

## ***Chapter 10***

# **The Evolutionary Argument Against Naturalism**

My overall thesis: there is superficial conflict but deep concord between science and religion, and superficial concord but deep conflict between science and naturalism. So far I have developed the first half of this theme; it is now time to turn to the second.

### **I SUPERFICIAL CONCORD**

**I suppose it isn't really necessary to argue that there is (at least) superficial concord between naturalism and science; the high priests of naturalism trumpet this loudly enough. Naturalists pledge allegiance to science; they nail their banner to the mast of science; they wrap themselves in the**

mantle of science like a politician in the flag. They confidently claim that naturalism is part of the “scientific world-view,” and that the advent of modern science has exposed supernaturalism as a tissue of superstition—perhaps acceptable and perhaps even sensible in a prescientific age, but now superseded. A particularly charming phrase, here, is the obligatory “as we now know”; we were previously wallowing in ignorance and superstition, but now, thanks to science, we finally know the truth.<sup>1</sup>

All of this, however, is error, and whopping error at that. Naturalists don’t ordinarily explain just why they think science guarantees or supports naturalism; they are usually content just to announce the fact. And ordinarily what they announce is not that, say, quantum mechanics, or general relativity, or the periodic table of the elements has dethroned theism and supernaturalism, but that Darwin has. According to Stephen J. Gould (see above [chapter 1](#)), “Before Darwin, we thought that a benevolent God had created us”; but now, after Darwin, we realize that “No intervening spirit watches lovingly over the affairs of nature.” George Gaylord Simpson seconds the motion: Man is the result of a purposeless and natural process that did not have him in mind.<sup>2</sup>

As we saw in [chapters 1](#) and [2](#), however, this is the result of confusion—a confusion between guided and unguided evolution, between sober science and philosophical or theological add-on. Let me briefly recapitulate. The scientific theory of evolution just as such is entirely compatible with the thought that God has guided and orchestrated the course of evolution, planned and directed it, in such a way as to achieve the ends he intends. Perhaps he causes the right mutations to arise at the right time; perhaps he preserves certain populations from extinction; perhaps he is active in many other ways.<sup>3</sup> On the one hand, therefore, we have the scientific theory, and on the other, there is the claim that the course of evolution is not directed or guided or orchestrated by anyone; it displays no teleology; it is blind and unforeseeing; as Dawkins says, it has no aim or goal in its mind's eye, mainly because it has no mind's eye.

This claim, however, despite its strident proclamation, is no part of the scientific theory as such; it is instead a metaphysical or theological add-on. On the one hand there is the scientific theory; on the other, the metaphysical add-on, according to which the process is unguided. The first is part of current science, and deserves the respect properly accorded to a pillar of science; but the first is entirely compatible with theism. The

second supports naturalism, all right, but is not part of science, and does not deserve the respect properly accorded science. And the confusion of the two—confusing the scientific theory with the result of annexing that add-on to it, confusing evolution as such with unguided evolution—deserves not respect, but disdain.

The fact is, as we saw in [chapter 9](#), science fits much better with theism than with naturalism. On balance, theism is vastly more hospitable to science than naturalism, a much better home for it. Indeed, it is theism, not naturalism, that deserves to be called “the scientific worldview.”

## II DEEP CONFLICT

In this chapter I'll take this line of thought further. I'll argue that despite the superficial concord between naturalism and science—despite all the claims to the effect that science implies, or requires, or supports, or confirms, or comports well with naturalism—the fact is science and naturalism don't fit together at all well. The fact is there is deep unease, deep discord, deep conflict between naturalism and science. I'll argue that there is a deep and irremediable conflict between naturalism

and *evolution*—and hence between naturalism and science.<sup>4</sup> My quarrel is certainly not with the scientific theory of evolution. Nor is it an argument for the conclusion that unguided evolution could not produce creatures with reliable belief-producing faculties; I very much doubt that it could, but that it *couldn't* is neither a premise nor the conclusion of my argument.<sup>5</sup> Still further, my argument will not be for the conclusion that naturalism is false, although of course I believe that it is.

What I *will* argue is that naturalism is in conflict with evolution, a main pillar of contemporary science. And the conflict in question is not that they can't both be true (the conflict is not that there is a contradiction between them); it is rather that one can't sensibly accept them both. By way of analogy: I can't sensibly believe that there aren't any beliefs, or that no one has true beliefs, or that my beliefs are all false. These things are all possible, but I can't sensibly believe them. In the same way, I mean to argue that one can't sensibly believe both naturalism and the scientific theory of evolution. If my argument is cogent, it follows that there is deep and serious conflict between naturalism and evolution, and hence deep conflict between naturalism and science.

Now it is not clear that naturalism, as it stands, is

a religion; there is enough vagueness around the edges of the concept of religion for it to be unclear whether naturalism does or doesn't belong there. But naturalism does serve one of the main functions of a religion: it offers a master narrative, it answers deep and important human questions. Immanuel Kant identified three great human questions: Is there such a person as God? Do we human beings have significant freedom? And can we human beings expect life after death? Naturalism gives answers to these questions: there is no God, there is no immortality, and the case for genuine freedom is at best dicey. Naturalism tells us what reality is ultimately like, where we fit into the universe, how we are related to other creatures, and how it happens that we came to be. Naturalism is therefore in competition with the great theistic religions: even if it is not itself a religion, it plays one of the main roles of a religion. Suppose we call it a "quasi-religion." I've already argued that there is no conflict between theistic religion and science; if my argument in this chapter is right, however, there *is* profound conflict between science and a quasi-religion, namely naturalism. So the real conflict lies not between science and Christian belief (or more generally theistic religion), but between science and naturalism. If we want to focus on the fact that naturalism is a quasi-religion, the truth is that there is a science-religion conflict, all right, but it is

between science and naturalism, not science and theistic religion.

### III THE ARGUMENT

My argument will center on our *cognitive faculties*: those faculties, or powers, or processes that produce beliefs or knowledge in us. Among these faculties is *memory*, whereby we know something of our past. There is also *perception*, whereby we know something about our physical environment—for the most part our immediate environment, but also something about distant objects such as the sun, the moon, and stars. Another is what is often called “a priori *intuition*,” by virtue of which we know truths of elementary arithmetic and logic. By way of a priori intuition we also perceive deductive connections among propositions; we can see which propositions logically follow from which other propositions. In this way, starting from a few elementary axioms, we can explore the great edifices of contemporary logic and mathematics.

There are still other cognitive faculties: Thomas Reid spoke of *sympathy*, which enables us to

know the thoughts and feelings of other people, *introspection* (reflection), whereby we know about our own mental life, *testimony* whereby we can learn from others, and *induction*, whereby we can learn from experience. Many would add that there is a *moral sense*, whereby we know right from wrong; and believers in God may add that there is also John Calvin's *sensus divinitatis* or Thomas Aquinas's "natural but confused knowledge of God" whereby we know something of God.<sup>6</sup> These faculties or powers work together in complex and variegated ways to produce a vast battery of beliefs and knowledge, ranging from the simplest everyday beliefs—it's hot in here, I have a pain in my right knee—to less quotidian beliefs such as those to be found in philosophy, theology, history, and the far reaches of science. In science, clearly enough, many of these faculties work together—perception, memory, testimony, sympathy, induction, a priori intuition are all typically involved. There is also the whole process of theory building, which may or may not be reducible to the previous abilities.

My argument will concern the *reliability* of these cognitive faculties. My memory, for example, is reliable only if it produces mostly true beliefs—if, that is, most of my memorial beliefs are true. What proportion of my memorial beliefs must be true



for my memory to be reliable? Of course there is no precise answer; but presumably it would be greater than, say, two-thirds. We can speak of the reliability of a particular faculty—memory, for example—but also of the reliability of the whole battery of our cognitive faculties. And indeed we ordinarily think our faculties *are* reliable, at any rate when they are functioning properly, when there is no cognitive malfunction or disorder or dysfunction. (If I get drunk and suffer from delirium tremens, my perception will be impaired and all bets are off with respect to its reliability.) We also think they are more reliable under some circumstances than others. Visual perception of middle-sized objects (medium-sized dry goods, as J. L. Austin called them) close at hand is more reliable than perception of very small objects, or middle-sized objects at some distance (a mountain goat from six hundred yards, for example). Beliefs about where I was yesterday are ordinarily more likely to be true than the latest high-powered scientific theories.

Now the natural thing to think, from the perspective of theism, is that our faculties are indeed for the most part reliable, at least over a large part of their range of operations. According to theistic religion (see [chapter 9](#)), God has created us in his image; an important part of this image consists in our resembling God in that like

him, we can have knowledge. In [chapter 9](#) we saw that Thomas Aquinas put it as follows: “Since human beings are said to be in the image of God in virtue of their having a nature that includes an intellect, such a nature is most in the image of God in virtue of being most able to imitate God.”<sup>7</sup> When Thomas speaks of our nature as including an intellect, he clearly means to endorse the thought that our cognitive faculties are for the most part reliable. But suppose you are a naturalist: you think that there is no such person as God, and that we and our cognitive faculties have been cobbled together by natural selection. Can you then sensibly think that our cognitive faculties are for the most part reliable?

I say you can't. The basic idea of my argument could be put (a bit crudely) as follows. First, the probability of our cognitive faculties being reliable, given naturalism and evolution, is low. (To put it a bit inaccurately but suggestively, if naturalism and evolution were both true, our cognitive faculties would very likely not be reliable.) But then according to the second premise of my argument, if I believe both naturalism and evolution, I have a *defeater* for my intuitive assumption that my cognitive faculties are reliable. If I have a defeater for *that* belief,

however, then I have a defeater for *any* belief I take to be produced by my cognitive faculties. That means that I have a defeater for my belief that naturalism and evolution are true. So my belief that naturalism and evolution are true gives me a defeater for that very belief; that belief shoots itself in the foot and is self-referentially incoherent; therefore I cannot rationally accept it. And if one can't accept both naturalism and evolution, that pillar of current science, then there is serious conflict between naturalism and science.

So much for an initial and rough statement of the argument; now we must proceed to develop it more carefully. The first premise, as I say, is something like the worry or doubt that our cognitive faculties would not be reliable if both naturalism and evolution (or perhaps just naturalism) were true. This worry has some eminent advocates. For example, there is Friederich Nietzsche. Ordinarily what Nietzsche says inspires little confidence, but in the following he may be on to something:

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.<sup>8</sup>**

**To leap to the present, there is the philosopher Thomas Nagel, himself no friend of theism: “If we came to believe that our capacity for objective theory [true beliefs, e.g.] were the product of natural selection, that would warrant serious skepticism about its results.”<sup>9</sup> According to another philosopher, Barry Stroud (again, no friend of theism), “There is an embarrassing absurdity in [naturalism] that is revealed as soon as the naturalist reflects and acknowledges that he believes his naturalistic theory of the world.... I mean he cannot say it and consistently regard it as true.”<sup>10</sup> As Patricia Churchland, an eminent naturalistic philosopher, puts it in a justly famous passage:**

**Boiled down to essentials, a nervous system enables the organism to succeed in the four F’s: feeding, fleeing, fighting and reproducing. The principle chore of nervous systems is to get the body parts where they should be in order that the organism may**

survive..... Improvements in sensorimotor control confer an evolutionary advantage: a fancier style of representing is advantageous *so long as it is geared to the organism's way of life and enhances the organism's chances of survival.* Truth, whatever that is, definitely takes the hindmost.<sup>[11](#)</sup>

Churchland's point, clearly, is that (from a naturalistic perspective) what evolution guarantees is (at most) that *we behave* in certain ways—in such ways as to promote survival, or more exactly reproductive success. The principal function or purpose, then, (the “chore” says Churchland) of our cognitive faculties is not that of producing true or verisimilitudinous (nearly true) beliefs, but instead that of contributing to survival by getting the body parts in the right place. What evolution underwrites is only (at most) that our *behavior* is reasonably adaptive to the circumstances in which our ancestors found themselves; hence it does not guarantee mostly true or verisimilitudinous beliefs. Our beliefs *might* be mostly true or verisimilitudinous (hereafter I'll omit the “verisimilitudinous”); but there is no particular reason to think they *would* be: natural selection is interested, not in truth, but in appropriate behavior. What Churchland

therefore suggests is that naturalistic evolution—that is, the conjunction of metaphysical naturalism with the view that we and our cognitive faculties have arisen by way of the mechanisms and processes proposed by contemporary evolutionary theory—gives us reason to doubt two things: (a) that a *purpose* of our cognitive systems is that of serving us with true beliefs, and (b) that they *do*, in fact, furnish us with mostly true beliefs.

Indeed, Darwin himself expresses serious doubts along these lines: “With me the horrid doubt always arises whether the convictions of man’s mind, which has been developed from the mind of the lower animals, are of any value or at all trustworthy. Would any one trust in the convictions of a monkey’s mind, if there are any convictions in such a mind?”<sup>12</sup>

#### IV THE FIRST PREMISE: DARWIN’S DOUBT

Nietzsche, Nagel, Stroud, Churchland, and Darwin, nontheists all, seem to concur: (naturalistic) evolution gives one a reason to doubt that human cognitive faculties produce for the most part true beliefs. Since Darwin is the standout among this

group, call this thought “Darwin’s doubt.” How shall we construe Darwin’s doubt? Can we state it a bit more exactly?

Here the idea of *conditional probability* will be useful. This is a familiar idea, one we constantly employ. The conditional probability of one proposition  $p$  on another proposition  $q$  is the probability that  $p$  is true *given that*, on the condition that,  $q$  is true.

Consider the probability that Mr. A will live to be eighty years old, given that he is now thirty-five, smokes heavily, is grossly overweight, eats only junk food, never exercises, and had grandparents all of whom died by the age of fifty: this probability is pretty low. Contrast this probability with the probability that Mr. B will live to be eighty, given that Mr. B is now seventy, has never smoked, watches his diet like a hawk, runs ten miles a day, and has grandparents all of whom lived to be over one hundred; that probability is much higher. With this notion of conditional probability in hand, we can put Darwin’s doubt as follows: the conditional probability that our cognitive faculties are reliable, given naturalism together with the proposition that we have come to be by way of evolution, is low. This is quite a mouthful: we can abbreviate it as

(1)  $P(R/N\&E)$  is low.

**“R” is the proposition that our cognitive faculties are reliable, “N” is naturalism, and “E” is the proposition that we and our cognitive faculties have come to be in the way proposed by the contemporary scientific theory of evolution. “P(..../\_\_\_)” is shorthand for “the probability of... given \_\_\_”. (1), that is, Darwin’s doubt, is the first premise of my argument.**

**All of the above luminaries apparently endorse something like Darwin’s doubt; nevertheless (oddly enough) there are those who seem to disagree. In what follows, therefore, I’ll explain why Darwin’s doubt seems eminently sensible and indeed correct.**

### ***A. Naturalism and Materialism***

First, we must note that nearly all naturalists are also *materialists* with respect to human beings; they hold that human beings are material objects. From this perspective a human person is not (contrary to Descartes and Augustine) an immaterial substance or self that is connected with or joined to (has?) a material body. Nor is it the case that a human being is a composite that has an



immaterial component; human beings do not have an immaterial soul or mind or ego. Instead, so the materialist thinks, a person *just is* her body, or perhaps some part of her body (so that talk about “my body” is misleading). I *am* my body (or maybe my brain, or its left hemisphere, or some other 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*.<sup>13</sup> Most of these objections (including Dennett’s) are astonishingly weak; no one not already convinced of materialism would (or at any rate should) find them at all persuasive.<sup>14</sup> Still, they are often trotted out as showing that we are all obliged, these enlightened days, to be materialists.

A second and somewhat 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 (for naturalists), but not wholly conclusive. That is because of the vagueness of the concept of naturalism. According to naturalism, there isn't anything *like* God; but just how much similarity to God is tolerable, from a naturalistic perspective? After all, everything resembles God in *some* respect; how much similarity to God can a decently sensitive naturalist manage to accept? 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 allow such a thing? That's not entirely easy to say. Far be it from me as an outsider, however, to intrude upon a delicate family dispute among naturalists; I hereby leave naturalists to settle this issue for themselves.

A third reason is as follows. Naturalists will ordinarily endorse Darwinian evolution; but how, they ask, 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."<sup>15</sup> 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. Though most mutations of this sort are lethal, 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, they ask, 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, but for an immaterial self?<sup>16</sup> That seems doubtful.

These reasons clearly aren't conclusive, but most naturalists find them (or perhaps other arguments for materialism) at least reasonably compelling. For these reasons and perhaps others, most naturalists are materialists about human beings. For present purposes, therefore, I propose to assimilate materialism to naturalism; henceforth I'll think of naturalism as including materialism, and what I'll be arguing against is the conjunction of current evolutionary theory and naturalism, the latter including materialism.

## ***B. Beliefs as Neural Structures***

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 hold the belief 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 100–200 billion neurons. These neurons, furthermore, are connected with other neurons via synapses; a single neuron, on the average, is connected with seven thousand of other neurons. 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—that is, 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, for example, 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 as well as to muscles and glands. But if this is the sort of thing beliefs are, if they are neuronal events or structures, they will have two quite different sorts of properties. On the one hand they will have *electro-chemical* or *neuro-physiological* 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 a *content*.<sup>17</sup> It will 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*. My belief that naturalism is vastly overrated has as content the proposition *naturalism is vastly overrated*. (That same proposition is the content of the Chinese speaker's belief that naturalism is vastly overrated, even though she expresses this belief by uttering a very different sentence; beliefs, unlike sentences, do not come in different languages.) It is in virtue of having content 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; Hitler's belief that the Third Reich would last a thousand years was false, because the proposition that constituted its content is (was) false.

Given materialism, therefore, beliefs are (ordinarily) long-standing neural events. As such, they have NP properties, but also content properties: each belief will have the property of having such and such a proposition as its content. NP properties are *physical* properties; on the

other hand content properties—for example the property of having as content the proposition *all men are mortal*—are *mental* properties. Now how, according to materialism, are mental and physical properties related? In particular, how are content properties related to NP properties—how is the content property of a particular belief related to the NP properties of that belief?

### ***C. Reductive and Nonreductive Materialism***

Materialists offer fundamentally two theories about the relation between physical and mental properties (and hence two theories about the relation between NP properties and content properties): reductive materialism and nonreductive materialism. According to Sir Francis Crick: “your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behaviour of a vast assembly of nerve cells and their associated molecules.”<sup>18</sup> This is a pretty good statement of *reductive* materialism, according to which (naturally enough) mental content properties are *reducible to* NP properties; according to *nonreductive* materialism, content properties are not reducible to NP properties, but are *determined by* (supervene on) NP properties.<sup>19</sup> We could put it like this: according to

reductive materialism, there is only one kind of property in the neighborhood: NP properties, some of which are also mental properties. According to nonreductive materialism, on the other hand, there are *two* kinds of properties, NP properties and also mental properties, which are not NP properties, but are determined by NP properties.

Suppose we think first about reductive materialism. Consider the property of having as content the proposition *naturalism is vastly overrated*, and call this property “C.” On reductive materialism, 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”)

$$P_1 \vee P_2 \vee P_3 \vee P_8 \vee \dots \vee P_n$$

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

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

If complex combinations of NP properties are themselves

NP properties, content properties, on reductive materialism, are really just a special kind of NP property. According to reductive materialism, therefore content properties—for example, the property of having *naturalism* is vastly overrated as content—are or are reducible to NP properties.

That's one of the two proposals made by materialists. The other is that a content property isn't an NP property, and can't be reduced to NP properties, but is nevertheless *determined by* NP properties. Here the basic idea is this: for any particular mental property M you pick, there is a physical property P such that necessarily, if a thing has M, then it has P, and if a thing has P, then it has M. <sup>21</sup> So take any mental property—for example, the property of being in pain: there will be some physical property P (presumably an NP property), such that it's true in every possible world that whatever has P is in pain, and, conversely, whatever is in pain has P. <sup>22</sup> Specified to content and NP properties, the idea is that for any content property C that a neural structure can have, there is an NP property P such that if a neural structure has that content property C, it has P, and conversely, any neural structure that has P also has that content property C.

According to both reductive and nonreductive materialism, mental properties are determined by physical properties (and indeed according to reductive materialism mental properties just *are* physical properties). 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 bacteria; presumably they have no beliefs at all. At the other end of the scale there are human beings, who have a rich and varied store of beliefs and whose brains contain many billions of neurons connected in complex and multifarious ways, so that the number of different possible brain states is more than astronomical. And 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 there arises something we can properly call a belief, something that is true or false. At a certain level of complexity, these neural structures start to display belief content. Perhaps this starts gradually and early on—perhaps it is with *C. elegans*, a small but charismatic beast that enjoys the distinction of having its nervous system completely mapped. Possibly *C. elegans* displays just the merest glimmer of consciousness and just the merest glimmer of actual belief content, or perhaps belief content shows up further up the scale; that doesn't matter. What does matter is that at a certain level of complexity, neural structures begin to display content and the creatures that harbor those structures have beliefs. This is true whether content properties are reducible to NP properties or supervene on them.

So (given materialism) some neural structures, at a certain level of complexity of NP properties, acquire content; at that level of complexity, NP properties determine belief content, and the structures in question are beliefs. And the question I want to ask is this: what is the likelihood,

*given evolution and naturalism* (construed as including materialism about human beings), that the content thus arising is in fact *true*? In particular, what is the likelihood, given N&E, that the content associated with *our* neural structures is true? What is the likelihood, given N&E, that our cognitive faculties are reliable, thereby producing mostly true beliefs?

## **V THE ARGUMENT FOR PREMISE (1)**

**We are now ready to state the reasons for the first premise of the main argument, which, as you recall is**

**(1)  $P(R/N\&E)$  is low.**

**Of course 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 I was last night, that I've just had oatmeal for breakfast, that my elder son's name is not Archibald, that a year ago I didn't live in the 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 assumption, and should give it up. I don't mean to argue that this natural assumption is false; like everyone else, I believe that our cognitive faculties *are*, in fact, mostly reliable. What I do mean to argue is that the *naturalist*—at any rate a naturalist who accepts evolution—is rationally obliged to give up this assumption.**

### ***A. The Argument and Nonreductive Materialism***

As you recall, we are thinking of naturalism as including materialism, which, as we've seen, comes in two varieties: reductive and nonreductive. Let's think first about the

question from the point of view of nonreductive materialism. Return to the evolutionary scale and *C. elegans*, that celebrated little worm, and suppose that it is in *C. elegans* that we first get belief. No doubt such belief will be primitive *in excelsis* (and if you don't think *C. elegans* has beliefs, you can simply go up the scale until you encounter creatures you think do have beliefs), but let's suppose members of this species have beliefs. Now given that *C. elegans* has survived for millions of years, we may assume that its behavior is adaptive. This behavior is produced or caused by the neurological structures in the *C. elegans* nervous system; we may further assume, therefore, that this neurology is adaptive. This underlying neurology causes adaptive behavior; as Churchland says, it gets the body parts where they must be in order to survive. But (in line with nonreductive materialism) it also determines belief content. As a result, these creatures have beliefs, which of course have a certain content.

And here's the question: what reason is there for supposing that this belief content is *true*? There isn't any. The neurology causes adaptive behavior and also causes or determines belief content: but there is no reason to suppose that the belief content thus determined is true. All that's required for survival and fitness is that the neurology cause adaptive behavior; this neurology also determines belief content, but whether or not that content is *true* makes no difference to fitness. Certain NP properties are selected for, because they contribute to fitness. These NP properties also cause or determine belief content; they associate a

content or proposition with each belief. The NP properties are selected, however, not because they cause the content they do, but because they cause adaptive behavior. If the content, the proposition determined by the neurology (the NP properties of the belief) is true, fine. But if it is false, that's no problem as far as fitness goes.

Objection: consider a frog on a lily pad. A fly buzzes by; the frog's tongue flicks out and captures the fly. If this frog is to behave successfully, adaptively, there must be mechanisms in it that register the distance to the fly at each moment, its size, speed, direction, and so on. Aren't these mechanisms part of the frog's cognitive faculties? And don't they have to be accurate in order for the frog to behave adaptively? And isn't it therefore the case that the frog's cognitive mechanisms must be accurate, reliable, if the frog is to survive and reproduce? Or consider an animal, maybe a zebra, grazing on the veldt; a lion approaches. The zebra notices the predator; this noticing consists in part of some neural structure arising in its brain, perhaps a certain pattern of firing of neurons in the optical portion of its brain, and perhaps this pattern ordinarily arises in response to the appearance of a predator in the middle distance. If this structure isn't properly correlated with the presence of predators, the zebra won't be long for this world. And wouldn't this structure, furthermore, be part of the creature's cognitive mechanisms? And don't those mechanisms have to be accurate, reliable, if the zebra is to survive?

Reply: that frog clearly does have "indicators," neural

structures that receive input from the frog's sense organs, are correlated with the path of the insect as it flies past, and are connected with the frog's muscles in such a way that it flicks out its tongue and captures that unfortunate fly. The same goes for the zebra: if it is to behave adaptively (evade predators, for example) it too will have to have indicators, neural structures that monitor the environment, that are correlated (for example) with the presence of predators, and are connected with its muscles in such a way as to cause it to flee when a predator threatens.

Now if we like, we can include these indicators under the rubric "cognitive faculties." The important point to see here, however, is that indication of this sort does not require *belief*. In particular, it does not require belief having to do with the state of affairs indicated; indeed it is entirely compatible with belief *inconsistent* with that state of affairs. For example, anaerobic marine bacteria (so the story goes) contain magnetosomes, tiny internal magnets that indicate magnetic north; in the oceans of the northern hemisphere, this direction is down, towards the oxygen-free depths.<sup>23</sup> These indicators are connected with the propulsion devices of the bacteria in such a way as to cause these creatures, which can't flourish in the oxygen-rich surface water, to move towards the deeper water. But this in no way requires that the bacteria form *beliefs*. Fleeing predators, finding food and mates—these things require cognitive devices that in some way track crucial features of the environment, and are appropriately connected with muscles; but they do not require true belief,

or even belief at all. The long-term survival of organisms of a certain species certainly makes it likely that its members enjoy cognitive devices that are successful in tracking those features of the environment—indicators, as I've been calling them. Indicators, however, need not be or involve beliefs. In the human body there are indicators for blood pressure, temperature, saline content, insulin level, and much else; in these cases neither the blood, nor its owner, nor anything else in the neighborhood ordinarily holds beliefs on the topic. The objector is therefore right in pointing out that fitness requires accurate indication; but nothing follows about reliability of belief.

Returning to the main line of our argument, we are considering nonreductive materialism and asking about  $P(R/N\&E)$ , given nonreductive materialism. (Another way to put this: we are considering  $P(R/N\&E \& \text{nonreductive materialism})$ .) In order to avoid automatically introducing into the argument our ordinary assumptions about our own mental life, 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 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. Suppose further that a certain content, a certain proposition, is determined by the NP properties of this structure. This structure, therefore, will be a belief, and will have a certain proposition  $p$  as its content.

But now for the crucial question: what is the probability (given N&E) 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 sufficient for the possession of that particular content. We are assuming also that this structure arises in response to the presence of that predator. We can assume further, if we like, that this structure is a reliable indicator of that kind of predator: it arises when and only when there is a such a predator in the middle distance. But why think it is a *true* proposition that is determined by those NP properties? These NP properties determine a proposition: but why think that proposition is true? Natural selection selects for adaptive NP properties; those NP properties determine content; but natural selection just has to take pot-luck with respect to the propositions or content determined by those adaptive NP properties. It does not get to influence or modify the function from NP properties to content properties: that's just a matter of logic or causal law, and natural selection can't modify either. Indeed, the content generated by the NP properties of this structure, on



this occasion, need have nothing to do with that predator, or with anything else in the environment. True: the structure is correlated with the presence of a predator and indicates that presence; but indication is not belief. Indication is one thing; belief content is something else altogether, and we know of no reason (given materialism) why the one should follow the other. We know of no reason why the content of a belief should match what that belief (together, perhaps, with other structures) indicates. 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 indicates, if anything. Indeed, the proposition constituting that content need not be so much as *about* that predator; it certainly need not be true.

What, then, is the likelihood that this proposition, this content, is true? Given just this much, shouldn't we suppose that the proposition in question is as likely to be false as true? Here's the picture: the NP properties of a belief are adaptive in that they cause adaptive behavior. Those NP properties also determine a content property. But as long as the NP properties are adaptive, it doesn't matter, for survival and reproduction, what content is determined by those NP properties. It could be true content; it could be false content; it doesn't matter. The fact that these creatures have survived and evolved, that their cognitive equipment was good enough to enable their ancestors to survive and reproduce—that fact would tell us nothing at all about the *truth* of their beliefs or the reliability of their cognitive faculties. It would tell something about the

*neurophysiological* properties of a given belief; it would tell us that by virtue of these properties, that belief has played a role in the production of adaptive behavior. But it would tell us nothing about the truth of the *content* of that belief: its content might be true, but might with equal probability be false. So shouldn't we suppose that the proposition in question has a probability of roughly .5? 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 should estimate the probability of its being true as .5.

The probability we are thinking of here is objective, not the personalist's subjective probability, and also not epistemic probability.<sup>24</sup> (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.) But then, in suggesting the first attitude above, am I not relying upon the notorious Principle of Indifference? And hasn't that principle been discredited?<sup>25</sup> Not really. The Bertrand paradoxes show that certain incautious statements of the principle of indifference 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. Still, the fact is we project properties all the time, and do so perfectly sensibly. And the fact is we also regularly employ a principle of indifference in ordinary

reasoning, and do so quite properly. We also use it in science—for example in statistical mechanics.<sup>26</sup>

Given that the probability, for any belief on the part of these creatures, is about .5, what is the probability that their cognitive faculties are *reliable*? Well, what proportion of my beliefs must be true, if my faculties are reliable? The answer will have to be vague; perhaps a modest requirement would be that a reliable cognitive faculty must deliver at least 3 times as many true beliefs as false: the proportion of true beliefs in its output is at least three-quarters. If so, then the probability that their faculties produce the preponderance of true beliefs over false required by reliability is very small indeed. If I have one thousand independent beliefs, for example, the probability (under these conditions) that three quarters or more of these beliefs are true will be less than  $10^{-58}$ .<sup>27</sup> And even if I am running a modest epistemic establishment of only one hundred beliefs, the probability that three-quarters of them are true, given that the probability of any one's being true is one half, is very low, something like .000001. So the chances that this creature's true beliefs substantially outnumber its false beliefs are small. The conclusion to be drawn is that it is very unlikely that the cognitive faculties of those creatures are reliable. But of course the same will go for us:  $P(R/N\&E)$  specified not to them but to us, will also be very low.

## ***B. The Argument and Reductive Materialism***

That's how things stand for nonreductive materialism:  $P(R/N\&E\&\text{nonreductive materialism})$  is low. We can deal more briefly with  $P(R/N\&E\&\text{reductive materialism})$ , the probability of R given naturalism and evolution and reductive materialism. On reductive materialism, mental properties are complex combinations of physical properties; more briefly, taking complex combinations of physical properties to be themselves physical properties, mental properties just are physical properties. What is the probability of R on N&E and reductive materialism?

Here we get the very same results as with nonreductive materialism. To see why, consider, again, any given belief on the part of a member of that hypothetical group of creatures—say the belief *naturalism is vastly overrated*. That belief 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. Since the event is a belief, some subset of these NP

properties together constitute its having the content it does in fact display. That is, there will be some proposition that is the content of the belief; the belief will therefore have the property of having that proposition as its content; and that property, the property of having such and such a proposition as its content, will be a (no doubt complex) NP property of the belief.

Now what is the probability that this content is *true*? What is the probability that this proposition, whatever it is, is true? 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 adaptive arrangement of NP properties also constitutes having that particular content. But again: it would be a piece of serendipity if this content, this proposition, were *true*; it could just as well be false. These NP properties, including those that constitute its having that content, are adaptive just as long as they cause adaptive behavior. They also constitute the property of having that particular content; but it doesn't matter at all, so far as adaptivity goes, whether that content is true. So take any particular belief on the part of one of those creatures. We may suppose (given that these creatures have come to be by way of evolution) that having this belief is adaptive; its NP properties cause adaptive behavior. These NP properties also constitute the property of having such and such content; but, clearly enough, it doesn't matter (with respect to the adaptivity of these properties) whether the content they constitute is true.

It could be true: fair enough; but it could equally well be false. If these properties had constituted different content, they still would have had the same causal effect with respect to behavior. Hence the probability that the content of this belief is true would have to be rated at about one-half, just as in the case of nonreductive materialism. If this is true for each of the independent beliefs of the organism in question, however, the probability that the cognitive faculties of these creatures are reliable (on N&E& reductive materialism), would have to be rated as low. The conclusion to be drawn so far, then, is that given N&E (N including materialism), it is unlikely that these creatures have reliable cognitive faculties.

### ***C. Objection***

Isn't it just obvious that true beliefs will facilitate adaptive action? A gazelle who mistakenly believes that lions are friendly, overgrown house cats won't be long for this world. The same goes for a rock climber who believes that jumping from a two-hundred-foot cliff will result in a pleasant and leisurely trip down with a soft landing. Isn't it obvious both that true beliefs are much more likely to be adaptive than false beliefs? Isn't it obvious, more generally, that true beliefs are more likely to be successful than false beliefs? I

want to go from New York to Boston: won't I be more likely to get there if I believe that Boston is north of New York than if I believe it's to the south?

Yes, certainly. This is indeed true. But it is also irrelevant. We are not asking about how things *are*, but about *what things would be like if both evolution and naturalism (construed as including materialism) were true*. We are asking about  $P(R/N\&E)$ , not about  $P(R/\text{the way things actually are})$ . Like everyone else, I believe that our cognitive faculties are for the most part reliable, and that true beliefs are more likely to issue in successful action than false. But that's not the question. The question is what things would be like if N&E were true; and in this context we can't just assume, of course, that if N&E, N including materialism, were true, then things would still be the way they are. That is, we can't assume that if materialism were true, it would still be the case that true beliefs are more likely to cause successful action than false beliefs. And in fact, if materialism were true, it would be unlikely that true beliefs mostly cause successful action and false belief unsuccessful action.

Here you may ask, "Why think a thing like that? What has materialism to do with this question?" Here's what. We ordinarily think true belief leads to successful action because we also think that beliefs cause (part-cause) actions, and do so *by virtue of their content*. I want a beer; I believe there is one in the fridge, and this belief is a (part) cause of my going over to the fridge. We think it is by virtue of the *content* of that belief that it causes me to go over to

the fridge; it is because this belief has as content that there is a beer in the fridge that it causes me to go to the fridge rather than, say, the washing machine. More generally, we think it is by virtue of the content of a belief B that B part-causes the behavior that it does cause.

But now suppose materialism were true: then, as we've seen, my belief will be a neural structure that has both NP properties and also a propositional content. It is by virtue of the NP properties, however, not the content, that the belief causes what it does cause. It is by virtue of *those* properties that the belief causes neural impulses to travel down the relevant efferent nerves to the relevant muscles, causing them to contract, and thus causing behavior. It isn't by virtue of the content of this belief; the content of the belief is irrelevant to the causal power of the belief with respect to behavior.

Consider an analogy. I am playing catch with my granddaughter, and in a vainglorious attempt to show off, I throw the ball too hard; it whistles over her head and shatters a neighbor's window. It is clear that the ball breaks the window *by virtue of* its mass, velocity, hardness, size, and the like. If it had been much less massive, been traveling at a lower rate of speed, had been as soft as a bunch of feathers, it would not have broken the window. If you ask "Why did the window shatter upon being hit by the ball?" the correct answer will involve the ball's having those properties (and of course also involve the window's having a certain degree of brittleness, tensile strength, and the like). As it happens, the ball was a birthday present; but it



does not break the window by virtue of being a birthday present, or being purchased at Sears and Roebuck, or costing \$5.00. Examples of this sort, clearly enough, can be multiplied endlessly; but examples of other kinds also bound. Sam has the right to fire the city manager by virtue of his being mayor, not by virtue of his being nice to his wife. Aquinas was a great philosopher by virtue of his acumen and insight and prodigious industry, not by virtue of his being called “the Dumb Ox.”<sup>28</sup>

Going back to materialism and the content of belief, then, it is by virtue of the NP properties of a belief B, not by virtue of its content, that the belief causes the behavior it does cause. Among B’s NP properties are such properties as that of involving many neurons working in concert: as we learn from current science, these neurons send a signal through effector nerves to the relevant muscles, causing those muscles to contract and thereby causing behavior. It is by virtue of these NP properties that it causes those muscles to contract. If the belief had had the same NP properties but different content, it would have had the same effect on behavior.

Objection: you claim that

(1) If the belief B had had the same NP properties but different content, it still would have had the same causal effects with respect to behavior;

but it *couldn't* have had the same NP properties but different content. (1) is not merely counterfactual; it's counterpossible. If the property of having C as content supervenes on neurophysiological properties, then (given strong super-venience) there will be a neurophysiological property equivalent to C in the broadly logical sense; hence it won't be so much as possible that the antecedent of (1) hold. Given the usual semantics for counterfactuals, the conclusion to be drawn is that (1) is true, all right, but so is any counterfactual with the same antecedent, including, for example

(2) if B had had the same content but different neurophysiological properties, B would *not* have had the same causal effects with respect to behavior.

Right. But *is* the usual semantics for counterfactuals correct? This is hardly the place to address that particular (and large) can of worms, but in fact (so I think) it isn't. It is true that if 2 had been greater than 3, then 3 would have been less than 2; it is not true that if 2 had been greater than 3, then 3 would have been greater than 2. It is not true that if 2 had been greater than 3, then the moon would have been made of green cheese. Even given that God is necessarily omniscient, it isn't true that if God had not been omniscient, he would have known that he doesn't exist. If I proved Gödel wrong, logicians everywhere would be

astonished; it is false that if I proved Gödel wrong, logicians would yawn in boredom.

Furthermore, philosophers regularly and quite properly use counterpossibles in arguing for their views. Consider the philosophical view that what I really am is a member of a series of momentary person stages. One argues against this view by pointing to the truth of

(3) if this were true, I wouldn't be responsible for anything that happened more than a moment ago (a new legal defense strategy?)

Even though the view in question is noncontingent—necessarily true or necessarily false—you take that counterpossible to be true and its mate

(4) if this were true, I *would* be responsible for much that happened more than a moment ago

false. A dualist might claim that if materialism were true, the content of one's beliefs wouldn't enter the causal chain leading to behavior; a materialist might claim that if (interactive) dualism were true, an immaterial substance would (implausibly) cause effects in the hard, heavy, massy material world. One of these counterfactuals has an impossible antecedent; both, however, are properly used in

the dispute between materialists and dualists.

The truth of (1) gives us some reason to think that B doesn't cause that action A by virtue of its content. As I say, however, this isn't the place to look into the difficult matter of figuring out how to reason with counterpossibles; that would take us far afield. But we can also address our question directly: is it by virtue of its content that B causes A? I should think the answer, clearly, is that it is not. It is by virtue of its neurophysiological properties that B causes A; it is by virtue of *those* properties that B sends a signal along the relevant nerves to the relevant muscles, causing them to contract, and thus causing A. It isn't by virtue of its having that particular content C that it causes what it does cause.

So once again: suppose N&E were true. Then materialism would be true in either its reductive or its nonreductive form. In either case, the underlying neurology is adaptive, and determines belief content. But in either case it doesn't matter to the adaptiveness of the behavior (or of the neurology that causes that behavior) whether the content determined by that neurology is true.<sup>[29](#)</sup>

## VI THE REMAINING PREMISES

**Now we're ready for the next step: the naturalist who sees that  $P(R/N\&E)$  is low has a *defeater* for R,**

and for the proposition that his own cognitive faculties are reliable. A defeater for a belief B I hold—at any rate this kind of defeater—is another belief B\* I come to hold which is such that, given that I hold B\*, I can no longer rationally hold B.<sup>30</sup> 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 guide book is notorious for giving the wrong date for the foundation of the University. (Actually it was established in 1495.) My new belief that the University was established in 1495 is a defeater for my old belief. In the same way, if I accept naturalism and see that  $P(R/N\&E)$  is low, then I have a defeater for R; I can no longer rationally believe that my cognitive faculties are reliable.

So the second premise of the argument:

(2) Anyone who accepts (believes) N&E and sees that  $P(R/N\&E)$  is low has a defeater for R.

It isn't that someone who believed N&E wouldn'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 isn't possible to acquire evidence for R, 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. If I do that, however, I am already assuming R to be true, at least for the faculties or processes that produce in me belief in the premises of the argument, and the belief that if the premises are true, so is the conclusion. 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.

So the belief that my cognitive faculties are reliable is one for which I don't need evidence or argument—that is, I don't need evidence or argument in order to be rational in believing it. I can be fully and entirely rational in believing this even though I have no evidence or argument for it at all. This is a belief such that it is rational to hold it in the

*basic* way, that is, not on the basis of argument or evidence from other things I believe. But that doesn't mean it isn't possible to acquire a defeater for it. Even if a belief is properly basic, it can still be defeated. 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 we can reuse an example from [chapter 6](#) 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 facades (indistinguishable, from the highway, from real barns) erected by the local inhabitants in a dubious effort to make themselves look more prosperous. 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.

And this is what happens when I believe N&E, and come to see that  $P(R/N\&E)$  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.

Consider an analogy. Suppose there is a drug—call it XX—that destroys cognitive reliability. I know that 95 percent of those who ingest XX become cognitively unreliable within two hours of ingesting it; they then believe more false propositions than true. Suppose further that I come to believe both that I've ingested XX a couple of hours ago and that  $P(R/I've\ ingested\ XX\ a\ couple\ of\ hours\ ago)$  is low; taken together, these two beliefs give me a defeater for my initial belief that my cognitive faculties are reliable.<sup>31</sup> 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 now just as suspect or compromised 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.

Objection: why should we think that premise (2) is true? Some propositions of that form are true, but some aren't. I believe that I've ingested XX, and that the probability that I am reliable, given that I've ingested XX is low; this gives me a defeater for the proposition that I am reliable. But I also believe that the probability that I live in Michigan, given that the earth revolves around the sun, is low, and I believe



that the earth revolves around the sun; this does not give me a defeater for my belief that I live in Michigan. Why think the case of N&E and R is more like the first than like the second?<sup>32</sup>

Reply: Right: not every proposition of that form is true. This one is, however. What's at issue, I think, is the question what else I believe (more exactly what else is such that I believe it and can legitimately conditionalize on it in this context). If the only thing I knew, relevant to

(a) my living in Michigan,

is that this is unlikely given that

(b) the earth revolves around the sun,

then my belief that (b) and that (a) is unlikely on (b) *would* give me a defeater for (a). But of course I know a lot more: for example, that I live in Grand Rapids, which is in Michigan. I quite properly conditionalize not just on (b), but on much else, on some of which (a) has a probability of 1. But now think about N&E and R. We agree that  $P(R/N\&E)$  is low. Do I know something else X, in addition to N&E, such that (a) I can properly conditionalize on

**X, and (b)  $P(R/ N\&E\&X)$  is high? This is the conditionalization problem, which I address briefly on pages 346.**

**This brings us to the third premise:**

**(3) Anyone who has a defeater for R has a defeater for any other belief she thinks she has, including N&E itself.**

**(3) is pretty obvious. If you have a defeater for R, you will also have a defeater for any belief you take to be produced by your cognitive faculties, any belief that is a deliverance of your cognitive faculties. But *all* of your beliefs, as I'm sure you have discovered, are produced by your cognitive faculties. Therefore you have a defeater for any belief you have.**

**Still, even if you realize you have a defeater for every belief you hold, you are unlikely to give up all or perhaps even any of your beliefs. 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 dismissive Humean thoughts about, say, induction when clinging unroped (you're free-soloing) to a rock face five hundred feet up the East Buttress of El**

**Capitan. (You won't find yourself saying, "Well, naturally 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 a really crushing skepticism, and it is this skepticism to which the naturalist is committed.**

**The final premise of the argument is**

**(4) If one who accepts N&E thereby acquires a defeater for N&E, N&E is self-defeating and can't rationally be accepted.**

**The entire argument, therefore, goes as follows:**

**(1)  $P(R/N\&E)$  is low.**

- (2) Anyone who accepts (believes) N&E and sees that  $P(R/N\&E)$  is low has a defeater for R.
- (3) Anyone who has a defeater for R has a defeater for any other belief she thinks she has, including N&E itself.
- (4) If one who accepts N&E thereby acquires a defeater for N&E, N&E is self-defeating and can't rationally be accepted. Conclusion: N&E can't rationally be accepted.

This argument shows that if someone accepts N&E and sees that  $P(R/N\&E)$  is low, then she have a defeater for N&E, a reason to reject it, a reason to doubt or be agnostic with respect to it.

Of course defeaters can themselves be defeated; so couldn't you get a defeater for this defeater—a defeater-defeater? Maybe by doing some science—for example, determining by scientific means that her faculties really are reliable? Couldn't she go to the MIT cognitive-reliability laboratory for a check-up?<sup>33</sup> Clearly that won't help. Obviously that course would *presuppose* that her faculties are reliable; she'd be relying on the accuracy of her faculties in believing that there is such a thing as MIT, that she has in fact consulted its scientists, that they have given her a clean bill of cognitive health, and so on. The great Scottish philosopher Thomas Reid put it like this:

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.<sup>34</sup>

Is there any sensible way at all in which she can argue for R? It is hard to see how. Any argument she might produce will have premises; these premises, she claims, give her good reason to believe R. But of course she has the very same defeater for each of those premises that she has for R, and she has the same defeater for the belief that if the premises of that argument are true, then so is the conclusion. So it looks as if this defeater can't be defeated. Naturalistic evolution gives its adherents a reason for doubting that our beliefs are mostly true; chances are they are mostly mistaken. If so, it won't help to *argue* that they can't be mostly mistaken; for the very reason for mistrusting our cognitive faculties *generally*, will be a reason for mistrusting the faculties that produce belief in the goodness of that argument.

**This defeater, therefore, can't be defeated. Hence the devotee of N&E has an undefeated defeater for N&E. N&E, therefore, cannot rationally be accepted—at any rate by someone who is apprised of this argument and sees the connections between N&E and R.**

## **VII TWO CONCLUDING COMMENTS**

**First, a comment on premise (2), according to which anyone who accepts (or believes) N&E and sees that  $P(R/N\&E)$  is low, has a defeater for R. Now obviously the person who believes N&E also believes a lot of other propositions. Perhaps some of those other propositions are such that by virtue of her believing *them* she doesn't get a defeater for R when she believes N&E. Perhaps she has a *defeater-deflector* for the looming defeat of R threatened by  $P(R/N\&E)$  is low and N&E. This could happen if, for example, there were some proposition X she also believes, such that  $P(R/N\&E\&X)$  is not low. Here's an example of a defeater-deflector. Go back to the sheep in the field example of a few paragraphs back. I see what I take to be a sheep in the field: the farmer**

who owns the field comes along and tells me that there are no sheep in that field, but adds that he has a sheep dog who looks like a sheep from this distance. That gives me a defeater. But suppose the farmer's wife had told me earlier on that her husband has developed a thing about sheep and sheep dogs, and tells everyone that there are no sheep in the field, even though there often are. Her telling me this is a *defeaterdeflector*: because I believe what she says, the farmer's comments about sheep and sheep dogs don't give me a defeater for my belief that I see a sheep—a defeater the owner's remarks would otherwise would have given me.

Returning to N&E and R, is there a defeater deflector for the defeat of R threatened by N&E and  $P(R/N\&E)$  is low? Is there a belief X the naturalist might have such that  $P(R/N\&E\&X)$  is not low? Well, it certainly looks as if there are: what about R itself? That's presumably something the naturalist believes.  $P(R/N\&E\&R)$  is certainly not low; it's 1. 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.<sup>35</sup> Which beliefs are such that they can properly function as defeater-deflectors? Which beliefs are admissible in this

context—that is, which beliefs  $X$  are such that if  $P(R/N\&E\&X)$  is not low, then  $X$  is a defeater-deflector for  $R$  and  $N\&E$  and  $P(R/N\&E)$  is low? This is the *conditionalization problem*.<sup>36</sup> It isn't easy to give a complete answer, but we can say at least the following.<sup>37</sup> First, neither  $R$  itself nor any proposition equivalent to it—for example,  $(R \vee (2+1=4)) \& \sim(2+1=4)$ —is admissible as a defeater-deflector here. Second, conjunctions of  $R$  with other propositions  $P$  the naturalist believes—for example,  $(2+1=3) \& R$ —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  $P$  is a defeater-deflector.<sup>38</sup> Finally, no proposition  $P$  that is evidentially dependent upon  $R$  for  $S$ —that is, such that  $S$  believes  $P$  only 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  $R$  (since I believe naturalism is false), as is *either  $R$  or Friesland is larger than the United States*, and *there is some true proposition  $P$  such that  $P(R/N\&P)$  is high*. There is much more to be said, but instead of saying it here, I will refer the interested reader to my paper “Content and Natural Selection.”<sup>39</sup>



**Second final comment: there is a slightly different version of this argument that has somewhat weaker premises; some might find that version appealing on that account.<sup>40</sup> The argument as I presented it above has as a premise that  $P(R/N\&E)$  is low: it is unlikely that our cognitive faculties are reliable, given naturalism and the proposition that we and those faculties have come to be by way of evolution. Here we are speaking of *all* of our cognitive faculties. But perhaps there are interesting distinctions to be made among them. Perhaps some are less likely than others to be reliable, given N&E. Perhaps those faculties that produce beliefs that appear to be relevant to survival and reproduction are more likely to be reliable than those faculties that produce beliefs of other kinds. For example, one might think that perceptual beliefs are often more likely to be relevant to adaptive behavior than beliefs about, say, art criticism, or postmodernism, or string theory. So consider *metaphysical* beliefs—for example, beliefs about the ultimate nature of our world, about whether there are both concrete and abstract objects, about the nature of abstract objects (if any), and about whether there is such a person as God. Metaphysical beliefs don't seem to be relevant to survival and reproduction. And**

of course naturalism is just such a metaphysical belief. This belief doesn't seem relevant to survival and reproduction: it is only the occasional member of the Young Atheist's Club whose reproductive prospects are enhanced by holding the belief that naturalism is true.

So consider the faculty (or subfaculty), whatever it is, that produces metaphysical beliefs, and call it "M." And now we can ask the following question: given N&E, what is the probability that M is reliable? What is  $P(MR/N\&E)$ , where MR is the proposition that metaphysical beliefs are reliably produced and are mostly true? Some people may think this probability is clearly low, even if they aren't so sure about  $P(R/N\&E)$ . If that's how you think about the matter, I propose that you replace the first premise of the argument by

(1\*)  $P(MR/N\&E)$  is low;

everything else can go on as before.

It is time to bring this chapter and indeed this book to a close. I argued in the earlier portions of the book that there are areas of conflict between theism and science (evolutionary psychology for example), but that the conflict is merely

superficial. I went on to argue in [chapter 9](#) that there is deep concord between science and theistic belief; science fits much better with theism than with naturalism. Turning to naturalism, clearly there is superficial concord between science and naturalism—if only because so many naturalists trumpet the claim that science as a pillar in the temple of naturalism. As I argue in this chapter, they are mistaken: one can't rationally accept both naturalism and current evolutionary theory; that combination of beliefs is self-defeating. But then there is a deep conflict between naturalism and one of the most important claims of current science. My conclusion, therefore, is that there is superficial conflict but deep concord between science and theistic belief, but superficial concord and deep conflict between science and naturalism. Given that naturalism is at least a quasi-religion, there is indeed a science/religion conflict, all right, but it is not between science and theistic religion: it is between science and naturalism. That's where the conflict really lies.