Why Think? ∷ *Evolution and the Rational Mind*

RONALD DE SOUSA

OXFORD
UNIVERSITY PRESS

2007

Contents

Chapter 1 ⁘ Introduction

Aristotle called human beings "rational animals." It is all too regrettably obvious, however, that we are frequently irrational. Yet it would be hasty to reject Aristotle's characterization outright. Much of this book is concerned with sorting out how to make sense of both our rationality and our irrationality. It is also about what's good about being rational, and why it's worth the trouble.

To make a start on the latter question, consider Jack and Jill. When Jill tackles a project, she is methodical and scrupulously careful. She tailors her means to her ends. She looks only to the best evidence and the soundest reasons. She is, then, you will agree, as rational as one could be. Yet she fails. Jack, on the contrary, is devoted to Non-Linear Thinking, which he interprets as requiring regular consultation of astrological charts, the hexagrams of the *Yi Jing*, and other magical omens. Rationality, he declares, is overrated. Irritatingly, he succeeds in what he attempts and loudly trumpets his success.

Such things do happen. When they do, isn't it enough to throw you right off the claims of rationality? What is it, actually, that is so good about rational thinking?

This may seem an idle question. Surely the advantages of thinking are obvious. And yet no product of human ingenuity can hold a candle to the subtle and economical complexity of a single living cell, let alone to the unfathomable organization of what is often termed the most complicated object in the universe, the human brain. Thought nowhere figures in the mechanisms of evolution that have shaped life itself. Nor does it play any part in the procedures used by most organisms to keep themselves alive. Such marvels are not the fruit of any computation or planning: they are merely the upshot of four billion years of natural selection, constrained by the laws of physics, chemistry, and probability. The precise details of the diversity of mechanisms involved in natural selection are still a matter of dispute, but in the main they are adequately summed up in the phrase made famous by the biologist Jacques Monod (1972): *chance* and *necessity.* Nature abounds in astonishing inventions such as the human eye, or the intricacies of the mechanisms that turn food into over three hundred different kinds of cells that make up our bodies. The proponents of the theory of Intelligent Design love to cite these, but they keep having to pick new candidates as science cracks one mystery after another. When a favorite example of the inexplicable is explained, it must be replaced with a new mystery. If the "irreducible complexity" of anything still unexplained had been consistently used to posit the intervention of an Intelligent Designer, evolutionary science would have been abandoned as a waste of time before it started. The wonder of nature's ingenuity rests precisely on the assumption that her most ingenious devices are all natural products of evolution, owing nothing to intelligence. What, then, is the point of thinking?

In approaching such questions, we should first remind ourselves that rationality does not guarantee success. Its advantage consists merely in increasing the *chances* of success. This brings rationality right into line with evolution, of which the very stuff, we might say without much exaggeration, is *probability*. Natural selection has perceptible effects only in the context of large numbers. At the level of statistical phenomena, probability governs the precise interactions of chance and necessity. As for individuals, no matter how well equipped they might be to seize opportunities and face the dangers that threaten them in every natural environment, survival is never guaranteed. What biologists call an organism's *fitness*, its *probable* survival and fertility, guarantees neither its survival nor its fertility. No more, for that matter, than success is assured even in the most minutely planned of intentional undertakings. In both cases, the most we can claim is that the best-adapted organism no less than the most elaborately worked-out plan will be the one most *likely* to succeed. This fact will translate into meaningful observable effects only in the long run, at the statistical level.

In this essential respect, then, the upshot of rational planning elaborated in intelligent thinking is the same as the upshot of natural selection: in the long run, individuals increase the chances of success in their respective undertakings. Furthermore, there is every reason to think that the methods used by rational beings such as we pride ourselves on being have themselves been shaped over millions of years by natural selection. This process took place over an enormous variety of circumstances—when our ancestors lived in the ocean and when they lived on land, when they had to succeed in catching prey and when they had to avoid their own predators. Should we then assume that our strategies of inference and discovery are invariably the best they could be? If not, can we fall back on the thought that they are generally adequate, if not actually optimal? Or should we, on the contrary, resign ourselves to the possibility that the most seemingly "natural" epistemic processes are often ill adapted to the circumstances of present day life? If the more pessimistic hypothesis is right, can we at least still count on our capacity for self-improvement? Given the way we came by our faculty of thought, what reasonable expectations are we entitled to?

The questions I have raised so far are of two kinds. A first batch takes the powers of rational thought for granted. Rational thought is set up as a benchmark, by reference to which we might assess the mechanisms of natural selection. The second batch, by contrast, takes natural selection as its point of departure, in order to question the viability and reliability of our modes of discovery, our rules of inference, and our standards of proof—in short, of all the epistemic strategies that natural selection has empowered us to devise and endorse. Thus the evolutionary point of view suggests two perspectives: one looks at the logic of natural selection that gave us adaptive functions, while the other scrutinizes the origins and the constraints on the rationality of thought and action that supposedly characterizes intelligent human beings. These two perspectives form the framework of what follows.

At the heart of both is the idea of rationality. Let me then begin by attempting to cast a little light on the significance of that notion. Rationality is generally thought to be a good thing, although the occasional dissenting voice is heard to deplore it as rigid, narrow, linear, or even—most horribly—"phallogocentric." What does *rationality* actually mean?

1.1 Two Senses of "Rationality"

At first sight it seems obvious that the ascription of rationality is confined, like its opposite *irrationality*, to thought and action and to organisms capable of both. Talk of rationality is not appropriate in connection with events governed purely by the laws of physics, even if such events involve a rational being. Suppose a man accidentally stumbles and falls into a clump of nettles. We wouldn't label him "irrational," for the incident was not a chosen act. It was a mere event, implicating the person not as an agent but merely as a physical object, subject to the laws of gravitation and inertia. We speak of a falling object as "obeying" the law of gravitation, to be sure, but *disobeying* the law of gravitation isn't really an option. That sort of obedience is neither rational nor irrational.

What this example brings out is that the word *rational* has two senses, marked by two different contraries. In the *categorial* sense, the contrary of *rational* is *arational,* a term that applies to behavior that is due neither to choice nor to thought. The notion of choice, in this context, implies nothing in particular about deliberation or free will, but merely refers to events that are caused in a certain way. For an occurrence to be a matter of choice in the sense intended, its causes must include *reasons*. Reasons, at a first rough level of approximation, provide explanations by appealing to certain goals, norms, or values.

The second, *normative* sense of the word *rational* contrasts with *irrational*. It implies that a belief or behavior was appropriately grounded in specific reasons, norms, or values. In this second sense, an agent who is not rational is in some sense defective in respect of thought or action. Irrationality is a normative notion: its ascription commonly involves a certain sort of reproach, complaint, or criticism. What sort of criticism is a question that will require close scrutiny. For one can criticize a landscape for being dull, or a fruit for being unripe, but complaints of that sort ascribe nothing like irrationality to landscapes or to fruit.

On pain of paradox, the word *rational* cannot be taken in its normative sense in Aristotle's characterization of humans as rational animals. The formula makes perfect sense, however, if it is interpreted in the categorial sense. Which is to say that if human beings can indeed be described as rational animals, it is precisely in virtue of the fact that humans, of all the animals, are the only ones capable of *irrational* thoughts and actions.

The distinction just drawn gives rise to a difficulty, however. If categorial rationality cannot appropriately be ascribed to events that are sufficiently explained in terms of natural laws, does this mean that human behavior escapes the determination of natural laws altogether? One might take this in either of two senses. On a more modest interpretation, it would mean that the laws of physics, chemistry, or any other science—including the laws of probability—that explain the behavior of inanimate objects are insufficient to explain that of rational beings. Rational behavior would then belong in a zone left fallow by the laws of nature and mathematics. Someone might offer the behavior of a chess player as an instance of something that can be explained only in terms of the rules of the game, and rules are not laws of nature. A stronger version would insist that the behavior of rational beings actually transgresses some natural laws. But that thesis would be absurd because to claim that a "transgression" of laws of nature has occurred is to posit a miracle. Or, more reasonably, it amounts to an admission that we hadn't got the alleged laws quite right in the first place.

Some philosophers, such as Kant and Bergson, have clung to the thought that free will transcends the natural world without actually violating the laws of nature. But this attempted solution is bred in bad faith. For talk of transcendence is generally a way of trying to paper over a contradiction with a spot of jargon. Better to acknowledge that regardless of intelligence or rationality, human beings are indeed subject, like everything else, to the laws of nature. The human difference must be sought among natural facts, and not in some hope that natural facts might be transcended.

The evolutionary perspective maintains that life arose about four billions years ago from chemical conditions that are still not fully understood, but of which one can safely presume that they included no phenomena that could be labeled either rational or irrational. It follows that at some point—or perhaps gradually, during a long transitional phase—phenomena that could be classed as rational succeeded others that could not reasonably be so labeled. By similar reasoning, a transition of the same kind must be supposed to take place in the course of development in every individual human organism. For each of us begins life as a single-celled organism, the zygote that results from the fusion of the parental gametes. As that cell and its descendents undergo successive divisions, according to the laws of physics and chemistry that govern those processes, they undergo a series of metamorphoses that at some point gives rise to an organism capable of reasoning, that is, a rational being in the categorial sense.

If we start from the thought that rationality is typically applicable to thought and to action, we can characterize two crucial metamorphoses, both in the evolutionary process and in the course of individual development. One took us from the mere detection of stimuli to the capacity to represent objects; the other took us from tropisms, or automatic behavioral responses, to the capacity to form and act on desires and intentions.

From Detection to Representation

Each living cell, and therefore every multicellular organism, is endowed with some capacity to detect what might be useful or harmful to it. One could call this "sensibility," but the notion I have in mind is meant precisely to contrast with the ideas of consciousness and knowledge evoked by this word. It is better to speak simply of a "detecting function" in order to underline the purely functional character of the faculty in question. The existence of a transition between the detecting function and its rational successor then raises the following questions: At what stage of phylogenetic evolution, and at what stage in the development of each adult to whom rationality is unquestionably ascribed, must we speak no longer of simple detection, but of belief, knowledge, or representations? What are the supplementary capacities that are crucial to this transformation, and how do they arise?

From Tropism to Desire

Every unicellular animal is equipped with a detecting function, on which some specific behaviors depend. In the simplest organisms, this will merely result in approach or avoidance. Although the terms *approach* and *avoidance* may seem to imply a greater measure of mobility than plants can claim, even plants react, if only with a simple change of orientation, the opening of some pores, or the tensing of certain fibers, such as enable the sunflower to track the position of the sun. What counts is that there should be some sort of differentiated behavior corresponding to the information detected. Behaviors of this sort are called *tropisms*, and they are typically triggered by a gradient of temperature, light, chemical concentration, or other stimuli in relation to which the organism orients itself.

Tropisms, like other adapted functions, are the creatures of natural selection. They fulfill tasks essential for the survival of the organism whose goals they serve. Explanation in terms of goals is called *teleological*, so this means that tropisms are liable to be explained in teleological terms. But that word, *teleology*, is rife with potential misunderstandings. When we think of biology in terms of *teleology* and *goals*, are we using these terms in the same sense as when they are used in connection with voluntary decisions and intentional behavior?

If one says of a cell that it *seeks* an environment at a certain temperature, or that it *desires* a certain chemical, one would surely be using these terms in a metaphorical sense. But why are we so sure? What really differentiates a full-fledged desire from a simple tropism? Or to put the question differently, what needs to be added to a tropism to turn it into a desire? This is just another way of posing the question just mentioned: when exactly—on the scale of phylogenetic evolution or on the scale of individual development and as a consequence of what changes in the capacities of an organism—does it become appropriate to speak of desires, projects, and intentions? What precisely make it legitimate to ascribe rationality, and not merely biological functionality, to a given process?

As we have just seen, the categorical notion of rationality implies the possibility of criticism. Three sorts of reproaches, in particular, are appropriate only when they are addressed to a rational agent: it makes sense to criticize a person, but not a cell, for having made a mistake in *computation*, or with having failed to *foresee* what should have been foreseen, or with having acted on *reasons* that fell short of the best set of reasons. We need to ask, then, about the baggage carried by that trio of words: *computation*, *foresight,* and *reasons*.

This last term is more likely to make trouble than to help. I will pass over it for now, noting only that its kinship with the Latin *ratio* evokes both *proportionality* and *accounting*.

As for the concept of foresight, it seems it could just about be stretched to apply to certain tropisms. A chemical gradient might be said to allow a cell to "foresee," if only in a metaphorical sense sufficient to license prediction of behavior, what it is likely to encounter in one direction or the other. The difference we seek is therefore not likely to be found in the idea of foresight. More likely to be helpful is the consideration that when complaining or criticizing is appropriate, some sort of *norm* must be involved, where a norm is roughly a notion of how things are *supposed* to be. It will therefore be in the neighborhood of this idea of *appropriate criticism* that we are most likely to locate the frontier of normativity, which will allow us to cross into the domain of rationality. We'll have occasion to look into this idea of appropriate criticism. But first, let us look further into the remaining concept in the trio just mentioned: computation.

1.2 What Is Computation?

These days we are used to computers regulating more and more aspects of our lives, so it no longer seems surprising that machines are able to effect computations. But René Descartes (1596–1650), one of the first philosophers who thought seriously about the difference between people and machines, would have been astonished. For he thought of computation as a manifestation of the faculty of reason, and he thought of reason as belonging exclusively in the province of the immaterial soul. It made no sense, Descartes maintained, to attribute the faculty of reason to any sort of material or mechanical device. Some three centuries later, our machines are rather bad at such animal functions as seeing, and they remain awkward in their attempts to get around on two feet. By contrast, they compute all kinds of things with ease, and the best machine is unbeatable at chess, the paradigm of games of intelligent computation.[1] When a machine effects a computation, should we think of it as computing in the very same sense as we might say of chess masters that they compute the next move? This is hotly denied by many champions of the unique human difference. But actually the question glosses over an important distinction between two very different sorts of computing machines: classical *digital* computers and *analog* computers.

To get a sense of that crucial difference, consider what we would think of someone who, after watching Galileo drop stones from the top of the tower of Pisa, offered the following account of the event:

> The stone *computes*, in accordance with the law soon to be formulated by Newton (a formula the stone knows innately), the speed it is to adopt at every instant of its fall.

> Simultaneously with this computation, the stone *implements* the motion determined by the result of the computation.

Most of us would assume that this description is meant to be metaphorical or just facetious. When the stone is said to "obey" the law of gravity, that simply means that its trajectory is adequately described by that law. We *use* the law to make the relevant calculations, determining the speed the stone will have reached when it hits the ground. But obviously the stone itself neither computes anything nor executes any plan.

An object that conforms to the law of gravitation can be used, however, to measure or compute something else. A pendulum's behavior is computable from the law of gravity with the help of some geometry and calculus. That provides us with a measure of time, which allows us to "compute" the interval elapsed between two given events. In this way, the pendulum provides a simple example of an analog computer. Another example—though not so obviously useful—is the humble soap bubble: its shape automatically minimizes the surface of a volume of gas, not on the basis of any digital computation, but merely by virtue of the implementation of a physical process.

Similar principles are embodied in self-regulating devices of various sorts. One particularly interesting example is James Watt's governor. This device solves the following problem: how to pinpoint the moment when the speed of a steam engine becomes dangerously high, and slow the engine down to prevent it from racing out of control. Nowadays, we might think of solving this problem by means of a computer equipped with three distinct modules. Call it SPEEDWATCH. A first module would detect the number of revolutions per minute effected by the machine. This module would pass on the information, in digital form, to a second module, which would compare the value acquired with a threshold programmed ahead of time. Once the threshold is reached, a message would be passed on to a third module that controls the pressure in the boiler. That module's task would be to lower the pressure and hence the engine speed. This would be, very roughly, an information-theoretic solution. But Watt's governor has nothing to do with information in any form. It does not detect, compare, or transmit information. Instead it functions purely mechanically.

Watt's governor consists in a revolving central shaft, to which are hinged two wings with weighted tips. As the speed of the machine increases, the central shaft revolves more quickly. The centrifugal

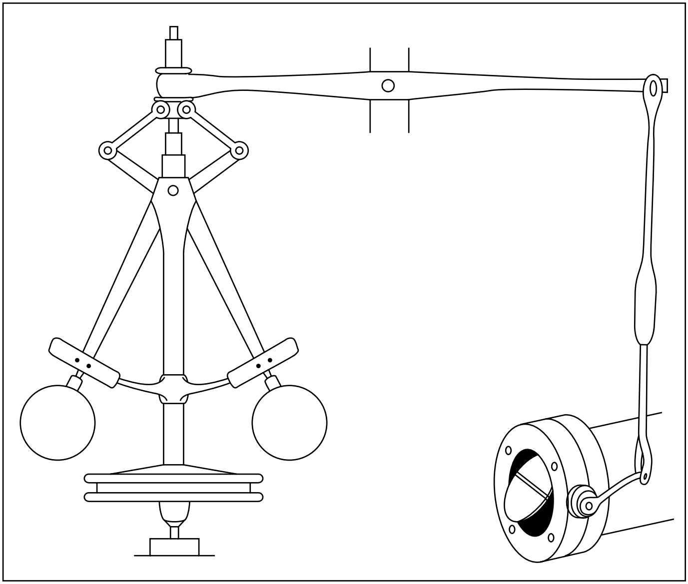

Figure 1.1. Watt's governor

force induced by this rotation lifts the wings, counteracting the pull of gravity that keeps them, when at rest, parallel with the central shaft. As the wings reach up toward the horizontal, they trigger the opening of a valve that allows steam to escape and lowers the pressure. The machine slows down. As the speed eases, the wings come down and the valve closes, which drives the pressure up again. The cycle begins again, keeping the machine close to the target speed.

This efficient and simple device can be thought of as an artificial tropism. Its elegance far surpasses the rather complicated digital-informational alternative described a moment ago. The comparison shows that digital devices are not necessarily superior in all domains, and that computation comes into its own only in certain specific circumstances. We'll have to look into what those circumstances are. The more general lesson to be drawn at this point is that a method appropriate to the solution of some problems at a given scale is not necessarily the best method in all other circumstances and at different scales.

That fact may have set an obstacle in the way of biological evolution. For past a given threshold, a once successful device may block further improvement. Natural selection may then have no means of switching to an entirely different strategy: it would be nice to fly unaided, for example, but we make too much good use of our arms: no amount of natural selection will cause them to morph into wings. That's why prosthetic wings were our only option for emulating birds. (But then those prosthetic wings easily outstripped the natural ones.)

Does that mean rationally designed technology can overcome any problem? No: rational deliberation can face a similar difficulty. In many practical circumstances it would be counterproductive to embark on minutely detailed planning. When a quick reaction in real time is imperative, an approximate but fast response will be more effective in the long run than an exact solution that takes time to work out. By the same token, however, the quick response may lead us astray. This is an inescapable dilemma of rationality. As we shall see, it has many interesting consequences.

Watt's governor is an analog machine. Strictly speaking, it performs no computations. Its functioning rests merely on physical laws that govern all dynamic systems—whatever involves forces, velocities, and masses, the relations among which are summed up in Newton's formula "Force = Mass × Acceleration." In a way, the calculating machine or computer does nothing different. It, too, has been conceived and built to behave in a certain way in the light of the laws of electricity and the logic of its circuitry. And yet surely the distinction between the two does not depend merely on the difference between dynamics and electricity. So what is the essential difference?

1.3 Digital and Analog

The best way to sum up the nature and the advantages of digital representation is to recall the paradoxical theory of resemblance advocated by Plato, known as the *Theory of Forms*. The best illustration of the core idea is provided by our humble roman alphabet. Take two tokens that resemble one another, say a and *a*.[2] Is the resemblance between them a two-term relation? Obviously yes, says common sense. But Plato held that it should really be treated as a three-term relation: what the resemblance between a and *a* consists in, he proposed, is the fact that both *derive* from a third entity, call it A, which is the "ideal" A and which is the depository of the essential nature of both a and *a*. There are, in fact, a thousand different fonts and styles of type, of which the tokens of 'a' resemble one another less than they might resemble other letters.[3] What all instances of the letter a have in common is just that they meet whatever norms it is that constitute them as concrete instances of that particular vowel. In practice, such norms are often ill conceived, and so we are apt to use contextual clues to avoid confusing the number 1, the uppercase ninth letter of the alphabet, I, and the lowercase twelfth letter, l. But in a well-planned and designed digital system, the elements of the system would be sufficiently well spaced as to make mistakes virtually impossible.

As everyone knows, computers of the sort many of us now use every day are based on a system of digital representation. What that means is that the voltage changes effected in the computer's circuitry are regarded, for the purposes of computation, as taking only one or another of a finite predetermined number of discrete values. From the purely physical point of view, of course, the voltage changes are effectively continuous. This illustrates the importance of distinguishing between the characterization of the physical processes themselves, and the way those characteristics are interpreted when they are set up to be part of a digital representational system.

Watt's engine governor shows that a mechanical or dynamic system, in which all effects derive simply from behavior that conforms to the laws of nature, can provide the most effective solution to certain classes of problems. By contrast, a device based on digital representation comes into its own whenever there is a need for a great many faithful reproductions of a given original. The reason is that all dynamical systems are subject to small variations and generate small errors of measurement at every stage. Whenever a sequence of copies is required, those tiny initial errors will cumulate and turn into serious discrepancies, as in successive xeroxes of xeroxes that quickly become illegible. In a digital system of representation, by contrast, you are never further than two steps away from the original. For in a digital setup, as in

Plato's theory, you don't really copy a copy, or indeed even the original. Instead, you copy the "idea," the paradigm, of which the original was itself only an instance or copy. Each copy derives, not from the product of previous copying operations, but from the original itself. Digital systems of representation and reproduction afford *almost perfect* copying fidelity in arbitrarily long series.

DNA, the digital "language" of genes, affords a remarkable illustration. Some sequences have remained almost the same for over a billion years. Some of our own most important genes, such as the "Hox genes" governing the basic body plan of bilateral creatures, have remained almost unchanged for some two billion years. They first operated in organisms that are the ancestors both of humans and of insects (Carroll 2005). (The *almost* is essential, however, for without the occasional copying error, we would all be single-celled organisms, like our very distant cousin the amoeba. Absolutely perfect copying would have allowed no mutations, and therefore no evolution.)

The degree of fidelity with which organisms reproduce is due in large part to the digital "language" of the genes, which guarantees an infinitesimally small mutation rate.[4] The analogies with actual language are striking. First, we can speak of an *alphabet*, constituted by the four bases of DNA or RNA—cytosine, thymine (or uracil in RNA), guanine, and adenine. Taken three at a time, these yield possible combinations that suffice to constitute enough *words* to specify each of the twenty-odd amino acids of which all proteins are made; those words, in turn, link into *sentences* that determine the immense variety of possible proteins. Finally, by analogy with *discourse* built up of sentences, we can think of the proteins making up all the works of biological nature; that is, of all living organisms, as well as the enzymes that are essential to the elaboration and differentiation of their organs.

1.4 Individuals and Communities

The elaboration of a mature organism, partly on the basis of the information provided by its DNA in the presence of favorable environmental conditions, requires a complex and precise collaboration between the individual cells that make it up. This collaboration shouldn't be taken for granted. How did some organisms come to diverge from the condition of their remote ancestors, living as independent individual cells fending for themselves, to arrive at the condition of such complex differentiated multicellular organisms? However it happened, that was assuredly one of the "major transitions" of evolution (Buss 1987; Maynard Smith and Szathmáry 1999).

This observation serves to remind us of another contrast, which cuts across that drawn above between the effects of natural selection and those of purposeful action. This is the contrast between individual rationality and group rationality. It often seems natural enough to apply the concepts of rationality and irrationality to groups or societies. Indeed, some philosophers have spoken of "collective selves" or "plural subjects," capable of formulating and executing genuine group intentions (Gilbert 1992). But as we shall see, the relation between individual rationality and group rationality raises difficult problems, no less at the theoretical than at the practical level.

Individual rationality seems more straightforward; yet we can also think of a single individual as a sort of community. Plato claimed that the soul has three parts, often at odds with one another. Freud favored a rather similar metaphor and distinguished ego, id, and superego. And at the level of the body of any multicellular organism, the viability of the whole presupposes a complex cooperative system linking the myriad cells of which it is composed. From that perspective, cancer can be thought of as the consequence of a break in that collaboration: the refusal, by a certain lineage of cells, to continue subordinating themselves to a coherent whole, for the sake of a proliferation that benefits only itself.

1.5 The Origins and Limits of Intelligence

At first sight, the difference between the products of intelligent thought and those of natural selection seems obvious enough. Thought boasts a crucial differentia. When pressed to declare what this is, some philosophers favor *intentionality*, others, *consciousness*. But most would accept that the differentia, however it should be characterized, amounts to the presence of *mind* or the *mental*. But what precisely is the mental?

To consider answering this question in a naturalistic framework is to entrust the analysis of the evolution of mind, as a product of natural selection, to mind itself. Should we worry that this makes mind into both judge and party? That might raise a doubt about the reliability of the naturalist perspective. Such is the charge made by the theist philosopher Alvin Plantinga.[5] His challenge can be paraphrased as follows:

> If indeed mind has its origin in natural selection, we have no reason to believe that our mental faculties are capable of uncovering the true nature either of mind itself or of the process of natural selection that allegedly gave rise to it. For it evolved under circumstances that couldn't possibly reward the ability to reveal such truths.

The ingenuity of this challenge lies in its attempt to use the very tools of naturalism against itself. It impugns not natural selection as such but the compatibility of natural selection with our capacity to know about it on the basis of rational scientific inference.

In reply, one might first point out that the theist alternative Plantinga touts has its own problems. For let us suppose we knew that our brain was fashioned by some creating intelligence rather than by natural selection. Nothing of interest could be inferred from that, unless we knew the *intentions* with which the creator fashioned it. But the theological premises required to deduce that the creator intended to favor us with a faculty for blanket truth-discovery would be entirely arbitrary.[6] Long before the hypothesis of natural selection became available to provide a concrete alternative, David Hume ([1779] 1947) had already perceived that the inevitable uncertainty about the putative creator's intentions suffices to knock theism out of the field of serious contending hypotheses.

Nevertheless, Plantinga's challenge remains all the more intriguing if it is regarded independently of the theological leap he urges us to make. Stripped of theology, we can reformulate it thus:

> The circumstances under which natural selection honed our mental equipment were presumably such as to afford our ancestors solutions to the practical problems of survival they confronted. We have no reason for thinking this same equipment is epistemologically reliable in situations that differ materially from those faced by our distant ancestors.

Actually Descartes had already come up with essentially the same worry over three centuries ago. In the fourth *Meditation*, he warned that we risk falling into error if we insist on inferring significant propositions about objective reality on the basis of the information provided by our senses (Descartes [1641] 1986). The only thing of which we may be reasonably certain (given God's benevolence) is that our senses will provide sound guidance for getting around in the practical world. But nothing can be inferred about the actual nature of the world itself.

Plantinga's challenge differs only in one particular from his predecessor's warning. For Descartes, God's goodness warrants two things. First, it allows us to count on the utility of the senses, but not on the truth of any ontological inference we might make from their deliverances. Second, it warrants the correctness of whatever we conceive clearly and distinctly. For Plantinga, even in my theology-free version, evolution can take over the first warranty but is very far from being able to underwrite the second. We are therefore in a weaker epistemological position than Descartes imagined. For we have no good reason to believe that what seems clearly obvious to us is actually true. Consequently, we can do no better in our quest for truth than to trust what seems to us most probable after we have looked into a question as thoroughly as possible. In any given case, we must admit the possibility that we might be mistaken. It is therefore crucial that we respond to Plantinga's challenge. If we cannot, we will remain massively vulnerable in all domains in which our reason ventures beyond the observable facts. We need to set out the reasons we have for trusting the methods of rational thought that evolution has bequeathed us.

1.6 How to Meet Plantinga's Challenge

Plantinga speaks to a view that is currently popular with a great many epistemologists, under the name of "reliabilism." Instead of searching for unshakable foundations of knowledge and rationality, reliabilists are merely committed to the idea that natural selection has so tuned our faculties as to make it likely that our search for truth will have a happy outcome in most practical and scientific contexts. More specifically, some have speculated that the reliability of reason is the fruit of an interaction between natural selection, which has shaped our individual intellectual capacities, and the effects of social intercourse, particularly as enhanced by the invention of language. On this view, the confidence we have in the opinions of others would play an ambiguous role, somewhat analogous to that played by our predisposition to make inferences from our experience about what lies beyond it. Induction—as such inferences are called—generates superstitions as well as correct predictions. But one cannot hope to avoid superstition without at the same time increasing the risk of missing certain significant correlations. Similarly, our tendency to believe what others tell us can be harmful as well as useful, depending on the circumstances. But if one were to try to protect oneself against the bad effects of credulity, one would immediately deprive oneself of the great advantages of submitting to mutual criticism. We need to tread a fine line: too much trust can be as harmful as too little. As we will see in chapter 4, it would be rash to expect natural selection to have set up a social contract that guarantees truth-telling. And in chapter 5, I shall explain how, although true belief is not the inevitable consequence of social conformity, conformism might nevertheless have secured the minimum of homogeneity to allow for some kind of group selection, which may at least sometimes favor strongly cooperative practices.

Despite its plausibility, I do not believe that reliabilism provides an adequate response to Plantinga's challenge. Without claiming a high degree of certainty, a slightly more specific hypothesis seems to me more promising. This hypothesis looks to mathematics, regarded as an extension of ordinary language. The argument rests on two propositions: (a) mathematics is unique from a methodological point of view; and (b) its usefulness to scientific and technological progress provides an independent argument in favor of its claim to present us with objectively correct representations of an external world.

Let me expand on each of these two points.

(a) On the methodological front, it is obvious that the development of our brain could not possibly have been influenced by selection pressure specifically favoring the faculty to do pure mathematics. Natural selection could not have gotten a grip on the ability to do mathematics as such unless that capacity manifested itself in some individuals while failing to do so in others. But mathematics is a recent invention: as far as we know, no one did pure mathematics until long after the human brain had attained essentially its present size and capacities, no more than a few thousand years ago. On the evolutionary scale, mathematics is part of our present rather than of our evolutionary past. It is therefore out of the question for mathematical talent as such to have been a factor in evolution by natural selection. Unless it is a faculty that somehow altogether escapes the constraints of biology, the mathematical faculty had to develop as a mere side effect of some other, more immediately useful, set of intellectual capacities in our ancestors. The mathematical gift remained long in the shadows, as it were, behind more visible talents, until the time it emerged fully mature, like Athena out of the head of Zeus.

(b) Once mathematics had emerged into the light of day, there was still nothing to guarantee that it could prove useful outside the domains in which our practical skills had already been operating for millennia. And yet, pure mathematics notoriously finds all kinds of startling applications in the solution of technological and scientific problems that our ancestors could not possibly have conceived of, and it does so by generating theories that would have remained wholly unintelligible to them. That strongly supports the idea that mathematics can uncover aspects of the universe of which neither the usefulness nor even the existence could possibly have been manifested in the environment of our evolutionary adaptation (EEA) in which the basic functions of the brain were being shaped by natural selection. As Wigner (1960) has argued, this constitutes at least prima facie evidence for the conclusion that the truths of mathematics do not merely reflect projective constructions of our brains, but probably correspond to an objective reality. This fact remains deeply puzzling, to be sure, since it can't have been a direct consequence of any sort of tuning of our brains to the world by natural selection. But the mystery is scarcely likely to be cleared up by being ascribed to divine intervention. The theological hypothesis is no real alternative. The upshot is that we can pursue the present inquiry without fear of its being reduced to absurdity either by Plantinga's attempt to show that the belief in natural selection undermines itself, or in virtue of the trickery of some evil demon.

Despite this methodological reassurance, there are compelling reasons to regard our thought processes as leaving much to be desired. On the basis of apparently solid premises, using unimpeachable rules of inference, we can be led to insoluble paradoxes, some of which have kept philosophers busy for centuries. How, asked Zeno of Elea, can I get from A to B in a finite time, since I must get halfway, then halfway again, an infinite number of times? If I assert: "What I am now saying is false," have I lied? Some types of problems, particularly those in which we are required to reason about probability, are liable to lure even individuals trained in the art of careful thinking into drawing systematically erroneous conclusions. Such pathologies of reasoning challenge us to inquire into the causes of such mistakes. In addition, they invite even more disconcerting questions: how could it be that natural selection should have shaped our minds in such a way that we are systematically prone to error? Does this supposition not contradict what we assume we know about natural selection, namely, that it leads to adaptation and thus optimizes fitness by perfecting the organs and capacities on which it works? Or must we fall back on the exculpatory supposition that, as a relatively young species, our faculties have still a long way to go before they reach the perfected state for which they are headed? Shall we need another few hundred thousand years for the kinks in our mental powers to get ironed out? That would be an intellectually flaccid way to explain away the mental deficiencies that have been observed in our species: let's hope for a more interesting approach.

1.7 Prospect

Adaptation, of organs to their function and of organisms to their ecological niches, is a particularly striking feature of the living world. That observation, however, often encourages a hasty inference to the conclusion that teleology is ineliminable from biology. It can seem undeniable that in order to explain the behavior of every organ and every member of a biological community—organism, hive, or ecological web—we must refer to the teleological aspects of its organization.

Chapter 2 will be devoted to the examination of that presumption of teleology. I shall argue that although teleology is not entirely banned from biology, it subsists only in what I shall refer to as a "vestigial" form. In that vestigial or degenerate form, it is still capable of explaining why nature produces such a powerful impression of being pervaded by inherent teleology. But teleology in that form is wholly distinct from that which is in question when we speak of the goals, purposes, and values of individual human agents.

In chapter 3, I shall return in more detail to the resemblances and differences between the "methods" of natural selection and those of reflective thought. These comparisons will lead us to consider the difference between digital and analog systems of representation. I shall argue that the distinctive characteristics of human thought rest essentially on the digital features of language, and that the novel capacity for explicit thought afforded by language enables the proliferation of individual values. This proliferation of values not only detaches human goals entirely from the vestigial "goals" of nature but also generates conflicts, both within and among human individuals as well as communities.

In chapter 4, I shall further explore the relation between individual and collective rationality. A sketch of some of the better-known models of emergent collective organization on the basis of purely individual interactions will serve to illustrate some of the advantages, problems, and limitations of the move from individual to collective rationality. That will bring me to one of the more heated controversies in recent theoretical biology. That is the debate concerning the units of selection, group selection, and the main explanations that have so far been adduced to explain the paradox of individual altruism: how is the Darwinian hypothesis of a universal "struggle for survival" compatible with the fact that human beings sometimes actually sacrifice themselves for a cause that benefits only others?

In chapter 5, my central protagonists—evolution and rationality—will flip over and trade places. Having looked, in previous chapters, at the aspects of natural selection that seem most akin to intelligent planning, I shall turn to the evolution of our rational faculty for intelligent planning. I will bring forward a sampling of some of the examples of systematic irrationality that have been brought to light by psychological research and will explain the apparent paradox that the much-vaunted human differentia of rationality, in the categorial sense, is actually grounded in our capacity to manifest irrational thought and action. I will show how even the most extreme case of irrationality presupposes a minimal level of normative rationality. We shall also see how the most notorious cases of irrationality can be explained on the basis of three factors: the modular organization of our capacities; the widening gap between the vestigial "goals" of natural selection and the concrete goals that human agents set up for themselves; and the indispensable yet ambiguous role played by emotions in the economy of rationality. Although our cognitive predispositions do entail serious deficiencies in a number of types of concrete situations, that is no reason to expect that future evolution will put this right by bringing a higher level of perfection to our cognitive powers. For in the great majority of cases, our faculties' defects are just the flip side of their virtues.