please: perhaps a self can't exist without being embodied; perhaps, even, an immaterial self supervenes on the body whose self it is. The idea would be that at a certain level of neurophysiological complication, an immaterial self simply arises. It is a (metaphysically or broadly logically) necessary truth that when that degree of complexity arises—when that particular configuration of properties is instantiated on the part of the brain or nervous system of a material organism—an immaterial self simply emerges. This immaterial self *S* is so related to the underlying biological body *B* that *B* can properly be said to *belong to S*, to be *S*'s body. That is, *S* sees through *B*'s eyes, feels pleasure and pain in *B*, can directly cause *B* to move in various ways, and the like. Furthermore, perhaps *S*'s mental life supervenes on *B*'s neurophysiological properties. If so, every mental act, every act of the self, would require a material substrate, a biological basis in the brain; and if a pair of such bodies exemplify the same NP properties, they will also exemplify the same mental states. But the self itself, so to speak, is not material; and thinking, believing, imagining, loving, hating, desiring—all of the mental activities in which we engage—are really activities on the part of this immaterial self.

Now: does this way of thinking enable the naturalist to avoid the virulent skepticism the materialistic naturalist is committed to? I can't see how. First, note that if the mental life of the self *supervenes* on the NP properties of the body, then the situation here is just as it was with respect to materialism. A certain group of NP properties—presumably a group that is adaptively useful—will give rise to a given belief; but what reason is there, given naturalism, for thinking that belief *true*? It isn't as if, as in theism, the person in question has been created in the image of a God one of whose outstanding characteristics is knowledge and understanding. Given naturalism, it seems that the belief in question would be as likely to be false as true. If so, the probability that this belief is true will have to be rated in the neighborhood of .5. The same goes for the other beliefs of the person in question. But then it is monumentally unlikely that the person's beliefs will display the preponderance of true belief required by her faculties' being reliable, and $P(R/N)$ will be low.

On the other hand, perhaps this person's mental life does not supervene, either logically or causally, upon the properties of her body. Perhaps her beliefs are not determined by the state of her body; they float free of her NP properties.

Then what is the probability, given naturalism, that such a free-floating belief should be true? Given theism, we'd expect that God would have created the self in question in such a way that her beliefs, at least in many areas, would be for the most part true. But given naturalism, there isn't, of course, any God who designs us so as to resemble him in holding true beliefs. It looks as if the probability of the belief in question's being true, given naturalism, would presumably be in the neighborhood of .5. If so, once again the probability that her cognitive faculties are reliable will be very low. If she sees this, she has a defeater for *R*, and hence for her other beliefs, thus falling into that skepticism. Dualistic naturalism does no better than materialistic naturalism in eluding this objection.

C. *Naturalism vs. belief*

My final criticism of naturalism: if you are a naturalist, then (so I say) you should reject the idea that anyone ever believes anything. This is no trivial matter; one of the most obvious things about us (of course) is that we believe many things. I believe that all men are mortal, that $7 + 5 = 12$, that I live in Indiana, that some of my children live in Brazil, that Paul Q. Zwier isn't much of a tennis player, and much else besides. I believe many things, and as far as I know am not idiosyncratic in so doing; the same goes for you and all other (normal, adult) human beings. What I propose to argue is that if naturalism is true, none of us believes any of these things or anything else. But I do have to offer a caveat. What I really propose to argue is that *materialism* (with respect to human beings) has no place for belief. Most naturalists, of course, are materialists; but it isn't obvious that naturalism implies materialism. I must concede that the present objection to naturalism can be avoided by any naturalist willing to embrace substance dualism. This isn't much of a concession, however. Most naturalists appear to be less than wildly enthusiastic about substance dualism; showing that a view leads to substance dualism, they typically think, is a *reductio ad absurdum* of it.

So most naturalists are materialists. But from the perspective of materialism, there is a real problem with such mental properties as *being conscious* and *being in pain*, and such mental acts as beliefs, desires, hopes, and the like. I'll concentrate on the latter, although the former is every bit as vexing, for a materialist. Beliefs, desires, hopes, fears, and the like, are said to be *propositional attitudes*, attitudes or stances one takes towards propositions. Thus I can hope that there is a dog in my house, fear that there is a dog there, believe that there is, and desire that there be one there. In each case I adopt a certain attitude towards the proposition *there is a dog in my house*. I can believe that proposition, or hope that it is true, or fear that it is, or desire that it be. And I say there is a real problem with beliefs (and these other intentional attitudes) from the

perspective of materialism. The problem is that there is no sensible way to think about belief from that perspective; if materialism were true, then (so I'll argue) there wouldn't *be* any beliefs. A materialist should really be an *eliminativist* with respect to beliefs, i.e., someone who thinks there actually aren't any such things as beliefs.⁶⁸ According to eliminativists, the thought that there are beliefs is part of what they call 'folk psychology,' a primitive theory (so they think) developed by early and unscientific human beings, and a theory that ought to be replaced by something more scientific and up to date. Materialists, I say, should agree with eliminativists in thinking there aren't any beliefs.

Some might think to finesse this problem by denying that there are any such things as beliefs, in the same spirit that 'adverbialists' with respect to sense data deny that there are sense data. On this way of thinking, there are people, and people believe propositions—e.g., that all men are mortal. It is not the case, however, that there are any such objects or entities as *beliefs*. When, as we say, I believe that all men are mortal, that is not to be thought of as involving *two* things—me and a belief—but only one thing: a person, who is behaving in a certain way or displaying a certain property (the property of believing that all men are mortal). The argument I give below can easily be recast so as to take account of this possibility.

The difficulty I have in mind is not a recent invention. You can find it in Plato, but Leibniz offers a famous and particularly forceful statement of it:

17. It must be confessed, moreover, that *perception*, and that which depends on it, are *inexplicable by mechanical causes*, that is by figures and motions. And supposing there were a machine so constructed as to think, feel and have perception, we could conceive of it as enlarged and yet preserving the same proportions, so that we might enter it as into a mill. And this granted, we should only find on visiting it, pieces which push one against another, but never anything by which to explain a perception. This must be sought for, therefore, in the simple substance and not in the composite or in the machine.⁶⁹

Now Leibniz uses the word 'perception' here; he's really thinking of mental life generally. His point, in this passage, is that thinking, mental life generally, cannot arise by way of the mechanical interaction of parts. Consider a bicycle (or, as Leibniz says, a mill): it does what it does by virtue of the mechanical interaction of its parts. Pushing down on the pedals causes the sprocket to

⁶⁸ For an example of eliminativism, see, e.g., Paul Churchland, "Eliminative Materialism and the Propositional Attitudes," in *Contemporary Materialism*, ed. Paul K. Moser and J.D. Trout (London: Routledge, 1995), p. 151.

⁶⁹ *Monadology* 17. There are many translations of the *Monadology*.

which they are attached to turn, which causes the chain to move, which causes the sprocket attached to the back wheel to turn, which causes the back wheel to rotate. By virtue of these mechanical interactions, the bicycle does what it does, i.e., moves from one place to another upon someone's pedaling it. And of course machines generally—jet aircraft, refrigerators, computers, centrifuges—do their things and accomplish their functions in the same way. So Leibniz's claim, here, is that thinking can't arise in this way. A thing can't think by virtue of the mechanical interaction of its parts.

Leibniz is thinking of mechanical interactions as interactions involving pushes and pulls, gears and pulleys, chains and sprockets. But I think he would say the same of other interactions studied in physics, those involving, for example, gravity, electricity, magnetism, the forces holding the nucleus of an atom together, and the like. Call these physical interactions. Leibniz's claim is that thinking can't arise by virtue of physical interaction among objects or parts of objects. According to current science, electrons and quarks are simple, without parts.⁷⁰ Presumably neither can think—neither can believe, doubt, want, fear, or feel pain. But then a proton composed of quarks won't be able to think either, at least by way of physical relations between its component quarks, and the same will go for an atom composed of protons and electrons, a molecule composed of atoms, a cell composed of molecules, and an organ (e.g., a brain) composed of cells. If electrons and quarks can't think, we won't find anything composed of them that *can* think by way of the physical interaction of its parts.

Leibniz is talking about thinking generally; suppose we narrow our focus to *belief*. Recall (above, p. 33) that from the viewpoint of materialism, a belief would be a neurophysiological event or structure of some kind, a structure or event involving many neurons connected to each other in various ways, with inputs and outputs from other parts of the nervous system. Furthermore, as we also saw, a belief will have at least two kinds of properties: on the one hand, there will be NP properties; on the other, there will be a property of a different kind: the property of having a certain content. Every belief is the belief that p for some proposition p : that proposition p is then the content of the belief. Thus the content of the belief that Proust is more subtle than L'Amour is the proposition *Proust is more subtle than L'Amour*; the content of the belief that $7 + 5 = 12$ is the proposition $7 + 5 = 12$.

And now the difficulty for materialism is this: how does it happen, how can it be, that an assemblage of neurons, a group of material objects firing away *has a content*? How can that happen? More poignantly, *what is it* for such an event to have a content? What is it for this structured group of neurons, or the

⁷⁰ Although there are speculative suggestions that quarks may in fact be composed of strings.

event of which they are a part, to be related to the proposition *Cleveland is a beautiful city* in such a way that the latter is its content? A single neuron (or quark, electron, atom, or whatever) presumably isn't a belief; but how can belief, content, arise from physical interaction among such material entities as neurons? How can such physical interaction bring it about that a group of neurons has content? We can examine this neuronal event as carefully as we please; we can measure the number of neurons it contains, their connections, their rates of fire, the strength of the electrical impulses involved, and the potential across the synapses, with as much precision as you could possibly desire; we can consider its electro-chemical, NP properties in the most exquisite detail; but nowhere, here, will we find so much as a hint of content. Indeed, none of this seems even vaguely *relevant* to its having content. None of this so much as slyly suggests that this bunch of neurons firing away is the belief that Proust is more subtle than Louis L'Amour, as opposed, e.g., to the belief Louis L'Amour is the most widely published author from Jamestown, North Dakota. Indeed, nothing we find here will so much as slyly suggest that it has a content of any sort. Nothing here will so much as slyly suggest that it is *about* something, in the way a belief about horses is about horses.

The fact is, we can't see how it *could* have a content. It's not that we see or know this is perfectly possible, but we just don't know how it's done. When light strikes photo-receptor cells in the retina, there is a complex cascade of electrical activity, resulting in an electrical signal to the brain. I have no idea how all that works; but of course I know it happens all the time. But the case under consideration is different. Here it's not merely that I don't know how physical interaction among neurons brings it about that an assemblage of neurons has content and is a belief. No, in this case, we can't see how such an event *could* have content—that is, it seems upon reflection that it could *not* have content. It's a little like trying to understand what it would be for the number seven, e.g., to weigh five pounds (or for an elephant to be a proposition). We can't see how that could happen; more exactly, we can see that it *couldn't* happen. A number just isn't the sort of thing that can have weight; there is no way in which that number or any other number could weigh anything at all. (The same goes for elephants and propositions.) Similarly, we can see, I think, that physical activity among neurons can't generate content. These neurons are clicking away, sending electrical impulses hither and yon. But what has this to do with content? How is content or aboutness supposed to arise from this neuronal activity? How can such a thing be a belief? You might as well say that thought arises from the activity of the wind or the waves. But then no neuronal event can as such have a content, can be *about* something, in the way in which my belief that the number seven is prime is about the number seven, or my belief that the oak tree in my backyard is without leaves is about that oak tree.

Here someone might object as follows. "You say we can't see how a neuronal event can have content; but in fact we understand this perfectly well, and

something similar happens all the time. For there is, after all, the computer analogy. A computer, of course, is a material object, an assemblage of wires, switches, relays, a hard disk, a keyboard, and the like. I can type in a sentence or indeed an entire document; in fact I *am* typing in an entire document. Now take any particular sentence in the document: say the sentence 'Naturalism is all the rage these days'. That sentence is represented and stored on the computer's hard disk. We don't have to know in exactly what *way* it's stored (it's pluses and minuses, or a magnetic configuration, or something else; it doesn't matter). Now the sentence 'Naturalism is all the rage these days' *expresses* the proposition *Naturalism is all the rage these days* (as does the German sentence 'Der Naturalismus ist diese Tage ganz gross in Mode' or any other sentence synonymous with this one). That sentence, therefore, has the proposition *Naturalism is all the rage these days* as its content. But then consider the analogue of that sentence on the computer disk: clearly it, too, expresses the same proposition as the sentence it represents. That bit of the computer disk, therefore, has propositional content. But of course that bit of the computer disk is also a material object (as is any inscription of the sentence in question). Contrary to your claim, therefore, a material object can perfectly well have propositional content; indeed, it happens all the time. But if a computer disk or an inscription of a sentence can have a proposition as content, why can't an assemblage of neurons? Just as a magnetic pattern has as content the proposition *Naturalism is all the rage these days*, so, too, a pattern of neuronal firing can have that proposition as content. Your claim to the contrary is completely bogus." Thus far the objector.

Well, if the sentence or the computer disk really did have content, then I guess the assemblage of neurons could too. But the fact is that neither does—or, rather, neither has the right kind of content: neither has *original* content. For how does it happen that the sentence has content? It's simply by virtue of the fact that we human beings *use* the sentence in a certain way, a way such that if a sentence is used in that way, then it expresses a certain proposition. Upon hearing that sentence, I think of, grasp, apprehend the proposition *Naturalism is all the rage these days* (and of course the same goes for the German sentence and speakers of German). You can get me to grasp, entertain, and perhaps believe that proposition by uttering that sentence. How exactly all this works is complicated and not at all well understood; but the point is that the sentence has content only because of something *we*, we who are already thinkers, do with it. We could put this by saying that the sentence has *secondary* or *derived* content; it has content only because we, we creatures whose thoughts and beliefs already have content, treat it in a certain way. The same goes for the magnetic pattern on the computer disk: it represents or expresses that proposition because we assign that proposition to that configuration. But of course that isn't how it goes (given materialism) with that pattern of neural firing. That pattern doesn't get its content by way of being used in a certain way by some other creatures whose thoughts and beliefs already have content. If that pattern has content at all, then, according to materialism,

it must have *original* or *primary* content. And what it is hard or impossible to see is how it could be that an assemblage of neurons (or a sentence, or a computer disk) could have original or primary content. To repeat: it isn't just that we can't see how it's done, in the way in which we can't see how the sleight-of-hand artist gets the pea to wind up under the middle shell. It is rather that we can see, to at least some degree, that it can't be done, just as we can see that an elephant can't be the number 7, and that the number 7 can't weigh seven pounds.

Peter van Inwagen agrees that it is indeed hard to see how physical interaction among material entities can produce thought: "... it seems to me that the notion of a physical thing that thinks is a mysterious notion, and that Leibniz's thought-experiment brings out this mystery very effectively."⁷¹ Now I am taking this fact as a reason to reject materialism, the idea that human beings are physical or material objects with no immaterial parts. I'm taking it as a reason for thinking materialism is false. But if materialism is false, immaterialism must be true; if a material object can't think, then whatever thinks must be an immaterial object. Hence a human being is really an immaterial object (or at least has an immaterial part or element). The simplest view here is substance dualism; this is the view that a human being is an immaterial object, a thing that can think, joined in a special way to a material body. I am an immaterial substance standing in a peculiarly intimate relation to a certain material thing, the thing I call my body. And the fact that it is hard to see how a material object can think (so I say) is a serious difficulty for materialism. It's an argument for substance dualism—but only, of course, if there is no similar difficulty for substance dualism itself.

Van Inwagen thinks there *is* a similar difficulty for dualism:

For it is thinking itself that is the source of the mystery of a thinking physical thing. The notion of a non-physical thing that thinks is, I would argue, equally mysterious. How any sort of thing could think is a mystery. It is just that it is a bit easier to see that thinking is a mystery when we suppose that the thing that does the thinking is physical, for we can form mental images of the operations of a physical thing and we can see that the physical interactions represented in these images—the only interactions that *can* be represented in these images—have no connection with thought or sensation, or none we are able to imagine, conceive or articulate. The only reason we do not readily find the notion of a non-physical thing that thinks equally mysterious is that we have no clear procedure for forming mental images of non-physical things. (p. 176)

So dualism is no better off than materialism; they both have the same problem. What is this problem, according to van Inwagen? The problem for materialism is that we can't *imagine* a material thing thinking; we can't form a mental

⁷¹ *Metaphysics* (Boulder, CO: Westview Press, 2002 (second edition)), p. 176. Hereafter page references given in the text.

image of a material thing thinking. But the same goes, says van Inwagen, for an immaterial thing: we also can't form a mental image of an immaterial thing thinking. Indeed, we can't form a mental image of any kind of thinking thing: "My point," he says, "is that nothing could possibly count as a mental image of a thinking thing" (p. 177). But then materialism and dualism are so far on a par; there is nothing here to incline us to dualism rather than materialism.

Thus far van Inwagen; but is he right? The thought of a physical thing thinking, he concedes, is mysterious; that is because we can't form a mental image of a physical thing thinking. But this seems to me to mislocate the problem. It is not just that we can't form a mental image of a physical thing thinking that inclines us to reject the idea. There are plenty of things of which we can't form a mental image, where we're not at all inclined to reject them. I can't form a mental image of the proposition *Proust is more subtle than L'Amour* or *Naturalism is all the rage these days*. I can't form a mental image of either of these propositions' being true (or being false). But I'm not in the least inclined on that account to reject the idea that the first, say, is in fact true. As Descartes pointed out, I can't form a mental image of a 1000-sided rectilinear plane figure (or at least an image that distinguishes it from a 100-sided rectilinear plane figure); that doesn't suggest that there can't be any such thing. I can't form a mental image of the number 79's being prime; that doesn't incline me to believe that it isn't prime. I don't believe that a proposition or a set could be red; but it's not because I can't form a mental image of a proposition's (or a set's) being red.

Well, what *is* it, then, that inclines me to think a proposition can't be red, or a horse be an even number? The answer, I think, is that one can just see upon reflection that these things are impossible. I can't form a mental image of a proposition's having members; but that's not why I think no proposition has members, because I also can't form a mental image of a *set's* having members. It's rather that one sees that a set is the sort of thing that has or can have members, and a proposition is not. It is the same with a physical thing's thinking. True, one can't imagine it. The reason for rejecting the idea, however, is not that one can't imagine it. It's rather that one can see that a physical object just can't do that sort of thing. This isn't as clear, perhaps, as that a proposition can't be red; some impossibilities are more clearly impossible than others. But one can see it to at least some degree.⁷² And the same *doesn't* go for an immaterial

⁷²Van Inwagen might be prepared to concede this; he says: ". . . Leibniz's thought experiment shows that when we carefully examine the idea of a material thing having sensuous properties, it seems to be an impossible idea." "Dualism and Materialism: Athens and Jerusalem?," *Faith and Philosophy* 12:4 (Oct. 1995), p. 478. That is (I take it), it seems to be *necessary* that material things don't have such properties. Van Inwagen's examples are such properties as *being in pain* and *'sensing redly'*; the same goes, I say, for properties like *being the belief that p* for a proposition *p*.

thing's thinking; we certainly can't see that no immaterial thing can think. (If we could, we'd have an argument against the existence of God: no immaterial thing can think; if there were such a person as God, he would be both immaterial and a thinker; therefore . . .)

Van Inwagen has a second suggestion as to why it's hard to conceive of a thinking thing:

In general, to attempt to explain how an underlying reality generates some phenomenon is to construct a representation of the working of that underlying reality, a representation that in some sense "shows how" the underlying reality generates the phenomenon. Essentially the same considerations as those that show that we are unable to form a mental image that displays the generation of thought and sensation by the workings of some underlying reality (whether the underlying reality involves one thing or many, and whether the things it involves are physical or non-physical) show that we are unable to form *any* sort of representation that displays the generation of thought and sensation by the workings of an underlying reality. (pp. 177–8)

The suggestion is that we can't form an image or other representation displaying the generation of thought by way of *the workings of an underlying reality*; hence we can't see how it can be generated by physical interaction among material objects such as neurons. This much seems right. Van Inwagen goes on to say, however, that this doesn't favor dualism over materialism, because we *also* can't see how thought can be generated by the workings of an underlying *non-physical* reality. And perhaps this is also right. But here there is an important dissimilarity between dualism and materialism. The materialist thinks of thought as generated by the physical interaction of such things as neurons; the dualist, however, typically thinks of an immaterial self, a soul, a thing that thinks, as *simple*. An immaterial self doesn't have any parts; hence, of course, thought isn't generated by the interaction of its parts. Say that a property *P* is *basic* to a thing *x* if *x* has *P*, but *x*'s having *P* is not generated by the interaction of its parts. Thought is then a basic property of selves, or, better, a basic *activity* of selves. It is also an *immediate* activity of selves, in that a self doesn't think by way of doing something else (in the way, for example, that the referee signals a touchdown by raising his arms). A self doesn't think by way of doing something else. It's not that (for example) there are various immaterial parts of a self whose interaction produces thought; nor is it that a self thinks by doing something else. Of course a self stands in causal relations to its body: retinal stimulation causes a certain sort of brain activity which in turn causes a certain kind of experience in the self. But there isn't any *way* in which the self produces a thought; it does so immediately. To ask 'How does a self produce thought?' is to ask an improper question. There isn't any *how* about it.

An analogy: consider the lowly electron. According to current science, electrons are simple, not composed of other things. An electron has basic properties such as spin and a negative charge. But then the question ‘How does an electron manage to have a charge?’ is an improper question. There’s no *how* to it; it doesn’t do something else that results in its having such a charge, and it doesn’t have parts by virtue of whose interaction it has such a charge. Its having a negative charge is rather a basic and immediate property of the thing. The same is true of a self and thinking: it’s not done by underlying activity or workings; it’s a basic and immediate activity of the self. But then the important difference, here, between materialism and immaterialism is that if a material thing managed to think, it would have to be by way of the activity of its parts; and we can’t see how that could happen (it seems upon reflection that it can’t happen). Not so for an immaterial self. Its activity of thinking is basic and immediate. And it’s not the case that we are inclined upon reflection to think this can’t happen—there’s nothing at all against it, just as there is nothing against an electron’s having a negative charge, not by virtue of the interaction of parts, but in that basic and immediate way. The fact of the matter, then, is that we can’t see how a material object can think—that is, upon reflection it seems to at least some degree that a material object *can’t* think. Not so for an immaterial self.

True, as van Inwagen says, thought can sometimes seem mysterious and wonderful, something at which to marvel. (Although from another point of view thought is more familiar than hands and feet). But there is nothing here to suggest that it can’t be done. Part of the mystery of thought is that it is wholly unlike things that are done by material objects; but of course that’s not to suggest that it can’t be done at all. Propositions are also mysterious and have wonderful properties: they manage to be about things; they are true or false; they can be believed; they stand in logical relations to each other. How do they manage to do those things? Well, certainly not by way of interaction among material parts. Sets manage, somehow, to have members—how do they do a thing like that? And why is it that a given set has just the members it has? How does the unit set of Socrates manage to have just him as a member? Why can’t I be a member of it? What mysterious force keeps me out of it? Well, it’s just the nature of sets to be like this. These properties can’t be explained by way of physical interactions among material parts, but that’s nothing at all against sets. Indeed, these properties can’t be *explained* at all. Of course if you *began* with the idea that everything has to be a material object, then thought (and propositions and sets) would indeed be mysterious and paradoxical. But why begin with that idea? Thought is seriously mysterious, I think, only when we assume that it would have to be generated in some physical way, by physical interaction among physical objects. *That* is certainly mysterious; indeed it goes far beyond mystery, all the way to apparent impossibility. But that’s not a problem for thought; it’s a problem for materialism.

Now of course this problem has not been lost on materialists, canny lot that they are. Their attempts to deal with the problem ordinarily take the form of suggestions as to how it might be that a neural object or event could have (original) content after all. Nearly all attempts to do so begin with what we might call *indicators*, or *indication*, or *indicator meaning*.⁷³ 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. We could speak here of 'natural signs': smoke is a natural sign of fire and the height of the mercury column signifies the ambient temperature. When one thing indicates or is a natural sign of another, there is ordinarily some sort of causal or nomic connection between them by virtue of which the first is reliably correlated with the second. Smoke causes fire, which is why it indicates fire; measles cause red spots on your face, which is why red spots on your face indicate measles; there is a causal connection between the height of the mercury column and the temperature.

The nervous systems of organisms often contain such indicators. A widely discussed example: whenever a frog sees a fly zooming by, there is (so we think) a certain pattern of neural firing in its brain; as a result these neurons, or patterns of firing, are sometimes called 'fly detectors.' Another famous example: some anaerobic marine bacteria have little internal magnets called 'magnetosomes.' These function like compass needles, indicating magnetic north. The direction to magnetic north (in the northern hemisphere) is downward; hence these bacteria, which can't flourish in the oxygen-rich surface water, move towards the more oxygen-free water at the bottom of the ocean. There are also such structures in human bodies. There are structures that respond in a regular way to blood pressure and temperature, to the amount of sugar in the blood, to its sodium content, to light of a certain pattern striking the retina, and the like. 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, there is a distinctive pattern of neural firing (or some other kind of structure) in your brain that is correlated with and caused by it.

The next step is to call these structures, the ones correlated with external or internal conditions of one kind or another, 'representations.' Indeed, the idea that such structures are representations has become so common that it is part of the current background assumptions in cognitive neuroscience. Those patterns of neural firing in the frog's brain are said to be representations of flies, or bugs (or small flying objects); those magnetosomes in anaerobic bacteria are said to represent north, or the direction towards oxygen-free water, or the lines of the earth's magnetic field (there is usually considerable latitude of choice

⁷³ See Fred Dretske's *Explaining Behavior* (Cambridge, MA: MIT Press, 1988), pp. 54ff.

as to what gets represented); the structures in your body that respond to the temperature of your blood are said to represent that temperature.

Now the terms 'represent' and 'representation' are multiply ambiguous. Webster's *Third International* gives a whole host of analogically connected meanings: you can send your representative to a meeting; your state or national representative represents your interests (we hope); an artist can produce a representation of a battle; a musical passage can represent a storm; x's and o's can represent football players and a dotted line can represent where the tight end is supposed to go; a scale model of Mt. Rainier can represent Mt. Rainier. This term is therefore something of a weasel word, and in typical philosophy of mind or cognitive science contexts it is used without definition. As a result, it is often hard to know just what is meant by calling those indicators 'representations'; shall we say that wherever you have causal or nomological correlation, you have representation? Shall we say that smoke represents fire (and fire represents smoke), that the rate at which the wheels of my car turn represent the speedometer reading, and that trees budding represent spring or warmer weather (and vice versa)? Well, I guess we can say these things if we like; it's a free country, and the term 'representation' is flexible enough to allow it.

But here the crucial next step: efforts to understand belief materialistically typically try, somehow, to promote these representations to *beliefs*. In so doing, they don't ordinarily try to solve Leibniz's problem—the fact that it looks as if a material thing can't think, or be a belief; they simply ignore it. But this procedure is also unpromising in its own right: representation of this sort is nowhere near sufficient for belief. The gas gauge on my car may represent the amount of gas (or the weight on the bolts holding the tank to the frame), or the volume of air in the tank, and other things as well; nothing in the neighborhood has beliefs on these scores. The thermostat may represent the temperature; but when the temperature drops and the thermostat starts the furnace, it doesn't believe that it's too cool in here (and neither does the furnace or anything else in the relevant neighborhood). Those magnetosomes perhaps represent the direction to oxygen-free water; neither they nor the bacteria that contain them believe that's the way to oxygen-free water. Certain internal structures indicate and thus represent your blood pressure; these structures don't believe that your blood pressure is thus and so, and neither (most of the time) do you. The thing to see is that no amount of this indication and representation, no matter how gussied up, is sufficient for *belief*. Clearly a material object *can* be a representation in *some* sense: Michelangelo's *David* for example, is a representation of David, and a few weird lines in a cartoon can represent Ted Kennedy. But it doesn't follow that a material structure can be a belief, or that it can have propositional content (original content). And I think we can see that it can't.

There are basically three ways in which materialist thinkers try to promote indicators to belief. First, there is the Millikan/Dennett proposal: an indicator

gets to be a belief when evolution confers on it the *function* of causing a certain sort of behavior. Second, there is Jerry Fodor's suggestion. It's natural to think there are certain brain structures that indicate cows; when I see a cow, presumably there will be in my brain a structure that is correlated with cows. But this structure can also be caused by other things—a moose in the twilight, for example, or maybe a very large cat, perhaps after I've had too many martinis. According to Fodor, what confers content on such a structure—the content *cow*—is that there being structures of that sort that are *not* caused by cows is asymmetrically dependent upon there being structures of that sort that *are* caused by cows: “But ‘cow’ means *cow* and not *cat* or *cow* or *cat* because *there being cat-caused ‘cow’ tokens depends on there being cow-caused ‘cow’ tokens, but not the other way around.*”⁷⁴ Third, there is Fred Dretske's work, perhaps the most sophisticated and accomplished attempt to explain belief from a materialist perspective.⁷⁵ I don't have the space to look into all of these; a brief examination of Dretske's efforts will have to suffice. But note that all three lines of approach ignore Leibniz's problem. All three simply assume that it is possible for a material thing to think and for a material assemblage of neurons to be a belief.

Dretske begins (as does nearly everyone undertaking this enterprise) with the notion of indication, correlation (perhaps nomic, perhaps causal) between events of one kind and events of another. His attempt to explain belief in terms of indication involves two additional ideas. First is the notion of *function*. All 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.”⁷⁶ So not all cases of indication are cases of representation: the fuel gauge in my automobile indicates the amount of gasoline in the tank, the weight 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 register the amount of gasoline in the tank. Hence it represents the amount of fuel in the tank and does not represent those other properties and quantities, interesting as they may be. This appeal to function enables Dretske to see representational contexts as like belief contexts in being intentional: it may be that it is the function of something or other to indicate a property *p*, while it isn't its function to indicate a nomically or logically equivalent property *q*.

⁷⁴ *A Theory of Content and Other Essays* (Cambridge, MA: MIT Press, 1990), p. 91 (original emphasis).

⁷⁵ See in particular *Explaining Behavior and Naturalizing the Mind* (Cambridge: MIT Press, 1995).

⁷⁶ *Naturalizing the Mind*, p. 2.

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 case of the noctuid moth, which, upon detecting the bursts of high-frequency sound emitted by the bat's radar, executes evasive maneuvers. Here we have representation; it is the function of those neural structures N 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".⁷⁷ Take a given moth and the neural circuit C whose firing causes those maneuvers M : the explanation of C 's causing M is not that C indicates the presence of bats, but the way the neural circuitry of this moth is deployed. The fact that in these moths, C represents the presence of bats may explain or help explain why moths of this type have survived and flourished; but the fact that in a given moth C represents bats does not explain why C causes M .

If we don't get belief here, where do we get it? Where there is *learning*, says Dretske. 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 a shadow on the walls of its cage. But then we reward it when it pecks at the red spot. Soon it 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, says Dretske, we do have 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.

As far as I can see, therefore, Dretske's complete account of belief can be put as follows:

- (D) x is a belief 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) whose function it is to indicate something F , (3) x is in the mode or state it is in when it indicates something F , (4) x causes some movement M , and (5) the explanation of x 's causing M is that it indicates F .

⁷⁷*Explaining Behavior*, p. 92.

A comment on (3): it's not necessary that on the occasion in question, x is actually indicating something F perhaps on this occasion x is misrepresenting. We fix red-colored spectacles on the bird: now its red spot indicator causes it to peck at any spot, red or not. But the red spot indicator is still on, as we might say, even when in fact the spot in front of the bird is black.

This is a complex and sophisticated account; some of its complexity can be accounted for in terms of the failure of earlier accounts. For example, someone might say, perhaps with Dennis Stampe, that a belief is any element of an indicator system that is indicating F , or perhaps with Millikan, that x is a belief if x is an element of an indicator system whose function it is to cause some adaptive motion M . These are clearly insufficient, as is shown by such internal indicator systems as those that register blood pressure, blood temperature, and sodium and sugar levels in blood. Here we have elements of indicator systems indicating something F , and having the function of causing some adaptive motion M (adjusting blood temperature, sodium or sugar level, etc.); but nothing in the neighborhood has the relevant beliefs.

Still, sophisticated as it is, Dretske's account, I think, won't anywhere nearly do the job. I've already argued that the notion of proper function can't be accommodated by naturalism and also that a material structure *can't* acquire or have content; for present purposes let's waive these more general objections and consider some that are a bit more specific. First, a couple of semi-technical objections. I believe that $7 + 5 = 12$; nothing, however, carries the information that $7 + 5 = 12$, and indeed $7 + 5$'s *being equal to 12* isn't information. That is because, according to Dretske's (Shannon) conception of information, information is always a matter of reduction of possibilities; but $7 + 5$'s *equaling 12* doesn't reduce the possibilities with respect to anything. The account is therefore too strong; it rules out beliefs that are logically necessary in either the broad or the narrow sense. And just what kind of possibilities are we thinking of here? If causal or nomic possibilities are relevant, then the account also fails to work for nomologically necessary beliefs, such as that (as current physics has it, anyway) nothing travels faster than light (more exactly, nothing accelerates from a velocity less than that of light to a velocity greater than that of light). This doesn't reduce the nomic possibilities. And what about beliefs about the past? Given that past propositions are 'accidentally necessary,' nothing *now* carries the information that Brutus stabbed Caesar (in Dretske's technical Shannon sense—obviously some textbooks carry that information in the ordinary sense).

Further, I believe that Proust is more subtle than L'Amour; is it even remotely plausible to suppose that I must therefore have a Proust-is-more-subtle-than-L'Amour-indicator, a neural structure whose function it is to indicate that Proust is more subtle than L'Amour? Or a structure that fires when one person is a more subtle writer than another? And even if there were such structures,

would they have to cause *motion* of one sort or another for me to believe that Proust is more subtle than L'Amour? Maybe I've always believed this, but never said so, or in any other way displayed this belief in my behavior.

Still further, return to that noctuid moth. Perhaps it was designed by God; and perhaps God designed it in such a way that *C*, the structure causing that evasive motion, causes that motion because *C* indicates the presence of bats. Then it would be true that *C* causes *M* because of what it indicates, and on Dretske's account, the moth would on the appropriate occasions believe that there are bats present. So if the moth came to be by undirected evolution it doesn't have beliefs (or at least doesn't have the belief that bats are present when its bat indicator is activated); if God has designed it, however, then it does have that belief on those occasions. Can that be right? In the same way there are all those internal indicators I mentioned a bit ago: structures whose function it is to indicate blood pressure, temperature, sodium level, sugar level, and the like. These indicators are in fact so constituted that they cause certain kinds of movements. If human beings have been designed by God, then presumably they cause those movements because of what they indicate; that's why God designed the system in such a way that they *do* cause those movements. But then on Dretske's account, these structures, or we who contain them, would hold the associated beliefs about our blood temperature, pressure, sodium level, sugar level, and the like. But we don't; if Dretske's account were right, therefore, this would constitute an argument against the existence of God. Clearly it doesn't.

Insofar as they can't accommodate necessary beliefs and beliefs about the past, Dretske's conditions are too strong: they aren't necessary for belief. But they are also too weak: they aren't sufficient either. If his account were correct, then if we have been designed by God, we hold all those beliefs about blood pressure, temperature, sodium content, and the like; but we don't. You may or may not think we have in fact been designed by God or anyone else; but even if we haven't it is certainly possible that we have; hence it's possible that Dretske's conditions hold when no beliefs are present. And really, why should the fulfillment of Dretske's conditions have anything at all to do with belief? So there is a structure that has the function of indicating something and causes what it does because of what it indicates; does that really so much as slyly suggest that something in the neighborhood of this structure holds the appropriate belief, or any belief at all? Consider again the lowly thermostat. The bimetallic strip indicates the temperature, and has the function of indicating it. Further, when it bends enough to close the circuit, thereby causing furnace ignition, it causes what it causes because of what it indicates; we designed the thermostat in such a way that when that strip indicates 67°F, it causes the furnace to ignite. The explanation of its causing that movement is that it is indicating that the temperature is 67°F. But neither the bimetallic strip nor the thermostat, nor the

furnace nor anything else need believe that the temperature is 67°F. Dretske's account, therefore, won't anywhere nearly serve as an explanation of how there could be beliefs if materialism about human beings is true.

That's the third problem for naturalism, construed so as to include materialism: if it were true, there would be no such thing as belief content, no such thing as primary intentionality, and no such thing as belief. There are, of course, various responses to this problem. As I said above, several materialists hold that there simply aren't any such things as beliefs; the fact is no one ever believes a proposition. These ideas—belief, content, intentionality, aboutness, and the like—belong to the infancy of our race. They belong to folk psychology, a way of thinking that is now outmoded, even if we all rely upon it. We may expect, so they tell us, that the categories of folk psychology will be replaced by more adequate ways of thought, categories, and concepts coming from science. Just as we no longer believe that the earth is flat, just as we no longer believe that the stars are slits in a giant canvas stretched over the earth every night to give us a good night's sleep, so (so the claim goes) we will at some point no longer believe in beliefs, desires, aboutness, and the like. But then the fact that naturalism has no room for belief is really nothing against it. This response seems to me the strongest a naturalist can muster. Of course it does have one real problem: it seems utterly crazy to think that people never hold beliefs.

Conclusion

What we've seen, so far, is that naturalism cannot accommodate proper function and the things that go with it: health, illness, flourishing, pathology, and the like. If naturalism is true, neither people nor other living things nor their systems or organs function properly (or improperly). Most naturalists are also materialists about human beings; if both naturalism and materialism are true, so I say, there aren't any such things as beliefs. Finally, we've also seen that naturalism is self-defeating; a reflective naturalist has a rationality defeater for naturalism itself, and is thus irrational in believing naturalism. Indeed, the reflective naturalist has a defeater for anything he believes and is thus thrown into that profound, many-layered, reflexive skepticism both feared and endorsed by David Hume.

So what's a naturalist to do: what options are available to him, once he recognizes these consequences? I see three basic possibilities. The first is a kind of fictionalism. Maybe there really aren't any desires or beliefs or other intentional phenomena (or alternatively the question whether there are has no answer); these things are mere fictions. Still, fiction has its uses. You can adopt the *intentional stance* towards (alleged) other persons and some machines:

you can treat them and think of them as if they really did possess desire and belief.⁷⁸ You can make with respect to them the predictions and predications you would make if you thought they really did possess those properties. If you do so, your predictions about their behavior and responses to what you do will be substantially enhanced and will go more easily and smoothly.⁷⁹

There is no doubt that fiction can sometimes enable us to achieve a level of understanding and control not otherwise available. In physics we think about frictionless planes, point particles, true vacuums. Perhaps there aren't any such things; even so, we can learn much by thinking about them. (Thinking about these fictions is also of practical benefit: we use them in the design and construction of space shuttles, linear accelerators, Olympic bobsleds, and so on.) Of course it requires a certain sophistication to see how fiction can help us gain genuine understanding; but the thought that it can goes back at least to Hobbes and Locke, with their fictional notion of an aboriginal contract signed and sealed by our remote ancestors hoping for relief from their miserable lives in the state of nature. The prime modern sources of this notion of useful fictions are perhaps Leibniz and Kant. Speaking of something like useful fiction with respect to some of the very ideas under consideration (purpose, goal, design plan, proper function and their colleagues) Kant has this to say:

... an object, or state of mind, or even an action is called purposive, although its possibility does not necessarily presuppose the representation of a purpose, merely because its possibility can be explained and conceived by us only so far as we assume for its ground a causality according to purposes, i.e. in accordance with a will which has regulated it according to the representation of a certain rule.⁸⁰

Kant's idea is that there are natural phenomena of which we can gain proper understanding only by way of such notions as purpose and function—despite the fact that nature itself can't properly be seen as displaying (or even covertly harboring) purpose or function. And perhaps the naturalist can follow Kant, adopting an intentional stance with respect to belief and also with respect to proper function and its colleagues.

Now these anti-realist stances are refined and highly sophisticated—in fact, a bit contorted. And, as any rock climber knows, unnatural stances become

⁷⁸ See Daniel Dennett's *Consciousness Explained* p. 76

⁷⁹ Another reason for adopting the intentional stance: perhaps for reasons of your own you want to preserve verbal agreement with those with whom you really disagree; you may want to speak with the vulgar but think with the learned.

⁸⁰ *Critique of Judgment*, tr. with an Introduction by J. H. Bernard (New York: Hafner Press, 1951), 55/6, pp. 54–5.

awkward and uncomfortable if held for any length of time. If in one way a fiction can help you understand a phenomenon, in another it can harm your understanding of it. You think the fact is there is no such thing as belief (or proper function); you find yourself nonetheless ineluctably compelled, in your non-philosophical life, to adopt a stance presupposing that there *are* such things. Of course no one really *adopts* such a stance, any more than you adopt your parents; we all take it utterly for granted from earliest consciousness that others have desires and beliefs. Similarly, we all take it for granted that there is such a thing as proper function for the heart, or kidney, or lung. More poignantly, there is such a thing as *mal* function for these things. The fictionalist stance is awkward: to adopt it you are to think that George's heart isn't *really* malfunctioning (or that there isn't any truth of the matter as to whether it is), but you are to treat it and think about it, somehow, as if it *were* malfunctioning. Can you really avoid doublethink and false consciousness? Alternatively, can you avoid what from your own perspective is illusion and error? Illusion, as Freud and Marx tell us, has its uses; but helping to achieve straightforward understanding is not among them.

This line of thought, therefore, is not really attractive. We can turn again to contemporary philosophy of mind for a second alternative. This is for the naturalist to follow the eliminativists in philosophy of mind, who say the same about belief, desire, hope, acceptance, and the other mental states recognized by what they call (in a disparaging tone of voice) 'folk psychology.'⁸¹ We ordinary folk organize our entire lives around the idea that people, ourselves included, believe some propositions, withhold others, and disbelieve still others. We all believe that people, ourselves included, desire some outcomes and hope to avoid others. The eliminativist, however, regards these notions—belief, desire, etc.—as part of an outmoded theory. A proper science of mind will have no place for them (or, presumably, for the notion of mind itself). The naturalist can follow suit, display the courage of his naturalistic convictions, and stoutly declare that there are no such things as belief and/or proper function. These declarations are simply the costs exacted by naturalism; if we want to be naturalists, we will have to pay the price.⁸²

But here we meet a very natural question: why should we *want* to be naturalists? If it requires giving up all these things, presumably we should weigh

⁸¹ See, e.g., Churchland, "Eliminative Materialism and the Propositional Attitudes," p. 151.

⁸² What about the Humean skepticism to which I said the naturalist falls prey? The eliminativist naturalist can perhaps reply that my argument is stated in terms of the categories of belief and defeaters; if, as he proposes, there simply aren't any such things as beliefs, my argument is of doubtful relevance. Of course the naturalist, even the eliminativist naturalist, will presumably still need something *like* the categories of belief and defeater; he'll still want to propose or advance or endorse certain propositions and reject others; he'll still hold that there are right and wrong

the cost of accepting naturalism against its proposed benefits. On the one side of the scale are the arguments for naturalism; on the other, our reasons for the ordinary beliefs we must give up if we wish to be naturalists. And it will be sensible to adopt naturalism, clearly enough, only if the arguments for it are stronger than the reasons for those ordinary beliefs. But where *are* the arguments for naturalism? Perhaps it would be sensible to give up all those ordinary ways of thinking if there were powerful arguments for naturalism. But where are those powerful arguments? As far as I can see, there aren't even any *decent* arguments, let alone powerful arguments, for naturalism. So I suggest a third possibility: give up naturalism, and perhaps accept instead some form of theism.⁸³

(or useful and fruitless) ways of doing this; and he'll still want to hold that one can come to endorse something such that endorsing it makes it right, or useful, or appropriate to stop endorsing something. His job will be to try to reconstruct something of the categories of folk psychology in terms of the categories he accepts; and perhaps the argument for Humean skepticism can also be reconstructed in terms of those categories.

⁸³ My thanks to Michaels Rea and Bergmann.