Also by Daniel C. Dennett

Content and Consciousness
Brainstorms
Elbow Room
The Intentional Stance
Consciousness Explained
Darwin's Dangerous Idea

KINDS OF MINDS

Toward an Understanding of Consciousness

DANIEL C. DENNETT

A Member of the Perseus Books Group

CHAPTER 3

THE BODY AND ITS MINDS

> In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation. Light will be thrown on the origin of man and his history.
>
> Charles Darwin, *The Origin of Species*

FROM SENSITIVITY TO SENTIENCE?

At last, let's take the journey. Mother Nature—or, as we call it today, the process of evolution by natural selection—has no foresight at all, but has gradually built beings with foresight. The task of a mind is to produce future, as the poet Paul Valéry once put it. A mind is fundamentally an anticipator, an expectation-generator. It mines the present for clues, which it refines with the help of the materials it has saved from the past, turning them into anticipations of the

future. And then it acts, rationally, on the basis of those hard-won anticipations.

Given the inescapable competition for materials in the world of living things, the task facing any organism can be considered to be one version or another of the childhood game of hide-and-seek. You seek what you need, and hide from those who need what you have. The earliest replicators, the macromolecules, had their needs and developed simple—*relatively* simple!—means of achieving them. Their seeking was just so much random walking, with a suitably configured grabber at the business end. When they bumped into the right things, they grabbed them. These seekers had no plan, no "search image," no representation of the sought-for items beyond the configuration of the grabbers. It was lock-and-key, and nothing more. Hence the macromolecule did not know it was seeking, and did not need to know.

The "need to know" principle is most famous in its application in the world of espionage, actual and fictional: No agent should be given any more information than he absolutely needs to know to perform his part of the project. Much the same principle has been honored for billions of years, and continues to be honored in a trillion ways, in the design of every living thing. The agents (or microagents or pseudoagents) of which a living thing is composed—like the secret agents of the CIA or KGB—are vouchsafed only the information they need in order to carry out their very limited specialized tasks. In espionage, the rationale is security; in nature, the rationale is economy. The cheapest, least intensively designed system will be "discovered" first by Mother Nature, and myopically selected.

It is important to recognize, by the way, that the cheapest design may well not be the most efficient, or the smallest. It may often be cheaper for Mother Nature to throw in—or leave in—lots of extra, nonfunctioning stuff, simply because such stuff gets created by the replication-and-development process and cannot be removed without exorbitant cost. It is now known that many mutations insert a code that simply "turns off" a gene without deleting it—a much cheaper move to make in genetic space. A parallel phenomenon in the world of human engineering occurs routinely in computer programming. When programmers improve a program (creating, say, WordWhizbang 7.0 to replace WordWhizbang 6.1), the standard practice is to create the new source code adjacent to the old code, simply by copying the old code and then editing or mutating the copy. Then, before running or compiling the new code, they "comment out" the old code—they don't erase it from the source code file but isolate the old version between special symbols that tell the computer to skip over the bracketed stuff when compiling or executing the program. The old instructions remain in the "genome," marked so that they are never "expressed" in the phenotype. It costs almost nothing to keep the old code along for the ride, and it might come in handy some day. Circumstances in the world might change, for instance—making the old version better after all. Or the extra copy of the old version might someday get mutated into something of value. Such hard-won design should not be lightly discarded, since it would be hard to re-create from scratch. As is becoming ever more clear, evolution often avails itself of this tactic, reusing again and again the leftovers of earlier design processes. (I explore this principle of thrifty accumulation of design in more depth in *Darwin's Dangerous Idea*.)

The macromolecules had no need to know, and their single-celled descendants were much more complex but also had no need to know what they were doing, or why what they were doing was the source of their livelihood. For billions of years, then, there were reasons but no reason formulators, or reason representers, or even, in the strong sense, reason appreciators. (Mother Nature, the process of natural selection, shows her appreciation of good reasons tacitly, by

wordlessly and mindlessly permitting the best designs to prosper.) We late-blooming theorists are the first to *see* the patterns and divine these reasons—the free-floating rationales of the designs that have been created over the eons.

We describe the patterns using the intentional stance. Even some of the simplest design features in organisms—permanent features even simpler than ON/OFF switches—can be installed and refined by a process that has an intentional-stance interpretation. For instance, plants don't have minds by any stretch of the theorist's imagination, but over evolutionary time their features are shaped by competitions that can be modeled by mathematical game theory—it is *as if* the plants and their competitors were agents like us! Plants that have an evolutionary history of being heavily preyed upon by herbivores often evolve toxicity to those herbivores as a retaliatory measure. The herbivores, in turn, often evolve a specific tolerance in their digestive systems for those specific toxins, and return to the feast, until the day when the plants, foiled in their first attempt, develop further toxicity or prickly barbs, as their next move in an escalating arms race of measure and countermeasure. At some point, the herbivores may "choose" not to retaliate but rather to discriminate, turning to other food sources, and then other nontoxic plants may evolve to "mimic" the toxic plants, blindly exploiting a weakness in the discriminatory system—visual or olfactory—of the herbivores and thereby hitching a free ride on the toxicity defense of the other plant species. The free-floating rationale is clear and predictive, even though neither the plants nor the digestive systems of the herbivores have minds in anything like the ordinary sense.

All this happens at an achingly slow pace, by our standards. It can take thousands of generations, thousands of years, for a single move in this game of hide-and-seek to be made and responded to (though in some circumstances the pace is shockingly fast). The patterns of evolutionary change emerge so slowly that they are invisible at our normal rate of information uptake, so it's easy to overlook their intentional interpretation, or to dismiss it as mere whimsy or metaphor. This bias in favor of *our* normal pace might be called *timescale chauvinism*. Take the smartest, quickest-witted person you know, and imagine filming her in action in ultra-slow motion—say, thirty thousand frames per second, to be projected at the normal rate of thirty frames per second. A single lightning riposte, a witticism offered "without skipping a beat," would now emerge like a glacier from her mouth, boring even the most patient moviegoer. Who could divine the intelligence of her performance, an intelligence that would be unmistakable at normal speed? We are also charmed by mismatched timescales going in the other direction, as time-lapse photography has vividly demonstrated. To watch flowers growing, budding, and blooming in a few seconds, is to be drawn almost irresistibly into the intentional stance. See how that plant is striving upward, racing its neighbor for a favored place in the sun, defiantly thrusting its own leaves into the light, parrying the counterblows, ducking and weaving like a boxer! The very same patterns, projected at different speeds, can reveal or conceal the presence of a mind, or the absence of a mind—or so it seems. (Spatial scale also shows a powerful built-in bias; if gnats were the size of seagulls, more people would be sure they had minds, and if we had to look through microscopes to see the antics of otters, we would be less confident that they were fun-loving.)

In order for us to see things as mindful, they have to happen at the right pace, and when we do see something as mindful, we don't have much choice; the perception is almost irresistible. But is this just a fact about our bias as observers, or is it a fact about minds? What is the *actual* role of speed in the phenomenon of mind? Could there be minds,

as real as any minds anywhere, that conducted their activities orders of magnitude slower than our minds do? Here is a reason for thinking that there could be: if our planet were visited by Martians who thought the same sorts of thoughts we do but thousands or millions of times faster than we do, we would seem to them to be about as stupid as trees, and they would be inclined to scoff at the hypothesis that we had minds. If they did, they would be wrong, wouldn't they—victims of their own timescale chauvinism. So if we want to deny that there could be a radically slow-thinking mind, we will have to find some grounds other than our preference for the human thought rate. What grounds might there be? Perhaps, you may think, there is a minimum speed for a mind, rather like the minimum escape velocity required to overcome gravity and leave the planet. For this idea to have any claim on our attention, let alone allegiance, we would need a theory that says why this should be. What could it be about running a system faster and faster that eventually would "break the mind barrier" and create a mind where before there was none? Does the friction of the moving parts create heat, which above a certain temperature leads to the transformation of something at the chemical level? And why would that make a mind? Is it like particles in an accelerator approaching the speed of light and becoming hugely massive? Why would that make a mind? Does the rapid spinning of the brain parts somehow weave a containment vessel to prevent the escape of the accumulating mind particles until a critical mass of them coheres into a mind? Unless something along these lines can be proposed *and defended*, the idea that sheer speed is essential for minds is unappealing, since there is such a good reason for holding that it's the *relative* speed that matters: perception, deliberation, and action all swift enough—relative to the unfolding environment—to accomplish the purposes of a mind. Producing future is no use to any intentional system if its "predictions" arrive too late to be acted on. Evolution will always favor the quick-witted over the slow-witted, other things being equal, and extinguish those who can't meet their deadlines well on a regular basis.

But what if there were a planet on which the speed of light was 100 kilometers per hour, and all other physical events and processes were slowed down to keep pace? Since in fact the pace of events in the physical world can't be sped up or slowed down by orders of magnitude (except in philosophers' fantastic thought experiments), a relative speed requirement works as well as an absolute speed requirement. Given the speed at which thrown stones approach their targets, and given the speed at which light bounces off those incoming stones, and given the speed at which audible warning calls can be propagated through the atmosphere, and given the force that must be marshaled to get 100 kilograms of body running at 20 kilometers per hour to veer sharply to the left or right—given these and a host of other firmly fixed performance specifications, useful brains have to operate at quite definite minimum speeds, independently of any fanciful "emergent properties" that might also be produced only at certain speeds. These speed-of-operation requirements, in turn, force brains to use media of information transmission that can sustain those speeds. That's one good reason why it can matter what a mind is made of. There may be others.

When the events in question unfold at a more stately pace, something mindlike can occur in other media. These patterns are discernible in these phenomena only when we adopt the intentional stance. Over very long periods of time, species or lineages of plants and animals can be *sensitive* to changing conditions, and *respond* to the changes they sense in rational ways. That's all it takes for the intentional stance to find predictive and explanatory leverage. Over much shorter periods of time, individual plants can respond

appropriately to changes they sense in their environment, growing new leaves and branches to exploit the available sunlight, extending their roots toward water, and even (in some species) *temporarily* adjusting the chemical composition of their edible parts to ward off the *sensed onslaught* of transient herbivores.

These sorts of slow-paced sensitivity, like the artificial sensitivity of thermostats and computers, may strike us as mere second-rate imitations of the phenomenon that really makes the difference: *sentience*. Perhaps we can distinguish "mere intentional systems" from "genuine minds" by asking whether the candidates in question enjoy sentience. Well, what is it? "Sentience" has never been given a proper definition, but it is the more or less standard term for what is imagined to be the lowest grade of consciousness. We may wish to entertain the strategy, at about this point, of contrasting sentience with mere sensitivity, a phenomenon exhibited by single-celled organisms, plants, the fuel gauge in your car, and the film in your camera. Sensitivity need not involve consciousness at all. Photographic film comes in different grades of sensitivity to light; thermometers are made of materials that are sensitive to changes in temperature; litmus paper is sensitive to the presence of acid. Popular opinion proclaims that plants and perhaps "lower" animals—jellyfish, sponges, and the like—are sensitive without being sentient, but that "higher" animals are sentient. Like us, they are not *merely* endowed with sensitive equipment of one sort or another—equipment that responds differentially and appropriately to one thing or another. They enjoy some further property, called sentience—so says popular opinion. But what is this commonly proclaimed property?

What does sentience amount to, above and beyond sensitivity? This is a question that is seldom asked and has never been properly answered. We shouldn't assume that there's a good answer. We shouldn't assume, in other words, that it's a good question. If we want to use the concept of sentience, we will have to construct it from parts we understand. Everybody agrees that sentience requires sensitivity plus some further as yet unidentified factor *x*, so if we direct our attention to the different varieties of sensitivity and the roles in which they are exploited, keeping a sharp lookout for something that strikes us as a crucial addition, we may discover sentience along the way. Then we can add the phenomenon of sentience to our unfolding story—or, alternatively, the whole idea of sentience as a special category may evaporate. One way or another, we will cover the ground that separates conscious us from the merely sensitive, insentient macromolecules we are descended from. One tempting place to look for the key difference between sensitivity and sentience is in the materials involved—the *media* in which information travels and is transformed.

THE MEDIA AND THE MESSAGES

We must look more closely at the development I sketched at the beginning of chapter 2. The earliest control systems were really just body protectors. Plants are alive, but they don't have brains. They don't need them, given their lifestyle. They do, however, need to keep their bodies intact and properly situated to benefit from the immediate surroundings, and for this they evolved systems of self-governance or control that took account of the crucial variables and reacted accordingly. Their concerns—and hence their rudimentary intentionality—was either directed inward, to internal conditions, or directed to conditions at the all-important boundaries between the body and the cruel world. The responsibility for monitoring and making adjustments was distributed, not centralized. Local sensing of changing conditions could

be met by local reactions, which were largely independent of each other. This could sometimes lead to coordination problems, with one team of microagents acting at cross-purposes to another. There are times when independent decision making is a bad idea; if everybody decides to lean to the right when the boat tips to the left, the boat may well tip over to the right. But in the main, the minimalist strategies of plants can be well met by highly distributed "decision making," modestly coordinated by the slow, rudimentary exchange of information by diffusion in the fluids coursing through the plant body.

Might plants then just be "very slow animals," enjoying sentience that has been overlooked by us because of our timescale chauvinism? Since there is no established meaning to the word "sentience," we are free to adopt one of our own choosing, if we can motivate it. We could refer to the slow but reliable responsiveness of plants to their environment as "sentience" if we wanted, but we would need some reason to distinguish this quality from the mere sensitivity exhibited by bacteria and other single-celled life-forms (to say nothing of light meters in cameras). There's no ready candidate for such a reason, and there's a fairly compelling reason for reserving the term "sentience" for something more special: animals have slow body-maintenance systems rather like those of plants, and common opinion differentiates between the operation of these systems and an animal's sentience.

Animals have had slow systems of body maintenance for as long as there have been animals. Some of the molecules floating along in such media as the bloodstream are themselves *operatives* that directly "do things" for the body (for instance, some of them destroy toxic invaders in one-on-one combat), and some are more like *messengers*, whose arrival at and "recognition" by some larger agent tells the larger agent to "do things" (for instance, to speed up the heart rate

or initiate vomiting). Sometimes the larger agent is the entire body. For instance, when the pineal gland in some species detects a general decrease in daily sunlight, it broadcasts to the whole body a hormonal message to begin preparing for winter—a task with many subtasks, all set into motion by one message. Although activity in these ancient hormonal systems may be accompanied by powerful instances of what we may presume to be sentience (such as waves of nausea, or dizzy feelings, or chills, or pangs of lust), these systems operate independently of those sentient accompaniments—for instance, in sleeping or comatose animals. Doctors speak of brain-dead human beings kept alive on respirators as being in a "vegetative state," when these body-maintenance systems alone are keeping life and limb together. Sentience is gone, but sensitivity of many sorts persists, maintaining various bodily balances. Or at least that's how many people would want to apply these two words.

In animals, this complex system of biochemical packets of control information was eventually supplemented by a swifter system, running in a different medium: traveling pulses of electrical activity in nerve fibers. This opened up a space of opportunities for swifter reactions, but also permitted the control to be differently distributed, because of the different geometries of connection possible in this new system, the autonomic nervous system. The concerns of the new system were still internal—or, at any rate, immediate in both space and time: Should the body shiver now, or should it sweat? Should the digestive processes in the stomach be postponed because of more pressing needs for the blood supply? Should the countdown to ejaculation begin? And so forth. The interfaces between the new medium and the old had to be worked out by evolution, and the history of that development has left its marks on our current arrangements, making them much more complicated than one might have expected. Ignoring these complexities has often led theorists

of mind astray—myself included—so we should note them, briefly.

One of the fundamental assumptions shared by many modern theories of mind is known as *functionalism*. The basic idea is well known in everyday life and has many proverbial expressions, such as *handsome is as handsome does*. What makes something a mind (or a belief, or a pain, or a fear) is not what it is made of, but what it *can do*. We appreciate this principle as uncontroversial in other areas, especially in our assessment of artifacts. What makes something a spark plug is that it can be plugged into a particular situation and *deliver a spark when called upon*. That's all that matters; its color or material or internal complexity can vary ad lib, and so can its shape, as long as its shape permits it to meet the specific dimensions of its functional role. In the world of living things, functionalism is widely appreciated: a heart is something for pumping blood, and an artificial heart or a pig's heart may do just about as well, and hence can be substituted for a diseased heart in a human body. There are more than a hundred chemically different varieties of the valuable protein lysozyme. What makes them all instances of lysozyme is what makes them valuable: what they can do. They are interchangeable, for almost all intents and purposes.

In the standard jargon of functionalism, these functionally defined entities admit *multiple realizations*. Why couldn't artificial minds, like artificial hearts, be made real—realized—out of almost anything? Once we figure out what minds do (what pains do, what beliefs do, and so on), we ought to be able to make minds (or mind parts) out of alternative materials that have those competences. And it has seemed obvious to many theorists—myself included—that what minds do is *process information*; minds are the control systems of bodies, and in order to execute their appointed duties they need to gather, discriminate, store, transform, and otherwise process information about the control tasks they perform. So far, so good. Functionalism, here as elsewhere, promises to make life easier for the theorist by abstracting away from some of the messy particularities of performance and focusing on the work that is actually getting done. But it's almost standard for functionalists to oversimplify their conception of this task, making life *too* easy for the theorist.

It's tempting to think of a nervous system (either an autonomic nervous system or its later companion, a central nervous system) as an information network tied at various specific places—transducer (or *input*) nodes and effector (or *output*) nodes—to the realities of the body. A *transducer* is any device that takes information in one medium (a change in the concentration of oxygen in the blood, a dimming of the ambient light, a rise in temperature) and translates it into another medium. A photoelectric cell transduces light, in the form of impinging photons, into an electronic signal, in the form of electrons streaming through a wire. A microphone transduces sound waves into signals in the same electronic medium. A bimetallic spring in a thermostat transduces changes in ambient temperature into a bending of the spring (and that, in turn, is typically translated into the transmission of an electronic signal down a wire to turn a heater on or off). The rods and cones in the retina of the eye are the transducers of light into the medium of nerve signals; the eardrum transduces sound waves into vibrations, which eventually get transduced (by the hair cells on the basilar membrane) into the same medium of nerve signals. There are temperature transducers distributed throughout the body, and motion transducers (in the inner ear), and a host of other transducers of other information. An *effector* is any device that can be directed, by some signal in some medium, to make something happen in another "medium" (to bend an arm, close a pore, secrete a fluid, make a noise).

In a computer, there is a nice neat boundary between the "outside" world and the information channels. The input devices, such as the keys on the keyboard, the mouse, the microphone, the television camera, all transduce information into a common medium—the electronic medium in which "bits" are transmitted, stored, transformed. A computer can have internal transducers too, such as a temperature transducer that "informs" the computer that it is overheating, or a transducer that warns it of irregularities in its power supply, but these count as *input* devices, since they extract information from the (internal) environment and put it in the common medium of information processing.

It would be theoretically clean if we could insulate information channels from "outside" events in a body's nervous system, so that all the important interactions happened at identifiable transducers and effectors. The division of labor this would permit is often very illuminating. Consider a ship with a steering wheel located at some great distance from the rudder it controls. You can connect the wheel to the rudder with ropes, or with gears and bicycle chains, wires and pulleys, or with a hydraulic system of high-pressure hoses filled with oil (or water or whiskey!). In one way or another, these systems transmit to the rudder the energy that the helmsman supplies when turning the wheel. *Or* you can connect the wheel to the rudder with nothing but a few thin wires, through which electronic signals pass. You don't have to transduce the energy, just the information *about* how the helmsman wants the rudder to turn. You can transduce this information from the steering wheel into a signal at one end and put the energy in locally, at the other end, with an effector—a motor of some kind. (You can also add "feedback" messages, which are transduced at the motor-rudder end and sent up to control the resistance-to-turning of the wheel, so that the helmsman can sense the pressure of the water on the rudder as it turns. This feedback is standard, these days, in

power steering in automobiles, but was dangerously missing in the early days of power steering.)

If you opt for this sort of system—a pure signaling system that transmits information and almost no energy—then it really makes no difference at all whether the signals are electrons passing through a wire or photons passing through a glass fiber or radio waves passing through empty space. In all these cases, what matters is that the information not be lost or distorted because of the time lags between the turning of the wheel and the turning of the rudder. This is also a key requirement in the energy-transmitting systems—the systems using mechanical linkages, such as chains or wires or hoses. That's why elastic bands are not as good as unstretchable cables, even though the information eventually gets there, and why incompressible oil is better than air in a hydraulic system.*

In modern machines, it is often possible in this way to isolate the control system from the system that is controlled, so that control systems can be readily interchanged with no loss of function. The familiar remote controllers of electronic appliances are obvious examples of this, and so are electronic ignition systems (replacing the old mechanical linkages) and other computer-chip-based devices in automobiles. And up to a point, the same freedom from particular media is a feature of animal nervous systems, whose parts can be quite clearly segregated into the peripheral transducers and effectors and the intermediary transmission pathways. One way of going deaf, for instance, is to lose your auditory nerve to cancer. The

.....

*The example of the steering gear has an important historical pedigree. The term "cybernetics" was coined by Norbert Wiener from the Greek word for "helmsman" or "steerer." The word "governor" comes from the same source. These ideas about how control is accomplished by the transmission and processing of information were first clearly formulated by Wiener in *Cybernetics; or, Control and Communication in the Animal and the Machine* (1948).

sound-sensitive parts of the ear are still intact, but the transmission of the results of their work to the rest of the brain has been disrupted. This destroyed avenue can now be replaced by a prosthetic link, a tiny cable made of a different material (wire, just as in a standard computer), and since the interfaces at both ends of the cable can be matched to the requirements of the existing healthy materials, the signals can get through. Hearing is restored. It doesn't matter at all what the medium of transmission is, just as long as the information gets through without loss or distortion.

This important theoretical idea sometimes leads to serious confusions, however. The most seductive confusion could be called the Myth of Double Transduction: first, the nervous system transduces light, sound, temperature, and so forth into neural signals (trains of impulses in nerve fibers) and second, in some special central place, it transduces these trains of impulses into some *other* medium, the medium of consciousness! That's what Descartes thought, and he suggested that the pineal gland, right in the center of the brain, was the place where this second transduction took place—into the mysterious, nonphysical medium of the mind. Today almost no one working on the mind thinks there is any such nonphysical medium. Strangely enough, though, the idea of a second transduction into some special *physical* or *material* medium, in some yet-to-be-identified place in the brain, continues to beguile unwary theorists. It is as if they saw—or thought they saw—that since peripheral activity in the nervous system was mere sensitivity, there had to be some more central place where the sentience was created. After all, a live eyeball, disconnected from the rest of the brain, cannot *see*, has no *conscious visual experience*, so that must happen later, when the mysterious *x* is added to mere sensitivity to yield sentience.

The reasons for the persistent attractiveness of this idea are not hard to find. One is tempted to think that mere nerve

impulses couldn't be the stuff of consciousness—that they need translation, somehow, into something else. Otherwise, the nervous system would be like a telephone system without anybody home to answer the phone, or a television network without any viewers—or a ship without a helmsman. It seems as if there has to be some central Agent or Boss or Audience, to take in (to transduce) all the information and *appreciate* it, and then "steer the ship."

The idea that the network *itself*—by virtue of its intricate structure, and hence powers of transformation, and hence capacity for controlling the body—could assume the role of the inner Boss and thus harbor consciousness, seems preposterous. Initially. But some version of this claim is the materialist's best hope. Here is where the very complications that ruin the story of the nervous system as a pure information-processing system can be brought in to help our imaginations, by distributing a portion of the huge task of "appreciation" back into the body.

"MY BODY HAS A MIND OF ITS OWN!"

> Nature appears to have built the apparatus of rationality not just on top of the apparatus of biological regulation, but also *from* it and *with* it.
>
> Antonio Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*

The medium of information transfer in the nervous system is electrochemical pulses traveling through the long branches of nerve cells—not like electrons traveling through a wire at the speed of light, but in a much-slower-traveling chain reaction. A nerve fiber is a sort of elongated battery, in which

chemical differences on the inside and outside of the nerve cell's wall induce electric activities that then propagate along the wall at varying speeds—much faster than molecule packets could be shipped through fluid, but much, much slower than the speed of light. Where nerve cells come in contact with each other, at junctures called synapses, a microeffector/microtransducer interaction takes place: the electrical pulse triggers the release of neurotransmitter molecules, which cross the gap by old-fashioned diffusion (the gap is very narrow) and are then transduced into further electrical pulses. A step backward, one might think, into the ancient world of molecular lock-and-key. Especially when it turns out that in addition to the neurotransmitter molecules (such as glutamate), which seem to be more or less neutral all-purpose synapse crossers, there are a variety of neuromodulator molecules, which, when *they* find the "locks" in the neighboring nerve cells, produce all sorts of changes of their own. Would it be right to say that the nerve cells *transduce* the presence of these neuromodulator molecules, in the same way that other transducers "notice" the presence of antigens, or oxygen, or heat? If so, then there are transducers at virtually every joint in the nervous system, adding input to the stream of information already being carried along by the electrical pulses. And there are also effectors everywhere, secreting neuromodulators and neurotransmitters into the "outside" world of the rest of the body, where they diffuse to produce many different effects. The crisp boundary between the information-processing system and the rest of the world—the rest of the body—breaks down.

It has always been clear that wherever you have transducers and effectors, an information system's "media-neutrality," or multiple realizability, disappears. In order to detect light, for instance, you need something photosensitive—something that will respond swiftly and reliably to photons, amplifying their subatomic arrival into larger-scale events

that can trigger still further events. (Rhodopsin is one such photosensitive substance, and this protein has been the material of choice in all natural eyes, from ants to fish to eagles to people. Artificial eyes might use some other photosensitive element, but not just anything will do.) In order to identify and disable an antigen, you need an antibody that has the right shape, since the identification is by the lock-and-key method. This limits the choice of antibody building materials to molecules that can fold up into these shapes, and this severely restricts the molecules' chemical composition—though not completely (as the example of lysozyme varieties shows). In theory, every information-processing system is tied at both ends, you might say, to transducers and effectors whose physical composition is dictated by the jobs they have to do; in between, everything can be accomplished by media-neutral processes.

The control systems for ships, automobiles, oil refineries, and other complex human artifacts are media-neutral, as long as the media used can do the job in the available time. The neural control systems for animals, however, are not really media-neutral—not because the control systems have to be made of particular materials in order to generate that special aura or buzz or whatever, but because they evolved as the control systems of organisms that already were lavishly equipped with highly distributed control systems, and the new systems had to be built on top of, and in deep collaboration with, these earlier systems, creating an astronomically high number of points of transduction. We can occasionally ignore these ubiquitous interpenetrations of different media—as, for instance, when we replace a single nerve highway, like the auditory nerve, with a prosthetic substitute—but only in a fantastic thought experiment could we ignore these interpenetrations *in general*.

For example: The molecular keys needed to unlock the locks that control every transaction between nerve cells are

glutamate molecules, dopamine molecules, and norepinephrine molecules (among others); but "in principle" all the locks could be changed—that is, replaced with a chemically different system. After all, the function of the chemical depends on its fit with the lock, and hence on the subsequent effects triggered by the arrival of this turn-on message, and not on anything else. But the distribution of responsibility throughout the body makes this changing of the locks practically impossible. Too much of the information processing—and hence information storage—is already embedded in these particular materials. And that's another good reason why, when you make a mind, the materials matter. So there are two good reasons for this: speed, and the ubiquity of transducers and effectors throughout the nervous system. I don't think there are any other good reasons.

These considerations lend support to the intuitively appealing claim often advanced by critics of functionalism: that it really does matter what you make a mind out of. You couldn't make a *sentient* mind out of silicon chips, or wire and glass, or beer cans tied together with string. Are these reasons for abandoning functionalism? Not at all. In fact, they depend on the basic insight of functionalism for their force.

The *only* reason minds depend on the chemical composition of their mechanisms or media is that in order to do the things these mechanisms must do, they have to be made, as a matter of biohistorical fact, from substances compatible with the preexisting bodies they control. Functionalism is opposed to vitalism and other forms of mysticism about the "intrinsic properties" of various substances. There is no more anger or fear in adrenaline than there is silliness in a bottle of whiskey. These substances, per se, are as irrelevant to the mental as gasoline or carbon dioxide. It is only when their abilities to function as components of larger functional systems depend on their internal composition that their so-called "intrinsic nature" matters.

The fact that your nervous system, unlike the control system of a modern ship, is not an insulated, media-neutral control system—the fact that it "effects" and "transduces" at almost every juncture—forces us to think about the functions of their parts in a more complicated (and realistic) way. This recognition makes life slightly more difficult for functionalist philosophers of mind. A thousand philosophical thought experiments (including my own story, "Where am I?" [1978]) have exploited the intuition that *I* am not my body but my body's . . . owner. In a heart transplant operation, you want to be the recipient, not the donor, but in a brain transplant operation, you want to be the donor—you go with the brain, not the body. In principle (as many philosophers have argued), *I* might even trade in my current brain for another, by replacing the medium while preserving only the message. I could travel by teleportation, for instance, as long as the information was perfectly preserved. In principle, yes—but only because one would be transmitting information about the whole body, not just the nervous system. One cannot tear me apart from my body leaving a nice clean edge, as philosophers have often supposed. My body contains as much of *me*, the values and talents and memories and dispositions that make me who I am, as my nervous system does.

The legacy of Descartes's notorious dualism of mind and body extends far beyond academia into everyday thinking: "These athletes are prepared both mentally and physically," and "There's nothing wrong with your body—it's all in your mind." Even among those of us who have battled Descartes's vision, there has been a powerful tendency to treat the mind (that is to say, the brain) as the body's boss, the pilot of the ship. Falling in with this standard way of thinking, we ignore an important alternative: viewing the brain (and hence the mind) as one organ among many, a relatively recent usurper of control, whose functions cannot properly be understood until we see it not as the boss but as just one

more somewhat fractious servant, working to further the interests of the body that shelters and fuels it and gives its activities meaning.

This historical or evolutionary perspective reminds me of the change that has come over Oxford in the thirty years since I was a student there. It used to be that the dons were in charge, and the bursars and other bureaucrats, right up to the vice chancellor, acted under their guidance and at their behest. Nowadays the dons, like their counterparts on American university faculties, are more clearly in the role of employees hired by a central administration. But from where, finally, does the University get its meaning? In evolutionary history, a similar change has crept over the administration of our bodies. But our bodies, like the Oxford dons, still have some power of decision—or, at any rate, some power to rebel when the central administration acts in ways that run counter to the sentiments of "the body politic."

It is harder to think functionalistically about the mind once we abandon the crisp identification of the mind with the brain and let it spread to other parts of the body, but the compensations are enormous. The fact that our control systems, unlike those of ships and other artifacts, are so noninsulated permits our bodies themselves (as distinct from the nervous systems they contain) to harbor much of the wisdom that "we" exploit in the course of daily decision making. Friedrich Nietzsche saw all this long ago, and put the case with characteristic brio, in *Thus Spake Zarathustra* (in the section aptly entitled "On the Despisers of the Body"):

> "Body am I, and soul"—thus speaks the child. And why should one not speak like children? But the awakened and knowing say: body am I entirely, and nothing else; and soul is only a word for something about the body.
>
> The body is a great reason, a plurality with one sense, a war and a peace, a herd and a shepherd. An instrument of your body is also your little reason, my brother, which you call "spirit"—a little instrument and toy of your great reason. . . . Behind your thoughts and feelings, my brother, there stands a mighty ruler, an unknown sage—whose name is self. In your body he dwells; he is your body. There is more reason in your body than in your best wisdom. (Kaufmann translation, 1954, p. 146)

Evolution embodies information in every part of every organism. A whale's baleen embodies information about the food it eats, and the liquid medium in which it finds its food. A bird's wing embodies information about the medium in which it does its work. A chameleon's skin, more dramatically, carries information about its current environment. An animal's viscera and hormonal systems embody a great deal of information about the world in which its ancestors have lived. This information doesn't have to be copied into the brain at all. It doesn't have to be "represented" in "data structures" in the nervous system. It can be exploited by the nervous system, however, which is designed to rely on, or exploit, the information in the hormonal systems just as it is designed to rely on, or exploit, the information embodied in the limbs and eyes. So there is wisdom, particularly about preferences, embodied in the rest of the body. By using the old bodily systems as a sort of sounding board, or reactive audience, or critic, the central nervous system can be guided—sometimes nudged, sometimes slammed—into wise policies. Put it to the vote of the body, in effect. To be fair to poor old Descartes, we should note that even he saw—at least dimly—the importance of this union of body and mind:

> By means of these feelings of pain, hunger, thirst, and so on, nature also teaches that I am present to my body not merely in the way a seaman is present to his ship, but that I am tightly joined and, so to speak, mingled together

> with it, so much so that I make up one single thing with it. (Meditation Six)

When all goes well, harmony reigns and the various sources of wisdom in the body cooperate for the benefit of the whole, but we are all too familiar with the conflicts that can provoke the curious outburst "My body has a mind of its own!" Sometimes, apparently, it is tempting to lump together some of this embodied information into a *separate* mind. Why? Because it is organized in such a way that it can sometimes make somewhat independent discriminations, consult preferences, make decisions, enact policies that are in competition with *your* mind. At such times, the Cartesian perspective of a puppeteer self trying desperately to control an unruly body-puppet is very powerful. Your body can vigorously betray the secrets *you* are desperately trying to keep—by blushing and trembling or sweating, to mention only the most obvious cases. It can "decide" that in spite of *your* well-laid plans, right now would be a good time for sex, not intellectual discussion, and then take embarrassing steps in preparation for a coup d'état. On another occasion, to your even greater chagrin and frustration, it can turn a deaf ear on your own efforts to enlist it for a sexual campaign, forcing you to raise the volume, twirl the dials, try all manner of preposterous cajolings to *persuade* it.

But why, if our bodies already had minds of their own, did they ever go about acquiring additional minds—*our* minds? Isn't one mind per body enough? Not always. As we have seen, the old body-based minds have done a robust job of keeping life and limb together over billions of years, but they are relatively slow and relatively crude in their discriminatory powers. Their intentionality is short-range and easily tricked. For more sophisticated engagements with the world, a swifter, farther-seeing mind is called for, one that can produce more and better future.

CHAPTER 4

HOW INTENTIONALITY CAME INTO FOCUS

THE TOWER OF GENERATE-AND-TEST*

In order to see farther ahead in time, it helps to see farther into space. What began as internal and peripheral monitoring systems slowly evolved into systems that were capable of not just proximal (neighboring) but distal (distant) discrimination. This is where perception comes into its own. The sense of smell, or olfaction, relies on the wafting from afar of harbinger keys to local locks. The trajectories of these harbingers are relatively slow, variable, and uncertain, because of random dispersal and evaporation; thus information about the source they emanate from is limited. Hearing depends on sound waves striking the system's transducers, and because the paths of sound waves are swifter and more regular, perception can come closer to approximating "action at a distance." But sound waves can deflect and bounce in ways that obscure their source. Vision depends on

*This section is drawn, with revisions, from *Darwin's Dangerous Idea*.

the much swifter arrival of photons bounced off the things in the world, on definitively straight-line trajectories, so that with a suitably shaped pinhole (and optional lens) arrangement, an organism can obtain instantaneous high-fidelity information about events and surfaces far away. How did this transition from internal to proximal to distal intentionality take place? Evolution created armies of specialized internal agents to receive the information available at the peripheries of the body. There is just as much information encoded in the light that falls on a pine tree as there is in the light that falls on a squirrel, but the squirrel is equipped with millions of information-seeking microagents, specifically designed to take in, and even to seek out and interpret this information.

Animals are not just herbivores or carnivores. They are, in the nice coinage of the psychologist George Miller, *informavores*. And they get their epistemic hunger from the combination, in exquisite organization, of the specific epistemic hungers of millions of microagents, organized into dozens or hundreds or thousands of subsystems. Each of these tiny agents can be conceived of as an utterly minimal intentional system, whose life project is to ask a single question, over and over and over—"Is my message coming in NOW?" "Is my message coming in NOW?"—and springing into limited but appropriate action whenever the answer is YES. Without the epistemic hunger, there is no perception, no uptake. Philosophers have often attempted to analyze perception into the Given and what is then done with the Given by the mind. The Given is, of course, Taken, but the taking of the Given is not something done by one Master Taker located in some central headquarters of the animal's brain. The task of taking is distributed among all the individually organized takers. The takers are not just the peripheral transducers—the rods and cones on the retina of the eye, the specialized cells in the epithelium of the nose—but also all the internal

functionaries fed by them, cells and groups of cells connected in networks throughout the brain. They are fed not patterns of light or pressure (the pressure of sound waves and of touch) but patterns of neuronal impulses; but aside from the change of diet, they are playing similar roles. How do all these agents get organized into larger systems capable of sustaining ever more sophisticated sorts of intentionality? By a process of evolution by natural selection, of course, but not just one process.

I want to propose a framework in which we can place the various design options for brains, to see where their power comes from. It is an outrageously oversimplified structure, but idealization is the price one should often be willing to pay for synoptic insight. I call it the Tower of Generate-and-Test. As each new floor of the Tower gets constructed, it empowers the organisms at that level to find better and better moves, and find them more efficiently.

The increasing power of organisms to produce future can be represented, then, in a series of steps. These steps almost certainly don't represent clearly defined transitional periods in evolutionary history—no doubt such steps were taken in overlapping and nonuniform ways by different lineages—but the various floors of the Tower of Generate-and-Test mark important advances in cognitive power, and once we see in outline a few of the highlights of each stage, the rest of the evolutionary steps will make more sense.

In the beginning, there was Darwinian evolution of species by natural selection. A variety of candidate organisms were blindly generated, by more or less arbitrary processes of recombination and mutation of genes. These organisms were field-tested, and only the best designs survived. This is the ground floor of the tower. Let us call its inhabitants *Darwinian creatures*.

This process went through many millions of cycles, producing many wonderful designs, both plant and animal.

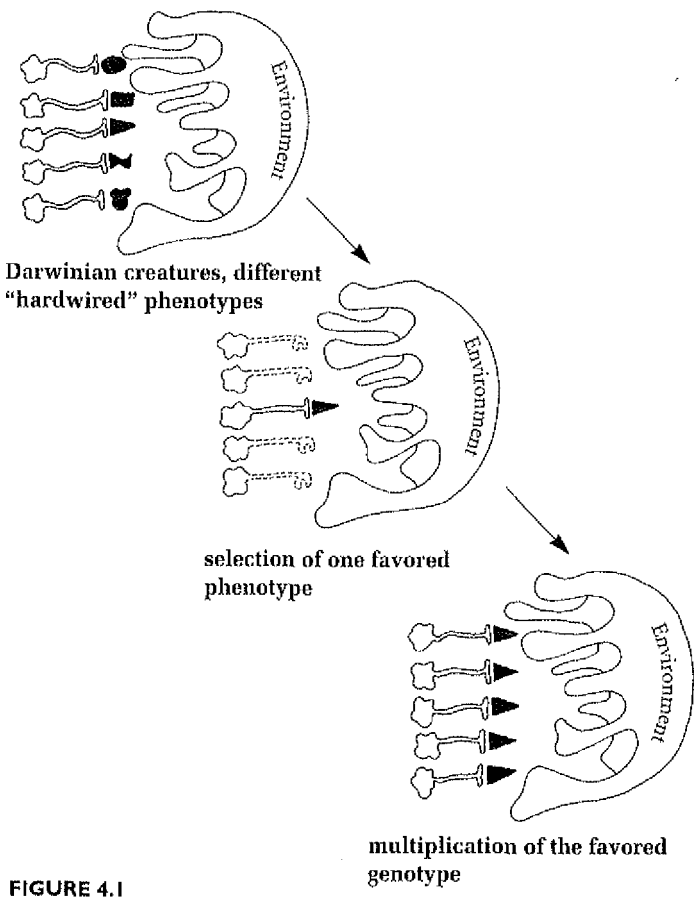


FIGURE 4.1

Eventually, among its novel creations were some designs with the property of *phenotypic plasticity*: that is, the individual candidate organisms were not wholly designed at birth; there were elements of their design that could be *adjusted by events that occurred during the field tests.* Some of these candidates, we may suppose, were no better off than their cousins, the hardwired Darwinian creatures, since they had no way of favoring (selecting for an encore) the behavioral options they were equipped to "try out." But others, we may suppose, were fortunate enough to have wired-in "reinforcers" that happened to favor Smart Moves—that is, actions that were better for the candidates than the available alternative actions. These individuals thus confronted the environment by generating a variety of actions, which they tried out, one by one, until they found one that worked. They detected that it worked only by getting a positive or negative signal from the environment, which adjusted the probability of that action's being reproduced on another occasion. Any creatures wired up wrong—with positive and negative reinforcement reversed—would be doomed, of course. Only those fortunate enough to be born with appropriate reinforcers would have an advantage. We may call this subset of Darwinian creatures *Skinnerian creatures*, since, as the behaviorist psychologist B. F. Skinner was fond of pointing out, such "operant conditioning" is not just analogous to Darwinian natural selection; it is an extension of it: "Where inherited behavior leaves off, the inherited modifiability of the process of conditioning takes over." (1953, p. 83)

The cognitive revolution that emerged in the 1970s ousted behaviorism from its dominant position in psychology, and ever since there has been a tendency to underestimate the power of Skinnerian conditioning (or its variations) to shape the behavioral competence of organisms into highly adaptive and discerning structures. The flourishing work on neural networks and "connectionism" in the 1990s, however, has demonstrated anew the often surprising virtuosity of simple networks that begin life more or less randomly wired and then have their connections adjusted by a simple sort of "experience"—the history of reinforcement they encounter.

The fundamental idea of letting the environment play a blind but selective role in shaping the mind (or brain or control system) has a pedigree even older than Darwin. The intellectual ancestors of today's connectionists and yester-

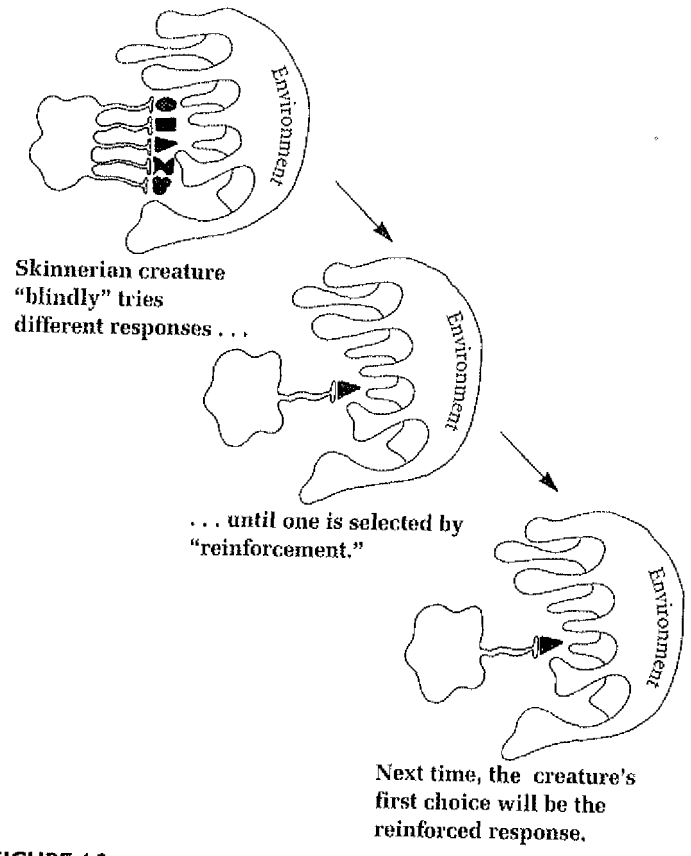


FIGURE 4.2

day's behaviorists were the associationists: such philosophers as David Hume, who tried in the eighteenth century to imagine how mind parts (he called them impressions and ideas) could become self-organizing without benefit of some all-too-knowing director of the organization. As a student once memorably said to me, "Hume wanted to get the ideas to think for themselves." Hume had wonderful hunches about how impressions and ideas might link themselves together by a process rather like chemical bonding, and then create beaten paths of habit in the mind, but these hunches

were too vague to be tested. Hume's associationism was, however, a direct inspiration for Pavlov's famous experiments in the conditioning of animal behavior, which led in turn to the somewhat different conditioning theories of E. L. Thorndike, Skinner, and the other behaviorists in psychology. Some of these researchers—Donald Hebb, in particular—attempted to link their behaviorism more closely to what was then known about the brain. In 1949, Hebb proposed models of simple conditioning mechanisms that could adjust the connections between nerve cells. These mechanisms—now called Hebbian learning rules—and their descendants are the engines of change in connectionism, the latest manifestation of this tradition.

Associationism, behaviorism, connectionism—in historical and alphabetical order we can trace the evolution of models of one simple kind of learning, which might well be called *ABC learning*. There is no doubt that most animals are capable of ABC learning; that is, they can come to modify (or redesign) their behavior in appropriate directions as a result of a long, steady process of training or shaping by the environment. There are now good models, in varying degrees of realism and detail, of how such a process of conditioning or training can be nonmiraculously accomplished in a network of nerve cells.

For many life-saving purposes (pattern recognition, discrimination, and generalization, and the dynamical control of locomotion, for instance), ABC networks are quite wonderful—efficient, compact, robust in performance, fault-tolerant, and relatively easy to redesign on the fly. Such networks, moreover, vividly emphasize Skinner's point that it makes little difference where we draw the line between the pruning and shaping by natural selection which is genetically transmitted to offspring (the wiring you are born with), and the pruning and shaping that later takes place in the individual (the rewiring you end up with, as a result of experience or training). Nature and nurture blend seamlessly together.

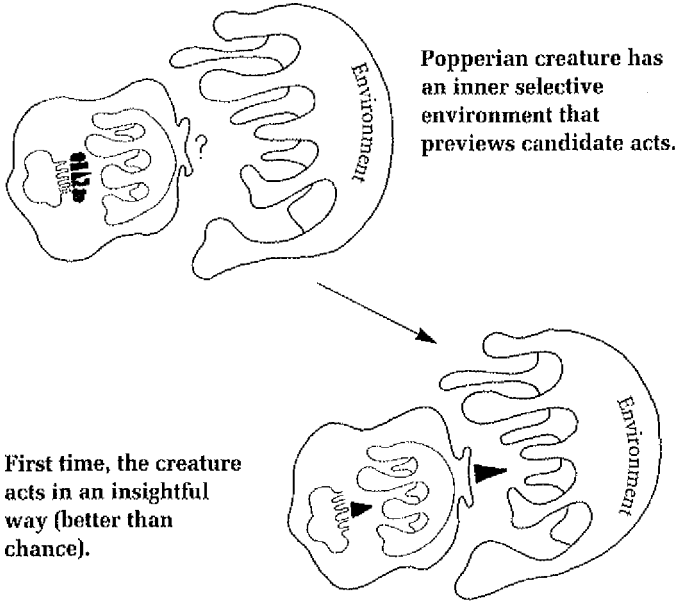
There are, however, some cognitive tricks that such ABC networks have not yet been trained to perform, and—a more telling criticism—there are some cognitive tricks that are quite clearly not the result of training at all. Some animals seem to be capable of "one-shot learning"; they can figure some things out without having to endure the arduous process of trial-and-error in the harsh world that is the hallmark of all ABC learning.

Skinnerian conditioning is a good thing as long as you are not killed by one of your early errors. A better system involves *preselection* among all the possible behaviors or actions, so that the truly stupid moves are weeded out before they're hazarded in "real life." We human beings are creatures capable of this particular refinement, but we are not alone. We may call the beneficiaries of this third floor in the Tower *Popperian creatures*, since, as the philosopher Sir Karl Popper once elegantly put it, this design enhancement "permits our hypotheses to die in our stead." Unlike the merely Skinnerian creatures, many of whom survive only because they make lucky first moves, Popperian creatures survive because they're smart enough to make better-than-chance first moves. Of course they're just lucky to be smart, but that's better than being just lucky.

How is this preselection in Popperian agents to be done? There must be a filter, and any such filter must amount to a sort of *inner environment,* in which tryouts can be safely executed—an inner something-or-other structured in such a way that the surrogate actions it favors are more often than not the very actions the real world would also bless, if they were actually performed. In short, the inner environment, whatever it is, must contain lots of *information* about the outer environment and its regularities. Nothing else (except magic) could provide preselection worth having. (One could always flip a coin or consult an oracle, but this is no improvement over blind trial and error—unless the coin or

oracle is systematically biased by someone or something that has true information about the world.)

The beauty of Popper's idea is exemplified in the recent development of realistic flight simulators used for training airplane pilots. In a simulated world, pilots can learn which moves to execute in which crises without ever risking their lives (or expensive airplanes). As examples of the Popperian trick, however, flight simulators are in one regard misleading: they reproduce the real world too literally. We must be very careful not to think of the inner environment of a Popperian creature as simply a replica of the outer world, with all the physical contingencies of that world reproduced. In such a miraculous toy world, the little hot stove in your head would be hot enough to actually burn the little finger in your head that you placed on it! Nothing of the sort needs to be supposed. The *information* about the effect of putting a



Popperian creature has an inner selective environment that previews candidate acts.

First time, the creature acts in an insightful way (better than chance).

FIGURE 4.3

finger on the stove has to be in there, and it has to be in there in a form that can produce its premonitory effect when called upon in an internal trial, but this effect can be achieved without constructing a replica world. After all, it would be equally Popperian to educate pilots just by having them read a book that explained to them all the contingencies they might encounter when they eventually climbed into the cockpit. It might not be as powerful a method of learning, but it would be hugely better than trial-and-error in the sky! The common element in Popperian creatures is that one way or another (either by inheritance or by acquisition) information is installed in them—accurate information about the world that they (probably) will encounter—and this information is in such a form that it can achieve the preselective effects that are its raison d'être.

One of the ways Popperian creatures achieve useful filtering is by putting candidate behavioral options before the bodily tribunal and exploiting the wisdom, however out-of-date or shortsighted, accumulated in those tissues. If the body rebels—for example, in such typical reactions as nausea, vertigo, or fear and trembling—this is a semireliable sign (better than a coin flip) that the contemplated act might not be a good idea. Here we see that rather than rewiring the brain to eliminate these choices, making them strictly unthinkable, evolution may simply arrange to respond to any thinking of them with a negative rush so strong as to make them highly unlikely to win the competition for execution. The information in the body that grounds the reaction may have been placed there either by genetic recipe or by recent individual experience. When a human infant first learns to crawl, it has an innate aversion to venturing out onto a pane of supportive glass, through which it can see a "visual cliff." Even though its mother beckons it from a few feet away, cajoling and encouraging, the infant hangs back fearfully, despite never having suffered a fall in its life. The

experience of its ancestors is making it err on the side of safety. When a rat has eaten a new kind of food and has then been injected with a drug that causes it to vomit, it will subsequently show a strong aversion to food that looks and smells like the food it ate just before vomiting. Here the information leading it to err on the side of safety was obtained from its own experience. Neither filter is perfect—after all, the pane of glass is actually safe, and the rat's new food is actually nontoxic—but better safe than sorry.

Clever experiments by psychologists and ethologists suggest other ways in which animals can try actions out "in their heads" and thereby reap a Popperian benefit. In the 1930s and 1940s, behaviorists demonstrated to themselves time and again that their experimental animals were capable of "latent learning" about the world—learning that was not specifically rewarded by any detectable reinforcement. (Their exercise in self-refutation is itself a prime example of another Popperian theme: that science makes progress only when it poses refutable hypotheses.) If left to explore a maze in which no food or other reward was present, rats would simply learn their way around in the normal course of things; then, if something they valued was introduced into the maze, the rats that had learned their way around on earlier forays were much better at finding it (not surprisingly) than the rats in the control group, which were seeing the maze for the first time. This may seem a paltry discovery. Wasn't it always obvious that rats were smart enough to learn their way around? Yes and no. It may have *seemed* obvious, but this is just the sort of testing—testing against the background of the null hypothesis—that must be conducted if we are going to be sure just how intelligent, how mindful, various species are. As we shall see, other experiments with animals demonstrate surprisingly stupid streaks—almost unbelievable gaps in the animals' knowledge of their own environments.

The behaviorists tried valiantly to accommodate latent learning into their ABC models. One of their most telling stopgaps was to postulate a "curiosity drive," which was satisfied (or "reduced," as they said) by exploration. There was reinforcement going on after all in those nonreinforcing environments. Every environment, marvelous to say, is full of reinforcing stimuli simply by being an environment in which there is something to learn. As an attempt to save orthodox behaviorism, this move was manifestly vacuous, but that does not make it a hopeless idea in other contexts; it acknowledges the fact that curiosity—epistemic hunger—must drive any powerful learning system.

We human beings are conditionable by ABC training, so we are Skinnerian creatures, but we are not *just* Skinnerian creatures. We also enjoy the benefits of much genetically inherited hardwiring, so we are Darwinian creatures as well. But we are more than that. We are Popperian creatures. Which other animals are Popperian creatures, and which are merely Skinnerian? Pigeons were Skinner's favorite experimental animals, and he and his followers developed the technology of operant conditioning to a very sophisticated level, getting pigeons to exhibit remarkably bizarre and sophisticated learned behaviors. Notoriously, the Skinnerians never succeeded in proving that pigeons were *not* Popperian creatures; and research on a host of different species, from octopuses to fish to mammals, strongly suggests that if there are any purely Skinnerian creatures, capable only of blind trial-and-error learning, they are to be found among the simple invertebrates. The huge sea slug (or sea hare) *Aplysia californica* has more or less replaced the pigeon as the focus of attention among those who study the mechanisms of simple conditioning.

We do not differ from all other species in being Popperian creatures then. Far from it; mammals and birds, reptiles, amphibians, fish, and even many invertebrates exhibit the

capacity to use general information they obtain from their environments to presort their behavioral options before striking out. How does the new information about the outer environment get incorporated into their brains? By perception, obviously. The environment contains an embarrassment of riches, much more information than even a cognitive angel could use. Perceptual mechanisms designed to ignore most of the flux of stimuli concentrate on the most useful, most reliable information. And how does the information gathered manage to exert its selective effect when the options are "considered," helping the animal design ever more effective interactions with its world? There are no doubt a variety of different mechanisms and methods, but among them are those that use the body as a sounding board.

THE SEARCH FOR SENTIENCE: A PROGRESS REPORT

We have been gradually adding elements to our recipe for a mind. Do we have the ingredients for sentience yet? Certainly the normal behavior of many of the animals we have been describing passes our intuitive tests for sentience with flying colors. Watching a puppy or a baby tremble with fear at the edge of an apparent precipice, or a rat grimacing in apparent disgust at the odor of supposedly toxic food, we have difficulty even entertaining the hypothesis that we are *not* witnessing a sentient being. But we have also uncovered substantial grounds for caution: we have seen some ways in which surprisingly mindlike behavior can be produced by relatively simple, mechanical, apparently unmindlike control systems. The potency of our instinctual responses to sheer speed and lifelikeness of motion, for instance, should alert us to the genuine—not merely philosophical—possibil-

ity that we can be fooled into attributing more subtlety, more understanding, to an entity than the circumstances warrant. Recognizing that observable behavior can enchant us, we can appreciate the need to ask further questions—about what lies behind that behavior.

Consider pain. In 1986, the British government amended its laws protecting animals in experiments, adding the octopus to the privileged circle of animals that may not be operated upon without anesthesia. An octopus is a mollusk, physiologically more like an oyster than a trout (let alone a mammal), but the behavior of the octopus and the other cephalopods (squid, cuttlefish) is so strikingly intelligent and—apparently—sentient that the scientific authorities decided to let behavioral similarity override internal difference: cephalopods (but not other mollusks) are officially presumed to be capable of feeling pain—just in case they are. Rhesus monkeys, in contrast, are physiologically and evolutionarily very close to us, so we tend to assume that they are capable of suffering the way we do, but they exhibit astonishingly different behavior on occasion. The primatologist Marc Hauser has told me in conversation that during mating season the male monkeys fight ferociously, and it is not uncommon to see one male pin another down and then bite and rip out one of its testicles. The injured male does not shriek or make a facial expression but simply licks the wound and walks away. A day or two later, the wounded animal may be observed mating! It is hard to believe that this animal was experiencing anything like the agonies of a human being similarly afflicted—the mind reels to think of it—in spite of our biological kinship. So we can no longer hope that the physiological and behaviorial evidence will happily converge to give us unequivocal answers, since we already know cases in which these two sorts of compelling if inconclusive evidence pull in opposite directions. How then can we think about this issue?

A key function of pain is negative reinforcement—the "punishment" that diminishes the likelihood of a repeat performance—and any Skinnerian creature can be trained by negative reinforcement of one sort or another. Is all such negative reinforcement pain? *Experienced* pain? Could there be unconscious or unexperienced *pain*? There are simple mechanisms of negative reinforcement that provide the behavior-shaping or pruning power of pain with apparently no further mindlike effects, so it would be a mistake to invoke sentience wherever we find Skinnerian conditioning. Another function of pain is to disrupt normal patterns of bodily activity that might exacerbate an injury—pain causes an animal to favor an injured limb until it can mend, for instance—and this is normally accomplished by a flood of neurochemicals in a self-sustaining loop of interaction with the nervous system. Does the presence of those substances then guarantee the occurrence of pain? No, for in themselves they are just keys floating around in search of their locks; if the cycle of interaction is interrupted, there is no reason at all to suppose that pain persists. Are these particular substances even necessary for pain? Might there be creatures with a different system of locks and keys? The answer may depend more on historical processes of evolution on this planet than on any intrinsic properties of the substances. The example of the octopus shows that we should look to see what variations in chemical implementation are to be found, with what differences in function, but without expecting these facts *in themselves* to settle our question about sentience.

What then about the other features of this cycle of interaction? How rudimentary might a pain system be and still count as sentience? What would be relevant and why? Consider, for instance, a toad with a broken leg. Is this a sentient being experiencing pain? It is a living being whose normal life has been disrupted by damage to one of its parts, preventing it from engaging in the behaviors that are its way of earning a

living. It is moreover in a state with powerful negative-reinforcement potential—it can readily be conditioned to avoid such states of its nervous system. This state is maintained by a cycle of interaction that somewhat disrupts its normal dispositions to leap—though in an emergency it will leap anyway. It is tempting to see all this as amounting to pain. But it is also tempting to endow the toad with a soliloquy, in which it dreads the prospect of such an emergency, yearns for relief, deplores its relative vulnerability, bitterly regrets the foolish actions that led it to this crisis, and so forth, and these further accompaniments are not in any way licensed by anything we know about toads. On the contrary, the more we learn about toads, the more confident we are becoming that their nervous systems are designed to carry them through life without any such expensive reflective capacities.

So what? What does *sentience* have to do with such fancy intellectual talents? A good question, but that means we must try to answer it, and not just use it as a rhetorical question to deflect inquiry. Here is a circumstance in which how we ask the questions can make a huge difference, for it is possible to bamboozle ourselves into creating a phantom problem at this point. How? By losing track of where we stand in a process of addition and subtraction. At the outset, we are searching for *x*, the special ingredient that distinguishes mere sensitivity from true sentience, and we work on the project from two directions. Working up from simple cases, adding rudimentary versions of each separate feature, we tend to be unimpressed: though each of these powers is arguably an essential component of sentience, there is surely more to sentience than that—a mere robot could well exhibit *that* without any sentience at all! Working down, from our own richly detailed (and richly appreciated) experience, we recognize that other creatures manifestly lack some of the particularly human features of our experience, so we subtract them as inessential. We don't want to be unfair to our

animal cousins. So while we recognize that much of what we think of when we think of the awfulness of pain (and why it matters morally whether someone is in pain) involves imagining just these anthropomorphic accompaniments, we generously decide that they are just accompaniments, not "essential" to the brute phenomenon of sentience (and its morally most significant instance, pain). What we may tend to overlook, as these two ships pass in the night, is the possibility that we are subtracting, on one path, the very thing we are seeking on the other. *If* that's what we're doing, our conviction that we have yet to come across *x*—the "missing link" of sentience—would be a self-induced illusion.

I don't say that we *are* making an error of this sort, but just that we *might well* be doing so. That's enough for the moment, since it shifts the burden of proof. Here, then, is a conservative hypothesis about the problem of sentience: There is no such *extra* phenomenon. "Sentience" comes in every imaginable grade or intensity, from the simplest and most "robotic," to the most exquisitely sensitive, hyper-reactive "human." As we saw in chapter 1, we do indeed have to draw lines across this multistranded continuum of cases, because having moral policies requires it, but the prospect that we will *discover* a threshold—a morally significant "step," in what is otherwise a ramp—is not only extremely unlikely but morally unappealing as well.

Consider the toad once again in this regard. On which side of the line does the toad fall? (If toads are too obvious a case for you one way or the other, choose whatever creature seems to occupy your penumbra of uncertainty. Choose an ant or a jellyfish or a pigeon or a rat.) Now suppose that "science confirms" that there is minimal genuine sentience in the toad—that a toad's "pain" is real, experienced pain, for instance. The toad now qualifies for the special treatment reserved for the sentient. Now suppose instead that the toad turns out not to have *x*, once we have determined what *x* is.

In this case, the toad's status falls to "mere automaton," something that we may interfere with in any imaginable way with no moral compunction whatever. Given what we *already* know about toads, does it seem plausible that there could be some *heretofore unimagined* feature the discovery of which could justify this enormous difference in our attitude? Of course, if we discovered that toads were really tiny human beings trapped in toad bodies, like the prince in the fairy tale, we would immediately have grounds for the utmost solicitude, for we would know that in spite of all behavioral appearances, toads *were* capable of enduring all the tortures and anxieties we consider so important in our own cases. But we already know that a toad is no such thing. We are being asked to imagine that there is some *x* that is nothing at all like being a human prince trapped in a toad skin, but is nevertheless morally compelling. We also already know, however, that a toad is not a simple wind-up toy but rather an exquisitely complex living thing capable of a staggering variety of self-protective activities in the furtherance of its preordained task of making more generations of toads. Isn't that already enough to warrant some special regard on our part? We are being asked to imagine that there is some *x* that is nothing at all like this mere sophistication-of-control-structure, but that nevertheless would command our moral appreciation when we discovered it. We are being asked, I suspect, to indulge in something beyond fantasy. But let us continue with our search, to see what comes next, for we are still a long way from human minds.

FROM PHOTOTAXIS TO METAPHYSICS

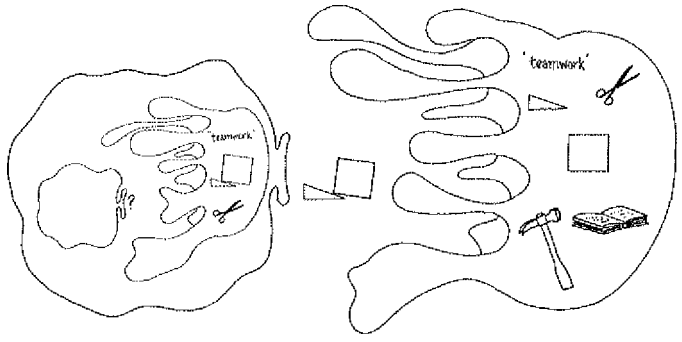
Once we get to Popperian creatures—creatures whose brains have the potential to be endowed, in inner environments,

with preselective prowess—what happens next? Many different things, no doubt, but we will concentrate on one particular innovation whose powers we can clearly see. Among the successors to mere Popperian creatures are those whose inner environments are informed by the *designed* portions of the outer environment. One of Darwin's fundamental insights is that design is expensive but copying designs is cheap; that is, making an all new design is very difficult, but redesigning old designs is relatively easy. Few of us could reinvent the wheel, but we don't have to, since we acquired the wheel design (and a huge variety of others) from the cultures we grew up in. We may call this sub-sub-subset of Darwinian creatures *Gregorian creatures*, since the British psychologist Richard Gregory is to my mind the preeminent theorist of the role of information (or more exactly, what Gregory calls Potential Intelligence) in the creation of Smart Moves (or what Gregory calls Kinetic Intelligence). Gregory observes that a pair of scissors, as a well-designed artifact, is not just a result of intelligence but an endower of intelligence (external potential intelligence), in a very straightforward and intuitive sense: when you give someone a pair of scissors, you enhance their potential to arrive more safely and swiftly at Smart Moves. (1981, pp. 311ff.)

Anthropologists have long recognized that the advent of tool use accompanied a major increase in intelligence. Chimpanzees in the wild go after termites by thrusting crudely prepared fishing sticks deep into the termites' underground homes and swiftly drawing up a stickful of termites, which they then strip off the stick into their mouths. This fact takes on further significance when we learn that not all chimpanzees have hit upon this trick; in some chimpanzee "cultures," termites are an unexploited food source. This reminds us that tool use is a two-way sign of intelligence; not only does it *require* intelligence to recognize and maintain a tool (let alone fabricate one), but a tool *confers* intelli-

gence on those lucky enough to be given one. The better designed the tool (the more information there is embedded in its fabrication), the more potential intelligence it confers on its user. And among the preeminent tools, Gregory reminds us, are what he calls mind tools: words.

Words and other mind tools give a Gregorian creature an inner environment that permits it to construct ever more subtle move generators and move testers. Skinnerian creatures ask themselves, "What do I do next?" and haven't a clue how to answer until they have taken some hard knocks. Popperian creatures make a big advance by asking themselves, "What should I think about next?" before they ask themselves, "What should I do next?" (It should be emphasized that neither Skinnerian nor Popperian creatures actually need to talk to themselves or think these thoughts. They are simply designed to operate *as if* they had asked themselves these questions. Here we see both the power and the risk of the intentional stance: The reason that Popperian creatures are smarter—more successfully devious, say—than Skinnerian creatures is that they are adaptively responsive



Gregorian creature imports mind tools from the (cultural) environment; these improve both the generators and the testers.

FIGURE 4.4

to more and better information, in a way that we can vividly if loosely describe from the intentional stance, in terms of these imaginary soliloquies. But it would be a mistake to impute to these creatures all the subtleties that go along with the ability to actually formulate such questions and answers on the human model of explicit self-questioning.) Gregorian creatures take a big step toward a human level of mental adroitness, benefiting from the experience of others by exploiting the wisdom embodied in the mind tools that those others have invented, improved, and transmitted; thereby they learn how to think better about what they should think about next—and so forth, creating a tower of further internal reflections with no fixed or discernible limit. How this step to the Gregorian level might be accomplished can best be seen by once more backing up and looking at the ancestral talents from which these most human mental talents must be constructed.

One of the simplest life-enhancing practices found in many species is *phototaxis*—distinguishing light from dark and heading for the light. Light is easily transduced, and given the way light emanates from a source, its intensity diminishing gradually as you get farther away, quite a simple connection between transducers and effectors can produce reliable phototaxis. In the neuroscientist Valentino Braitenberg's elegant little book *Vehicles*, we get the simplest model—the vehicle in figure 4.5. It has two light transducers, and their variable output signals are fed, crossed, to two effectors (think of the effectors as outboard motors). The more light transduced, the faster the motor runs. The transducer nearer the light source will drive its motor a bit faster than the transducer farther from the light, and this will always turn the vehicle in the direction of the light, till eventually it hits the light source itself or orbits tightly around it.

The world of such a simple being is graded from light to not-so-light to dark, and it traverses the gradient. It knows,

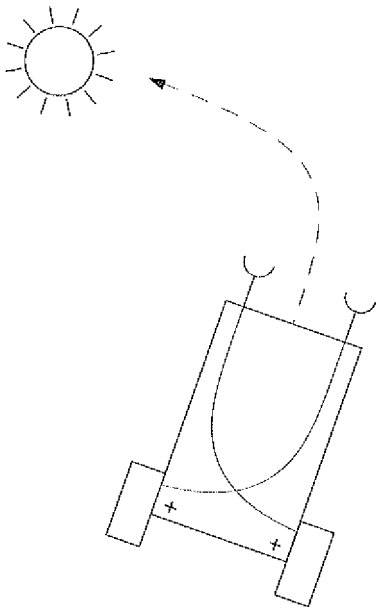


FIGURE 4.5

and needs to know, nothing else. Light *recognition* is almost for free—whatever turns on the transducer is light, and the system doesn't care whether it's the very *same* light that has returned or a new light. In a world with two moons, it might make a difference, ecologically, which moon you were tracking; moon recognition or identification could be an additional problem that needed a solution. Mere phototaxis would not be enough in such a world. In our world, a moon is not the sort of object that typically needs reidentifying by a creature; mothers, in contrast, often are.

Mamataxis—homing in on Mother—is a considerably more sophisticated talent. If Mama emitted a bright light, phototaxis might do the job, but not if there were other mothers in the vicinity, all using the same system. If Mama

then emitted a particular blue light, different from the light emitted by every other mother, then putting a particular everything-but-blue filter on each of your phototransducers would do the job quite well. Nature often relies on a similar principle, but using a more energy-efficient medium. Mama emits a signature odor, distinguishably different from all other odors (in the immediate vicinity). Mamataxis (mother-reidentification and homing) is then accomplished by odor-transduction, or olfaction. The intensity of odors is a function of the concentration of the molecular keys as they diffuse through the surrounding medium—air or water. A transducer can therefore be an appropriately shaped lock, and can follow the gradient of concentration by using an arrangement just like that in Braitenberg's vehicle. Such olfactory signatures are ancient, and potent. They have been overlaid, in our species, by thousands of other mechanisms, but their position in the foundation is still discernible. In spite of all our sophistication, odors *move* us without our knowing why or how, as Marcel Proust famously noted.*

Technology honors the same design principle in yet another medium: the EPIRB (Emergency Position Indicating Radio Beacon), a self-contained, battery-powered radio transmitter that repeats over and over again a particular signature at a particular frequency. You can buy one in a marine hardware store and take it with you on your sailboat. Then if you ever get in distress, you turn it on. Immediately

*Odors are not used only for identification signals. They often play powerful roles in attracting a mate or even suppressing the sexual activity or maturation of one's rivals. Signals from the olfactory bulb bypass the thalamus on their way to the rest of the brain, so in contrast to the signals arising in vision, hearing, and even touch, olfactory commands go directly to the old control centers, eliminating many middlemen. It is likely that this more direct route helps to explain the peremptory, nearly hypnotic power some odors have over us.

a worldwide tracking system senses your EPIRB's signal and indicates its position with a blip on an electronic map. It also looks up the signature in its giant table of signatures and thereby identifies your boat. Identification greatly simplifies search and rescue, since it adds redundancy: the beacon can be homed in on blindly by radio receivers (transducers), but as the rescuers get close it helps if they know whether they are looking (with their eyes) for a black fishing trawler, a small dark-green sailboat, or a bright-orange rubber raft. Other sensory systems can be brought in to make the final approach swifter and less vulnerable to interruption (should the EPIRB's battery run down, for instance). In animals, odor tracking is not the only medium of Mamataxis. Visual and auditory signatures are also relied on, as the ethologist Konrad Lorenz has notably demonstrated in his pioneering studies of "imprinting" in young geese and ducks. Chicks that are not imprinted shortly after birth with a proper Mama signature will fix on the first large moving thing they see and treat it as Mama thereafter.

Beacons (and their complement of beacon sensors) are good design solutions whenever one agent needs to track (recognize, reidentify) a particular entity—typically another agent, such as Mama—for a long time. You just install the beacon in the target in advance, and then let it roam. (Anti-car-theft radio beacons that you hide in your car and then remotely turn on if your car is stolen are a recent manifestation.) But there are costs, as usual. One of the most obvious is that friend and foe alike can use the tracking machinery to home in on the target. Predators are typically tuned to the same olfactory and auditory channels as offspring trying to stay in touch with Mama, for instance.

Odors and sounds are broadcast over a range that is not easily in the control of the emitter. A low-energy way of achieving a more selective beacon effect would be to put a particular blue spot (pigment of one sort or another) on

Mama, and let the reflected light of the sun create a beacon visible only in particular sectors of the world and readily extinguished by Mama's simply moving into the shadows. The offspring can then follow the blue spot whenever it is visible. But this setup requires an investment in more sophisticated photosensitive machinery: a simple eye, for instance—not just a pair of photocells.

The ability to stay in reliably close contact with one particular ecologically very important thing (such as Mama) does not require the ability to *conceive of* this thing as an enduring particular entity, coming and going. As we have just seen, reliable Mamataxis can be achieved with a bag of simple tricks. The talent is normally robust in simple environments, but a creature armed with such a simple system is easily "fooled," and when it is fooled, it trundles to its misfortune without any appreciation of its folly. There need be no capability for the system to monitor its own success or reflect on the conditions under which it succeeds or fails; that's a later (and expensive) add-on.

Cooperative tracking—tracking in which the target provides a handy beacon and thus simplifies the task for the tracker—is a step on the way toward competitive tracking, in which the target not only provides no unique signature beacon but actively tries to hide, to make itself untrackable. This move by prey is countered by the development in predators of general-purpose, track-anything systems, designed to turn *whatever aspects* a trackworthy thing reveals into a sort of private and temporary beacon—a "search image," created for the nonce by a gaggle of feature-detectors in the predator and used to correlate, moment by moment, the signature of the target, revising and updating the search image as the target changes, always with the goal of keeping the picked-out object in the cross-hairs.

It is important to recognize that this variety of tracking does not require categorization of the target. Think of a prim-

itive eye, consisting of an array of a few hundred photocells, transducing a changing pattern of pixels, which are turned on by whatever is reflecting light on them. Such a system could readily deliver a message of the following sort: "*X*, the whatever-it-is responsible for the pixel-clump currently under investigation, has just dodged to the right." (It would not have to deliver this message in so many words—there need be no words, no symbols, in the system at all.) So the only identification such a system engages in is a degenerate or minimal sort of moment-to-moment reidentification of the something-or-other being tracked. Even here, there is tolerance for change and substitution. A gradually changing clump of pixels moving against a more or less static background can change its shape and internal character radically and still be trackable, so long as it doesn't change too fast. (The *phi phenomenon*, in which sequences of flashing lights are involuntarily interpreted by the vision system to be the trajectory of a moving object, is a vivid manifestation of this built-in circuitry in our own vision systems.)

What happens when *X* temporarily goes behind a tree? The simpleminded solution is to keep the most recent version of the search image intact and then just scan around at random, hoping to lock back onto this temporary beacon once again when it emerges, if it ever does. You can improve the odds by aiming your search image at the likeliest spot for the reappearance of the temporary beacon. And you can get a better-than-a-coin-flip idea of the likeliest spot just by sampling the old trajectory of the beacon and plotting its future continuation in a straight line. This yields instances of producing future in one of its simplest and most ubiquitous forms, and also gives us a clear case of the arrow of intentionality poised on a nonexistent but reasonably hoped-for target.

This ability to "keep in touch with" another object (literally touching and manipulating it, if possible) is the prerequisite for high-quality perception. Visual recognition of a particular person or object, for instance, is almost impossible if the image of the object is not kept centered on the high-resolution fovea of the eye for an appreciable length of time. It takes time for all the epistemically hungry microagents to do their feeding and get organized. So the ability to maintain such a focus of information *about* a particular thing (the whatever-it-is I'm visually tracking right now) is a precondition for developing an identifying description of the thing.*

The way to maximize the likelihood of maintaining or restoring contact with an entity being tracked is to rely on multiple independent systems, each fallible but with overlapping domains of competence. Where one system lets down the side, the others take over, and the result tends to

.....

*This point about the primacy of tracking over description is, I think, the glimmer of truth in the otherwise forlorn philosophical doctrine that there are two varieties of belief—*de re* beliefs, which are somehow "directly" about their objects, and *de dicto* beliefs, which are about their objects only through the mediation of a *dictum*, a definite description (in a natural language, or in some "language of thought"). The contrast is illustrated (supposedly) by the difference between

believing that Tom (*that guy*, right over there) is a man,

and

believing that whoever it was that mailed this anonymous letter to me is a man.

The intentionality in the first case is supposed to be somehow more direct, to latch onto its object in a more primitive way. But, as we have seen, we can recast even in the most direct and primitive cases of perceptual tracking into the *de dicto* mode (the *x* such that *x* is whatever is responsible for the pixel-clump currently under investigation has just jumped to the right) in order to bring out a feature of the mechanism that mediates this most "immediate" sort of reference. The difference between *de re* and *de dicto* is a difference in the speaker's perspective or emphasis, not in the phenomenon. For more on this, see Dennett, "Beyond Belief" (1982).

be smooth and continuous tracking composed of intermittently functioning elements.

How are these multiple systems linked together? There are many possibilities. If you have two sensory systems, you can link them by means of an AND-gate: they both have to be turned ON by their input for the agent to respond positively. (An AND-gate can be implemented in any medium; it isn't a thing, but a principle of organization. The two keys that have to be turned to open a safe deposit box, or fire a nuclear missile, are linked by an AND-gate. When you fasten a garden hose to a spigot and put a controllable nozzle on the other end, these ON-OFF valves are linked by an AND-gate; both have to be open for water to come out.) Alternatively, you can link two sensory systems with an OR-gate: either one by itself, A *or* B (*or* both together), will evoke a positive response from the agent. OR-gates are used to include backup or spare subsystems in larger systems: if one unit fails, the extra unit's activity is enough to keep the system going. Twin-engined planes link their engines by an OR-gate: two in working order may be best, but in a pinch, one is enough.

As you add more systems, the possibility of linking them in intermediate ways looms. For instance, you can link them so that IF system A is ON, then if *either* B or C is ON, the system is to respond positively; otherwise, *both* systems B and C must be on to produce a positive response. (This is equivalent to a majority rule linking the three systems; if the majority—any majority—is ON, the system will respond positively.) All the possible ways of linking systems with AND-gates and OR-gates (and NOT-gates, which simply reverse or invert the output of a system, turning ON to OFF and vice versa) are called Boolean functions of those systems, since they can be precisely described in terms of the logical operators AND, OR, and NOT, which the nineteenth-century English mathematician George Boole first formalized. But there are also non-Boolean ways that systems can intermingle their effects. Instead of bringing all the contributors to a central voting place, giving them each a single vote (YES or NO, ON or OFF), and thereby channeling their contribution to behavior into a single vulnerable decision point (the summed effect of all the Boolean connections), we could let them maintain their own independent and continuously variable links to behavior and have the world extract an outcome behavior as the result of all the activity. Valentino Braitenberg's vehicle, with its two cross-wired phototransducers, is an utterly simple case in point. The "decision" to turn left or right emerges from the relative strength of the contributions of the two transducer-motor systems, but the effect is not efficiently or usefully represented as a Boolean function of the respective "arguments" of the transducers. (In principle, the input-output behavior of any such system can be approximated by a Boolean function of its components, suitably analyzed, but such an analytic stunt may fail to reveal what is important about the relationships. Considering the weather as a Boolean system is possible in principle, for instance, but unworkable and uninformative.)

By installing dozens or hundreds or thousands of such circuits in a single organism, elaborate life-protecting activities can be reliably controlled, all without anything happening inside the organism that looks like *thinking specific thoughts*. There is plenty of *as if* decision making, *as if* recognizing, *as if* hiding and seeking. There are also lots of ways an organism, so equipped, can "make mistakes," but its mistakes never amount to formulating a representation of some false proposition and then deeming it true.

How versatile can such an architecture be? It is hard to say. Researchers have recently designed and test-driven artificial control systems that produce many of the striking behavioral patterns we observe in relatively simple life-

forms, such as insects and other invertebrates; so it is tempting to believe that all the astonishingly complex routines of these creatures can be orchestrated by an architecture like this, even if we don't yet know how to design a system of the required complexity. After all, the brain of an insect may have only a few hundred neurons in it, and think of the elaborate engagements with the world such an arrangement can oversee. The evolutionary biologist Robert Trivers notes, for example:

> Fungus-growing ants engage in agriculture. Workers cut leaves, carry these into the nest, prepare them as a medium for growing fungus, plant fungus on them, fertilize the fungus with their own droppings, weed out competitive species by hauling them away, and, finally, harvest a special part of the fungus on which they feed. (1985, p. 172)

Then there are the prolonged and intricately articulated mating and child-rearing rituals of fish and birds. Each step has sensory requirements that must be met before it is undertaken, and then is guided adaptively through a field of obstacles. How are these intricate maneuvers controlled? Biologists have determined many of the conditions in the environment that are used as cues, by painstakingly varying the available sources of information in experiments, but it is not enough to know what information an organism can pick up. The next difficult task is figuring out how their tiny brains can be designed to put all this useful sensitivity to information to good use.

If you are a fish or a crab or something along those lines, and one of your projects is, say, building a nest of pebbles on the ocean floor, you will need a pebble-finder device, and a way of finding your way back to *your* nest to deposit the found pebble in an appropriate place before heading out again. This system need not be foolproof, however. Since impostor pebble-nests are unlikely to be surreptitiously erected in place of your own during your foray (until clever human experimenters take an interest in you), you can keep your standards for reidentification quite low and inexpensive. If a mistake in "identification" occurs, you probably go right on building, not just taken in by the ruse but completely incapable of recognizing or appreciating the error, not in the slightest bit troubled. On the other hand, if you happen to be equipped with a backup system of nest identification, and the impostor nest fails the backup test, you will be thrown into disarray, pulled in one direction by one system and in another by the other system. These conflicts happen, but it makes no sense to ask, as the organism rushes back and forth in a tizzy, "Just what is it thinking now? What is the *propositional content* of its confused state?"

In organisms such as us—organisms equipped with many layers of self-monitoring systems, which can check on and attempt to mediate such conflicts when they arise—it can sometimes be all too clear just what mistake has been made. A disturbing example is the Capgras delusion, a bizarre affliction that occasionally strikes human beings who have suffered brain damage. The defining mark of the Capgras delusion is the sufferer's conviction that a close acquaintance (usually a loved one) has been replaced by an impostor who looks like (and sounds like, and acts like) the genuine companion, who has mysteriously disappeared! This amazing phenomenon should send shock waves through philosophy. Philosophers have made up many far-fetched cases of mistaken identity to illustrate their various philosophical theories, and the literature of philosophy is crowded with fantastic thought experiments about spies and murderers traveling incognito, best friends dressed up in gorilla suits, and long-lost identical twins, but the real-life cases of Capgras delusion have so far escaped philosophers' attention.

What is particularly surprising about these cases is that they don't depend on subtle disguises and fleeting glimpses. On the contrary, the delusion persists even when the target individual is closely scrutinized by the agent, and is even pleading for recognition. Capgras sufferers have been known to murder their spouses, so sure are they that these look-alike interlopers are trying to step into shoes—into whole lives—that are not rightfully theirs! There can be no doubt that in such a sad case, the agent in question has deemed true some very specific propositions of nonidentity: *This man is not my husband*; this man is as qualitatively similar to my husband as ever can be, and yet he is not my husband. Of particular interest to us is the fact that people suffering from such a delusion can be quite unable to say why they are so sure.

The neuropsychologist Andrew Young (1994) offers an ingenious and plausible hypothesis to explain what has gone wrong. Young contrasts Capgras delusion with another curious affliction caused by brain damage: *prosopagnosia*. People with prosopagnosia can't recognize familiar human faces. Their eyesight may be fine, but they can't identify even their closest friends until they hear them speak. In a typical experiment, they are shown collections of photographs: some photos are of anonymous individuals and others are of family members and celebrities—Hitler, Marilyn Monroe, John F. Kennedy. When asked to pick out the familiar faces, their performance is no better than chance. But for more than a decade researchers have suspected that in spite of this shockingly poor performance, *something* in some prosopagnosics was correctly identifying the family members and the famous people, since their bodies react differently to the familiar faces. If, while looking at a photo of a familiar face, they are told various candidate names of the person pictured, they show a heightened galvanic skin response when they hear the right name. (The galvanic skin response is the measure of the skin's electrical conductance

and is the primary test relied on in polygraphs, or "lie detectors.") The conclusion that Young and other researchers draw from these results is that there must be two (or more) systems that can identify a face, and one of these is spared in the prosopagnosics who show this response. This system continues to do its work well, covertly and largely unnoticed. Now suppose, Young says, that Capgras sufferers have just the opposite disability: the overt, conscious face-recognition system (or systems) works just fine—which is why Capgras sufferers agree that the "impostors" do indeed look just like their loved ones—but the covert system (or systems), which normally provides a reassuring vote of agreement on such occasions, is impaired and ominously silent. The *absence* of that subtle contribution to identification is so upsetting ("Something's missing!") that it amounts to a pocket veto on the positive vote of the surviving system: the emergent result is the sufferer's heartfelt conviction that he or she is looking at an impostor. Instead of blaming the mismatch on a faulty perceptual system, the agent blames the world, in a way that is so metaphysically extravagant, so improbable, that there can be little doubt of the power (the political power, in effect) that the impaired system normally has in us all. When this particular system's epistemic hunger goes unsatisfied, it throws such a fit that it overthrows the contributions of the other systems.

In between the oblivious crab and the bizarrely mistaken Capgras sufferer there are intermediate cases. Can't a dog recognize, or fail to recognize, its master? According to Homer, when Ulysses returns to Ithaca after his twenty-year odyssey, disguised in rags as a beggar, his old dog, Argos, recognizes him, wags his tail, drops his ears, and then dies. (And Ulysses, it should be remembered, secretly wipes a tear from his own eye.) Just as there are reasons for a crab to (try to) keep track of the identity of its own nest, there are reasons for a dog to (try to) keep track of its master, among

many other important things in its world. The more pressing the reasons for reidentifying things, the more it pays not to make mistakes, and hence the more investments in perceptual and cognitive machinery will pay for themselves. Advanced kinds of learning depend, in fact, on prior capacities for (re-)identification. To take a simple case, suppose a dog sees Ulysses sober on Monday, Wednesday, and Friday, but sees Ulysses drunk on Saturday. There are several conclusions that are logically available to be drawn from this set of experiences: that there are drunk men and sober men, that one man can be drunk on one day and sober on another, that Ulysses is such a man. The dog could not—logically, could not—learn the second or third fact from this sequence of separate experiences unless it had some (fallible, but relied upon) way of reidentifying the man as the same man from experience to experience. (Millikan, forthcoming) (We can see the same principle in a more dramatic application in the curious fact that you can't—as a matter of logic—learn what you look like by looking in a mirror unless you have some *other* way of identifying the face you see as yours. Without such an independent identification, you could no more discover your appearance by looking in a mirror than you could by looking at a photograph that happened to be of you.)

Dogs live in a behavioral world much richer and more complex than the world of the crab, with more opportunities for subterfuge, bluff, and disguise, and hence with more benefits to derive from the rejection of misleading clues. But again, a dog's systems need not be foolproof. If the dog makes a mistake of identification (of either sort), we can characterize it as a case of mistaken identity without yet having to conclude that the dog is capable of *thinking* the proposition which it behaves as if it believes. Argos's behavior in the story is touching, but we mustn't let sentimentality cloud our theories. Argos might also love the smells of autumn, and respond with joy each year when the first whiff

of ripe fruit met his nostrils, but this would not show that he had any way of distinguishing between recurring season types, such as autumn, and returning individuals, such as Ulysses. Is Ulysses, to Argos, just an organized collection of pleasant smells and sounds, sights and feelings—a sort of irregularly recurring season (we haven't had one for twenty years!), during which particular behaviors are favored? It is a season that is usually sober, but some instances of it have been known to be drunk. *We* can see, from our peculiar human perspective, that Argos's success in this world will often depend on how closely his behavior approximates the behavior of an agent who, like us adult human beings, clearly distinguishes between individuals. So we find that when we interpret his behavior from the intentional stance, we do well to attribute beliefs to Argos that distinguish Ulysses from other people, strong rival dogs from weaker rival dogs, lambs from other animals, Ithaca from other places, and so forth. But we must be prepared to discover that this apparent understanding of his has shocking gaps in it—gaps inconceivable in a human being with our conceptual scheme, and hence utterly inexpressible in the terms of a human language.

Tales of intelligence in pets have been commonplace for millennia. The ancient Stoic philosopher Chrysippus reported a dog that could perform the following feat of reason: coming to a three-way fork, he sniffed down paths A and B, and *without sniffing* C, ran down C, having reasoned that if there is no scent down A and B, the quarry must have gone down C. People are less fond of telling tales of jaw-dropping stupidity in their pets, and often resist the implications of the gaps they discover in their pets' competences. Such a smart doggie, but can he figure out how to unwind his leash when he runs around a tree or a lamppost? This is not, it would seem, an unfair intelligence test for a dog—compared, say, with a test for sensitivity to irony in poetry,

or appreciation of the transitivity of *warmer-than* (if A is warmer than B, and B is warmer than C, then A is [warmer than? colder than?] C). But few if any dogs can pass it. And dolphins, for all their intelligence, are strangely unable to figure out that they could easily leap over the surrounding tuna net to safety. Leaping out of the water is hardly an unnatural act for them, which makes their obtuseness all the more arresting. As researchers regularly discover, the more ingeniously you investigate the competence of nonhuman animals, the more likely you are to discover abrupt gaps in competence. The ability of animals to generalize from their particular exploitations of wisdom is severely limited. (For an eye-opening account of this pattern in the investigation of the minds of vervet monkeys, see Cheney and Seyfarth, *How Monkeys See the World,* 1990.)

We human beings, thanks to the perspective we gain from our ability to reflect in our special ways, can discern failures of tracking that would be quite beyond the ken of other beings. Suppose Tom has been carrying a lucky penny around for years. Tom has no name for his penny, but we shall call it Amy. Tom took Amy to Spain with him, keeps Amy on his bedside table when he sleeps, and so forth. Then one day, on a trip to New York City, Tom impulsively throws Amy into a fountain, where she blends in with the crowd of other pennies, utterly indistinguishable, by Tom and by us, from all the others—at least, all the others that have the same date of issue as Amy stamped on them. Still, Tom can *reflect* on this development. He can recognize the truth of the proposition that one, and only one, of those pennies is the lucky penny that he had always carried with him. He can be bothered (or just amused) by the fact that he has irremediably lost track of something he has been tracking, by one method or another, for years. Suppose he picks up an Amy-candidate from the fountain. He can appreciate the fact

that one, and exactly one, of the following two propositions is true:

1. The penny now in my hand is the penny I brought with me to New York.
2. The penny now in my hand is not the penny I brought with me to New York.

It doesn't take a rocket scientist to appreciate that one or the other of these has to be true, even if neither Tom nor anybody else in the history of the world, past and future, can determine which. This capacity *we* have to frame, and even under most circumstances test, hypotheses about identity is quite foreign to all other creatures. The practices and projects of many creatures require them to track and reidentify individuals—their mothers, their mates, their prey, their superiors and subordinates in their band—but no evidence suggests they must appreciate that this is what they are doing when they do it. Their intentionality never rises to the pitch of metaphysical particularity that ours can rise to.

How do we do it? It doesn't take a rocket scientist to think such thoughts, but it does take a Gregorian creature who has language among its mind tools. But in order to use language, we have to be specially equipped with the talents that permit us to extract these mind tools from the (social) environment in which they reside.