# 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.

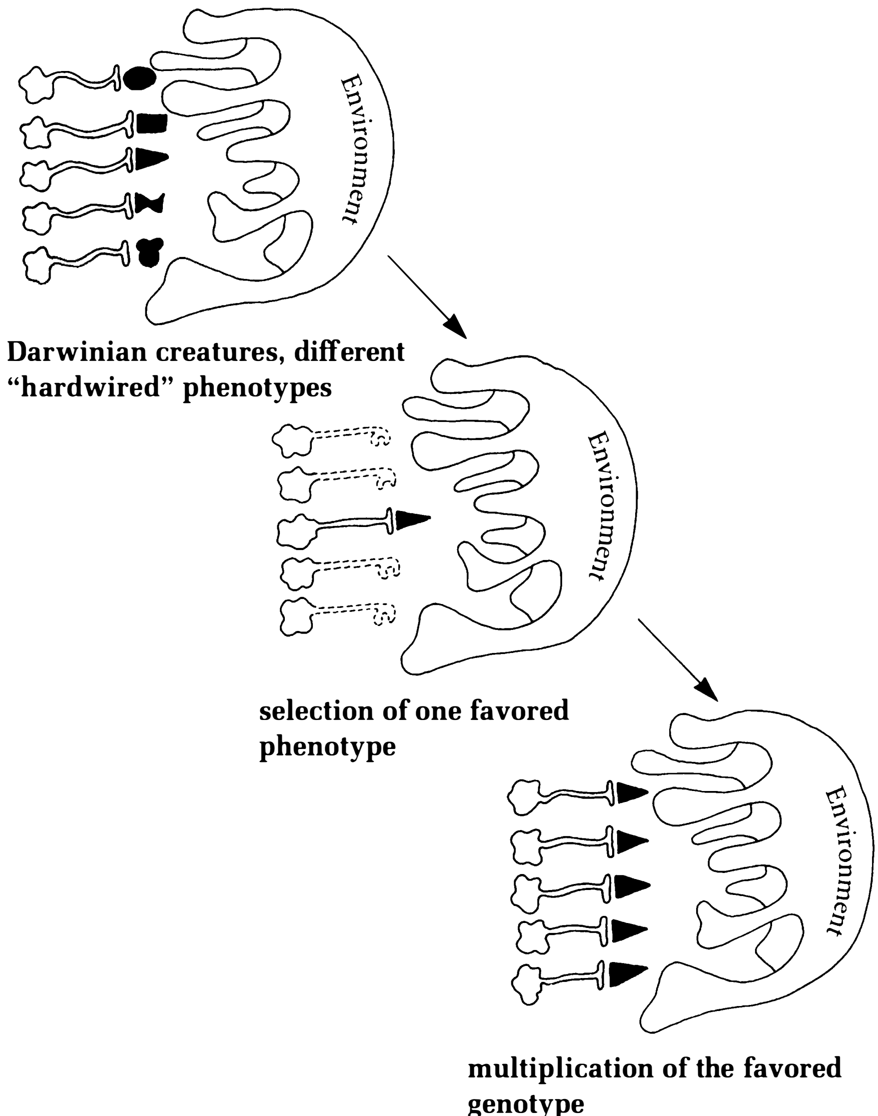


**FIGURE 4.1**

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-

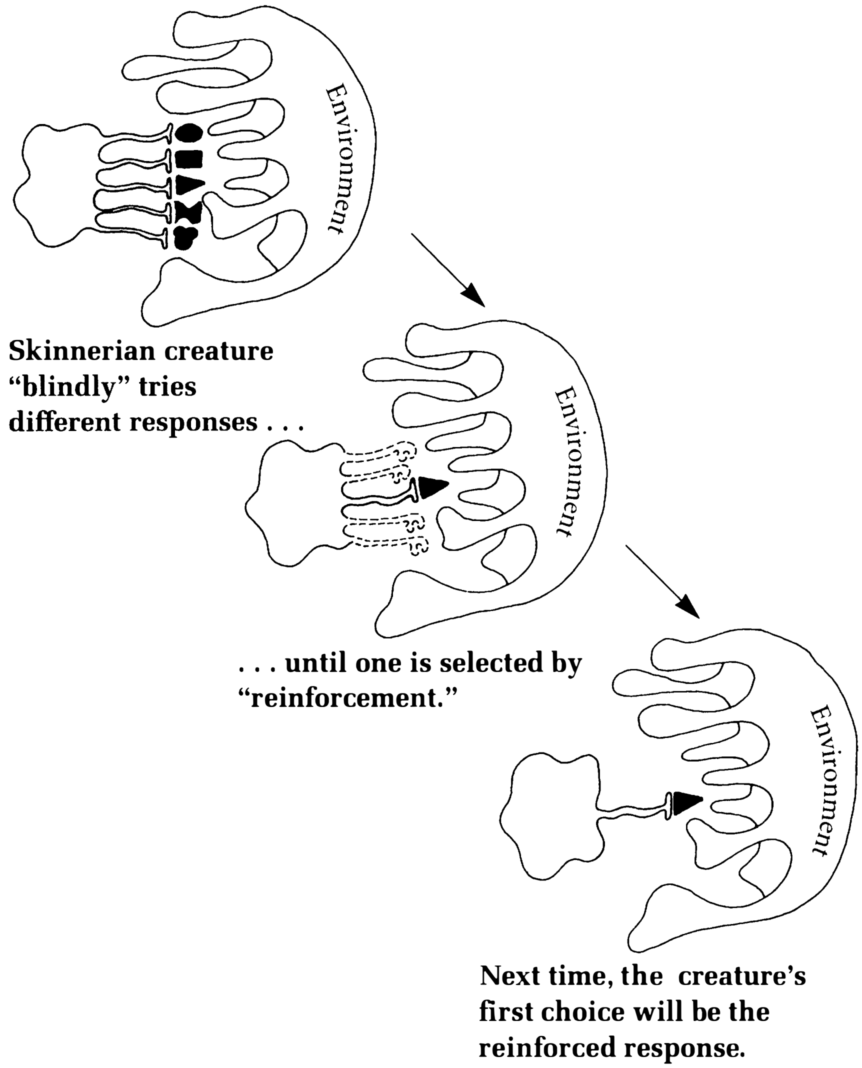


**FIGURE 4.2**

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

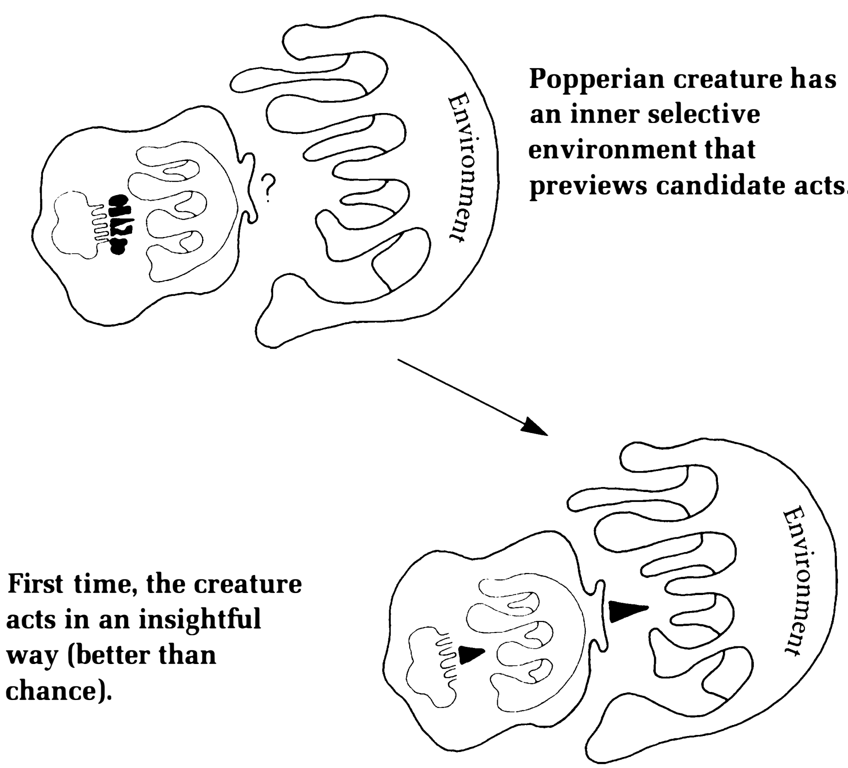


**FIGURE 4.3**

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-