# Being There
## Putting Brain, Body, and World Together Again

Andy Clark

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

# Introduction: A Car with a Cockroach Brain

Where are the artificial minds promised by 1950s science fiction and 1960s science journalism? Why are even the best of our "intelligent" artifacts still so unspeakably, terminally dumb? One possibility is that we simply misconstrued the nature of intelligence itself. We imagined mind as a kind of logical reasoning device coupled with a store of explicit data—a kind of combination logic machine and filing cabinet. In so doing, we ignored the fact that minds evolved to make things happen. We ignored the fact that the biological mind is, first and foremost, an organ for controlling the biological body. Minds make motions, and they must make them fast—before the predator catches you, or before your prey gets away from you. Minds are *not* disembodied logical reasoning devices.

This simple shift in perspective has spawned some of the most exciting and groundbreaking work in the contemporary study of mind. Research in "neural network" styles of computational modeling has begun to develop a radically different vision of the computational structure of mind. Research in cognitive neuroscience has begun to unearth the often-surprising ways in which real brains use their resources of neurons and synapses to solve problems. And a growing wave of work on simple, real-world robotics (for example, getting a robot cockroach to walk, seek food, and avoid dangers) is teaching us how biological creatures might achieve the kinds of fast, fluent real-world action that are necessary to survival. Where these researches converge we glimpse a new vision of the nature of biological cognition: a vision that puts explicit data storage and logical manipulation in its place as, at most, a secondary adjunct to the kinds of dynamics and complex response loops that couple real brains,

bodies, and environments. Wild cognition, it seems, has (literally) no time for the filing cabinet.

Of course, not everyone agrees. An extreme example of the opposite view is a recent $50 million attempt to instill commonsense understanding in a computer by giving it a vast store of explicit knowledge. The project, known as CYC (short for "encyclopedia"), aims to handcraft a vast knowledge base encompassing a significant fraction of the general knowledge that an adult human commands. Begun in 1984, CYC aimed at encoding close to a million items of knowledge by 1994. The project was to consume about two person-centuries of data-entry time. CYC was supposed, at the end of this time, to "cross over": to reach a point where it could directly read and assimilate written texts and hence "self-program" the remainder of its knowledge base.

The most noteworthy feature of the CYC project, from my point of view, is its extreme faith in the power of explicit symbolic representation: its faith in the internalization of structures built in the image of strings of words in a public language. The CYC representation language encodes information in units ("frames") such as the following:

Missouri
Capital: (Jefferson City)
Residents: (Andy, Pepa, Beth)
State of: (United States of America)

The example is simplified, but the basic structure is always the same. The unit has "slots" (the three subheadings above), and each slot has as its value a list of entities. Slots can reference other units (for example, the "residents" slot can act as a pointer to another unit containing still more information, and so on and so on). This apparatus of units and slots is augmented by a more powerful language (the CycL Constraint language) that allows the expression of more complex logical relationships, such as "For all items, if the item is an X then it has property Y." Reasoning in CYC can also exploit any of several simple inference types. The basic idea, however, is to let the encoded knowledge do almost all the work, and to keep inference and control structure simple and within the bounds of current technology. CYC's creators, Douglas Lenat and Edward Feigenbaum (1992, p. 192), argue that the bottleneck for adaptive intelligence is *knowledge*, not inference or control.



The CYC knowledge base attempts to make explicit all the little things we know about our world but usually wouldn't bother to say. CYC thus aims to encode items of knowledge we all have but seldom rehearse—items such as the following (ibid., p. 197):

Most cars today are riding on four tires.
If you fall asleep while driving, your car will start to head out of your lane pretty soon.
If something big is between you and the thing you want, you probably will have to go around it.

By explicitly encoding a large fraction of this "consensus reality knowledge," CYC is supposed to reach a level of understanding that will allow it to respond with genuine intelligence. It is even hoped that CYC will use analogical reasoning to deal sensibly with novel situations by finding partial parallels elsewhere in its vast knowledge base.

CYC is an important and ambitious project. The commonsense data base it now encodes will doubtless be of great practical use as a resource for the development of better expert systems. But we should distinguish two possible goals for CYC. One would be to provide the best simulacrum of commonsense understanding possible within a fundamentally unthinking computer system. The other would be to create, courtesy of the CYC knowledge base, the first example of a genuine artificial mind.

Nothing in the performance of CYC to date suggests that the latter is in the cards. CYC looks set to become a bigger, fancier, but still fundamentally brittle and uncomprehending "expert system." Adding more and more knowledge to CYC will not remedy this. The reason is that CYC lacks the most basic kinds of adaptive responses to an environment. This shortcoming has nothing to do with the relative paucity of the knowledge the system explicitly encodes. Rather, it is attributable to the lack of any fluent coupling between the system and a real-world environment posing real-world problems of acting and sensing. Even the lowly cockroach, as we shall see, displays this kind of fluent coupling—it displays a version of the kind of robust, flexible, practical intelligence that most computer systems so profoundly lack. Yet such a simple creature can hardly be

accused of commanding a large store of explicitly represented knowledge! Thus, the CYC project, taken as an attempt to create genuine intelligence and understanding in a machine, is absolutely, fundamentally, and fatally flawed. Intelligence and understanding are rooted not in the presence and manipulation of explicit, language-like data structures, but in something more earthy: the tuning of basic responses to a real world that enables an embodied organism to sense, act, and survive.

This diagnosis is not new. Major philosophical critics of AI have long questioned the attempt to induce intelligence by means of disembodied symbol manipulation and have likewise insisted on the importance of situated reasoning (that is, reasoning by embodied beings acting in a real physical environment). But it has been all too easy to attribute such doubts to some sort of residual mysticism—to unscientific faith in a soul-like mental essence, or to a stubborn refusal to allow science to trespass on the philosophers' favorite terrain. But it is now increasingly clear that the alternative to the "disembodied explicit data manipulation" vision of AI is not to retreat from hard science; it is to pursue some even harder science. It is to put intelligence where it belongs: in the coupling of organisms and world that is at the root of daily, fluent action. From CYC to cycle racing: such is the radical turn that characterizes the new sciences of the embodied mind.

Take, for example, the humble cockroach. The roach is heir to a considerable body of cockroach-style commonsense knowledge. At least, that is how it must appear to any theorist who thinks explicit knowledge is the key to sensible-looking real-world behavior! For the roach is a formidable escape artist, capable of taking evasive action that is shaped by a multitude of internal and external factors. Here is a brief list, abstracted from Ritzmann's (1993) detailed study, of the escape skills of the American cockroach, *Periplaneta americana*:

The roach senses the wind disturbance caused by the motion of an attacking predator.

It distinguishes winds caused by predators from normal breezes and air currents.

It does not avoid contact with other roaches.

When it does initiate an escape motion, it does not simply run at random. Instead, it takes into account its own initial orientation, the presence of obstacles (such as walls and corners), the degree of illumination, and the direction of the wind.



No wonder they always get away! This last nexus of contextual considerations, as Ritzmann points outs, leads to a response that is much more intelligent than the simple "sense predator and initiate random run" reflex that cockroach experts (for such there be) once imagined was the whole story. The additional complexity is nicely captured in Ritzmann's descriptions of a comparably "intelligent" automobile. Such a car would sense approaching vehicles, but it would ignore those moving in normal ways. If it detected an impeding collision, it would automatically initiate a turn that took its own current state (various engine and acceleration parameters) into account, took account of the road's orientation and surface, and avoided turning into other dangers. A car with the intelligence of a cockroach, it seems clear, would be way ahead of the current state of the automotive art. "Buy the car with the cockroach brain" does not immediately strike you as a winner of an advertising slogan, however. Our prejudice against basic forms of biological intelligence and in favor of bigger and fancier "filing cabinet/logic machines" goes all too deep.

How *does* the roach manage its escapes? The neural mechanisms are now beginning to be understood. Wind fronts are detected by two cerci (antenna-like structures located at the rear of the abdomen). Each cercus is covered with hairs sensitive to wind velocity and direction. Escape motions are activated only if the wind is accelerating at 0.6 $m/s^2$ or more: this is how the creature discriminates ordinary breezes from the lunges of attackers. The interval between sensing and response is very short: 58 milliseconds for a stationary roach and 14 milliseconds for a walking roach. The initial response is a turn that takes between 20 and 30 milliseconds (Ritzmann 1993, pp. 113–116). The basic neural circuitry underlying the turn involves populations of neurons whose locations and connectivity are now quite well understood. The circuitry involves more than 100 interneurons that act to modulate various turning commands in the light of contextual information concerning the current location of the roach and the state of the local environment. The basic wind information is carried by a population of ventral giant interneurons, but the final activity builds in the results of modulation from many other neuronal populations sensitive to these other contextual features.

Confronted with the cockroach's impressive display of sensible escape routines, a theorist might mistakenly posit some kind of stored quasi-linguistic database. In the spirit of CYC, we might imagine that the roach is accessing knowledge frames that include such items as these:

If you are being attacked, don't run straight into a wall.
If something big is between you and the food, try to go around it.
Gentle breezes are not dangerous.

As the philosopher Hubert Dreyfus (1991) and others have pointed out, the trouble is that real brains don't seem to use such linguaform, text-like resources to encode skillful responses to the world. And this is just as well, since such strategies would require vast amounts of explicit data storage and search and could thus not yield the speedy responses that real action requires. In fact, a little reflection suggests that there would be no obvious end to the "commonsense" knowledge we would have to write down to capture all that an adult human knows. Even the embodied knowledge of a cockroach would probably require several volumes to capture in detail!

But how else might AI proceed? One promising approach involves what has become known as *autonomous-agent theory*. An autonomous agent is a creature capable of survival, action, and motion in real time in a complex and somewhat realistic environment. Many existing artificial autonomous agents are real robots that are capable of insect-style walking and obstacle avoidance. Others are computer simulations of such robots, which can thus move and act only in simulated, computer-based environments. There are disputes between researchers who favor only real-world settings and real robots and researchers who are happy to exploit "mere" simulations, but the two camps concur in stressing the need to model realistic and basic behaviors and in distrusting overintellectualized solutions of the "disembodied explicit reasoning" stripe.

With this general image of autonomous-agent research in mind, let us return very briefly to our hero, the cockroach. Randall Beer and Hillel Chiel have created plausible computer and robot simulations of cockroach locomotion and escape. In modeling the escape response, Beer and Chiel set out to develop an autonomous-agent model highly constrained by ethological and neuroscientific data. The goal was, thus, to stay as close to the real biological data as is currently possible. To this end, they combined the autonomous-agent methodology with neural-network-style modeling. They also constrained this computational model in ways consistent with what is known about the actual neural organization of (in this case) the cockroach. They used a neural net to control the body of a simulated insect (Beer and Chiel 1993). The net circuitry was constrained by known facts about the neural populations and connectivities underlying the escape response in real cockroaches. After training, the neural network controller was able to reproduce in the simulated insect body all the main features of the escape response discussed earlier. In the chapters that follow, we shall try to understand something of how such successes are achieved. We shall see in detail how the types of research just sketched combine with developmental, neuroscientific, and psychological ideas in ways that can illuminate a wide range of both simple and complex behaviors. And we shall probe the surprising variety of adaptive strategies available to embodied and environmentally embedded agents—beings that move and that act upon their worlds.



These introductory comments set out to highlight a fundamental contrast: to conjure the disembodied, atemporal intellectualist vision of mind, and to lay beside it the image of mind as a controller of embodied action. The image of mind as controller forces us to take seriously the issues of time, world, and body. Controllers must generate appropriate actions, rapidly, on the basis of an ongoing interaction between the body and its changing environment. The classical AI planning system can sit back and take its time, eventually yielding a symbolically couched description of a plausible course of action. The embodied planning agent must take action fast—before the action of another agent claims its life. Whether symbolic, text-like encodings have any role to play in these tooth-and-claw decisions is still uncertain, but it now seem clear that they do not lie at its heart.

The route to a full computational understanding of mind is, to borrow a phrase from Lenat and Feigenbaum, blocked by a mattress in the road. For many years, researchers have swerved around the mattress, tried to finesse it away, done just about anything except get down to work to shift it. Lenat and Feigenbaum think the mattress is knowledge—that the puzzles of mind will fall away once a nice big knowledge base, complete with

explicit formulations of commonsense wisdom, is in place. The lessons of wild cognition teach us otherwise. The mattress is not knowledge but basic, real-time, real-world responsiveness. The cockroach has a kind of common sense that the best current artificial systems lack—no thanks, surely, to the explicit encodings and logical derivations that may serve us in a few more abstract domains. At root, our minds too are organs for rapidly initiating the next move in real-world situations. They are organs exquisitely geared to the production of actions, laid out in local space and real time. Once mind is cast as a controller of bodily action, layers upon layers of once-received wisdom fall away. The distinction between perception and cognition, the idea of executive control centers in the brain, and a widespread vision of rationality itself are all called into question. Under the hammer too is the methodological device of studying mind and brain with scant regard for the properties of the local environment or the opportunities provided by bodily motion and action. The fundamental shape of the sciences of the mind is in a state of flux. In the chapters to follow, we will roam the landscape of mind in the changing of the light.

# I

# Outing the Mind

Well, what do you think you understand with? With your head? Bah!
—Nikos Kazantzakis, *Zorba the Greek*

Ninety percent of life is just being there.
—Woody Allen

# 1

# Autonomous Agents: Walking on the Moon

## 1.1 Under the Volcano[1]

In the summer of 1994, an eight-legged, 1700-pound robot explorer named Dante II rapelled down a steep slope into the crater of an active volcano near Anchorage, Alaska. During the course of a six-day mission, Dante II explored the slope and the crater bed, using a mixture of autonomous (self-directed) and external control. Dante II is one product of a NASA-funded project, based at Carnegie Mellon University and elsewhere, whose ultimate goal is to develop truly autonomous robots for the purpose of collecting and transmitting detailed information concerning local environmental conditions on other planets. A much smaller, largely autonomous robot is expected to be sent to Mars in 1996, and the LunaCorp lunar rover, which is based on Dante II software, has a reserved spot on the first commercial moon shot, planned for 1997.

The problems faced by such endeavors are instructive. Robots intended to explore distant worlds cannot rely on constant communication with earth-based scientists—the time lags would soon lead to disaster. Such robots must be programmed to pursue general goals by exploring and transmitting information. For long missions, they will need to replenish their own energy supplies, perhaps by exploiting solar power. They will need to be able to function in the face of unexpected difficulties and to withstand various kinds of damage. In short, they will have to satisfy many (though by no means all) of the demands that nature made on evolving mobile organisms.

The attempt to build robust mobile robots leads, surprisingly quickly, to a radical rethinking of many of our old and comfortable ideas about the nature of adaptive intelligence.

## 1.2 The Robots' Parade

### Elmer and Elsie

The historical forebears of today's sophisticated animal-like robots (sometimes called "animats") were a pair of cybernetic "turtles" built in 1950 by the biologist W. Grey Walter. The "turtles"—named Elmer and Elsie[2]—used simple light and touch sensors and electronic circuitry to seek light but avoid intense light. In addition, the turtles each carried indicator lights, which came on when their motors were running. Even such simple onboard equipment led to thought-provoking displays of behavior, especially when Elmer and Elsie interacted both with each other (being attracted by the indicator lights) and with the local environment (which included a few light sources which they would compete to be near, and a mirror which led to amusing, self-tracking "dancing"). In a strange way, the casual observer would find it easier to read life and purpose into the behavior of even these shallow creations than into the disembodied diagnostics of fancy traditional expert systems such as MYCIN.[3]

### Herbert

One of the pioneers of recent autonomous-agent research is Rodney Brooks of the MIT Mobile Robot Laboratory. Brooks's mobile robots ("mobots") are real robots capable of functioning in messy and unpredictable real-world settings such as a crowded office. Two major characteristics of Brooks's research are the use of *horizontal* microworlds and the use of *activity-based decompositions* within each horizontal slice.

The contrast between horizontal and vertical microworlds is drawn in Clark 1989 and, in different terms, in Dennett 1978b. The idea is simple. A microworld is a restricted domain of study: we can't solve all the puzzles of intelligence all at once. A vertical microworld is one that slices off a small piece of human-level cognitive competence as an object of study. Examples include playing chess, producing the past-tense forms of English verbs, and planning a picnic, all of which have been the objects of past AI programs. The obvious worry is that when we human beings solve these advanced problems we may well be bringing to bear computational resources shaped by the other, more basic needs for which evolution equipped our ancestors. Neat, design-oriented solutions to these recent problems may thus be quite unlike the natural solutions dictated by the need to exploit existing machinery and solutions. We may be chess masters courtesy of pattern-recognition skills selected to recognize mates, food, and predators. A horizontal microworld, in contrast, is the complete behavioral competence of a whole but relatively simple creature (real or imaginary). By studying such creatures, we simplify the problems of human-level intelligence without losing sight of such biological basics as real-time response, integration of various motor and sensory functions, and the need to cope with damage.



Brooks (1991, p. 143) lays out four requirements for his artificial creatures:

> A creature must cope appropriately and in a timely fashion with changes in its dynamic environment.
> A creature should be robust with respect to its environment. . . .
> A creature should be able to maintain multiple goals. . . .
> A creature should do something in the world; it should have some purpose in being.

Brooks's "creatures" are composed of a number of distinct activity-producing subsystems or "layers." These layers do not create and pass on explicit, symbolic encodings or recodings of inputs. Instead, each layer is itself a complete route from input to action. The "communication" between distinct layers is restricted to some simple signal passing. One layer can encourage, interrupt, or override the activity of another. The resultant setup is what Brooks calls a "subsumption architecture" (because layers can subsume one another's activity but cannot *communicate* in more detailed ways).

A creature might thus be composed of three layers (Brooks 1991, p. 156):

> Layer 1: Object avoidance via a ring of ultrasonic sonar sensors. These cause the mobot to *halt* if an object is dead ahead and allow reorientation in an unblocked direction.
> Layer 2: If the object avoidance layer is currently inactive, an onboard device can generate random course headings so the mobot "wanders."
> Layer 3: This can surpass the wander layer and instead set up a distant goal to take the mobot into a whole new locale.

A key feature of the methodology is that layers can be added incrementally, each such increment yielding a whole, functional creature. Notice that such creatures do not depend on a central reservoir of data or on a central planner or reasoner. Instead, we see a "collection of competing behaviors" orchestrated by environmental inputs. There is no clear dividing line between perception and cognition, no point at which perceptual inputs are translated into a central code to be shared by various onboard reasoning devices. This image of multiple, special-purpose problem solvers orchestrated by environmental inputs and relatively simple kinds of internal signaling is, I shall argue in a later chapter, a neuroscientifically plausible model even of more advanced brains.

Herbert,[4] built at the MIT Mobot Lab in the 1980s, exploits the kind of subsumption architecture just described. Herbert's goal was to collect empty soft-drink cans left strewn around the laboratory. This was not a trivial task; the robot had to negotiate a cluttered and changing environment, avoid knocking things over, avoid bumping into people, and identify and collect the cans. One can imagine a classical planning device trying to solve this complex real-world problem by using rich visual data to generate a detailed internal map of the present surroundings, to isolate the cans, and to plan a route. But such a solution is both costly and fragile—the environment can change rapidly (as when someone enters the room), and rich visual processing (e.g. human-level object and scene recognition) is currently beyond the reach of any programmed system.

Subsumption architectures, as we saw, take a very different approach. The goal is to have the complex, robust, real-time behavior emerge as the result of simple interactions between relatively self-contained behavior-producing subsystems. These subsystems are, in turn, controlled rather directly by properties of the encountered environment.[5] There is no central control or overall plan. Instead, the environment itself will guide the creature, courtesy of some basic behavioral responses, to success. In Herbert's case, these simple behaviors included obstacle avoidance (stopping, reorienting, etc.) and locomotion routines. These would be interrupted if a table-like outline was detected by a simple visual system. Once Herbert was beside a table, the locomotion and obstacle-avoidance routines ceded control to other subsystems that swept the table with a laser and a video camera. Once the basic outline of a can was detected, the robot would rotate until the can-like object was in the center of its field of vision. At this point, the wheels stopped and a robot arm was activated. The arm, equipped with simple touch sensors, gently explored the table surface ahead. When Herbert encountered the distinctive shape of a can, grasping behavior ensued, the can was collected, and the robot moved on.



Herbert is thus a simple "creature" that commands no stored long-term plans or models of its environment. Yet, considered as an artificial animal foraging for cans in the sustaining niche provided by the Mobot Lab ecosystem, Herbert exhibits a kind of simple adaptive intelligence in which sensors, onboard circuitry, and external environment cooperate to ensure success.

### Attila

Rodney Brooks believes that robots smaller and more flexible than the lumbering Dante will better serve the needs of extraterrestrial exploration. Attila[6] weighs just $3\frac{1}{2}$ pounds and uses multiple special-purpose "mini-brains" ("finite-state machines") to control a panoply of local behaviors which together yield skilled walking: moving individual legs, detecting the forces exerted by the terrain so as to compensate for slopes, and so on. Attila also exploits infrared sensors to detect nearby objects. It is able to traverse rough terrain, and even to stand up again if it should fall on its back. Rodney Brooks claims that Attila already embodies something close to insect-level intelligence.

### Periplaneta Computatrix

This is the simulated cockroach mentioned above. Beer and Chiel (1993) describe a neural-network controller for hexapod locomotion. Each leg has a mini-controller that exploits a "pacemaker" unit—an idealized model neuron whose output oscillates rhythmically. The unit will fire at intervals determined by the tonic level of excitation from a command neuron and any additional inputs it receives. The idea, borrowed from a biological model developed by K. G. Pearson (1976), is to give each leg its own rhythmic-pattern generator but then to factor in modulatory local influences involving the different sensory feedbacks from each leg as the insect traverses uneven terrain. Coordination between legs is achieved by

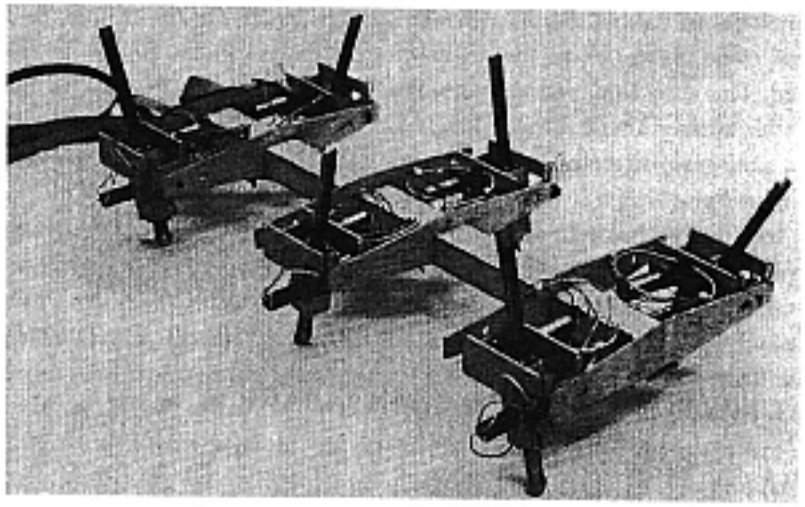

Figure 1.1
The first hexapod robot, built by Ken Espenschied at Case Western Reserve University under the supervision of Roger Quinn. Source: Quinn and Espenschied 1993. Reproduced by kind permission of K. Espenschied, R. Quinn, and Academic Press.

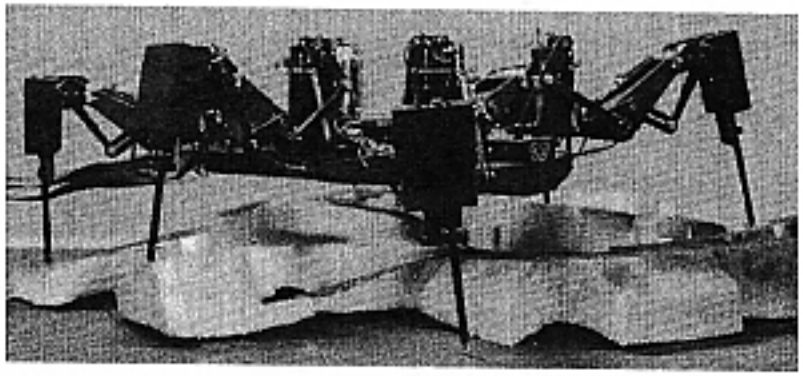

Figure 1.2
The second hexapod robot, built by Ken Espenschied at Case Western Reserve University under the supervision of Roger Quinn. Photograph courtesy of Randall Beer.



inhibitory links between neighboring pattern generators. Each leg has three motor neurons: one controls back swing, one controls forward swing, and one causes the foot to raise. The overall control circuit is again fully distributed. There is no central processor that must orchestrate a response by taking all sensory inputs into account. Instead, each leg is individually "intelligent," and simple inhibitory linkages ensure globally coherent behavior. Different gaits emerge from the interactions between different levels of tonic firing of the pacemaker units (the pattern generators) and local sensory feedback. The robot will adopt a *tripod* gait at high firing frequencies and will switch to a *metachronal* gait at lower ones. In a tripod gait, the front and back legs on one side swing in phase with the middle legs on the other side; in a metachronal gait, each leg begins its swing just after the leg behind it, in a kind of wave or ripple motion.

Although designed and tested as a pure computer simulation, the locomotion circuit has been used in a real robot body and has proved robust in the real world of friction, inertia, noise, delays, and so on. An early example of a real-world hexapod robot is shown is figure 1.1 and is further discussed in Beer and Chiel 1993 and in Quinn and Espenschied 1993. The locomotion circuit employed is also able (because it is so highly distributed) to preserved most of its functionality after damage to individual neurons or connections (Beer et al. 1992). Despite the complexity of the behavior it produces, the locomotion circuit itself is quite modest—just 37 "neurons," strategically deployed and interconnected. Nonetheless, videos of the robot hexapod and its successors provide an enthralling spectacle. One sequence shows a somewhat more complex successor robot (figure 1.2) tentatively making its way across the rough terrain provided by some fragments of polystyrene packing. A foot is extended and gently lowered. Finding no purchase (because of the local terrain), it is retracted and then placed in a slightly different location. Eventually a suitable foothold is discovered and the robot continues on its way. Such tentative exploratory behavior has all the flavor of real, biological intelligence.

### Brachiation Robot

Brachiation (figure 1.3) is the branch-to-branch swinging motion that apes use to traverse highly forested terrain. Saito and Fukuda (1994) describe a robot device that learns to brachiate using a neural-network

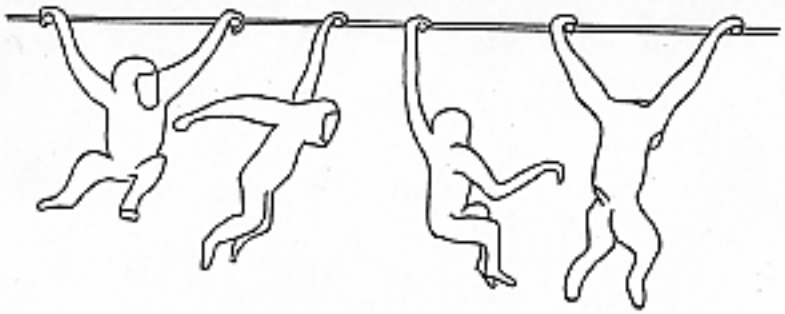

Figure 1.3
The brachiation of a gibbon. Source: Saito and Fukuda 1994. Used by kind permission of F. Saito, T. Fukuda, and MIT Press.

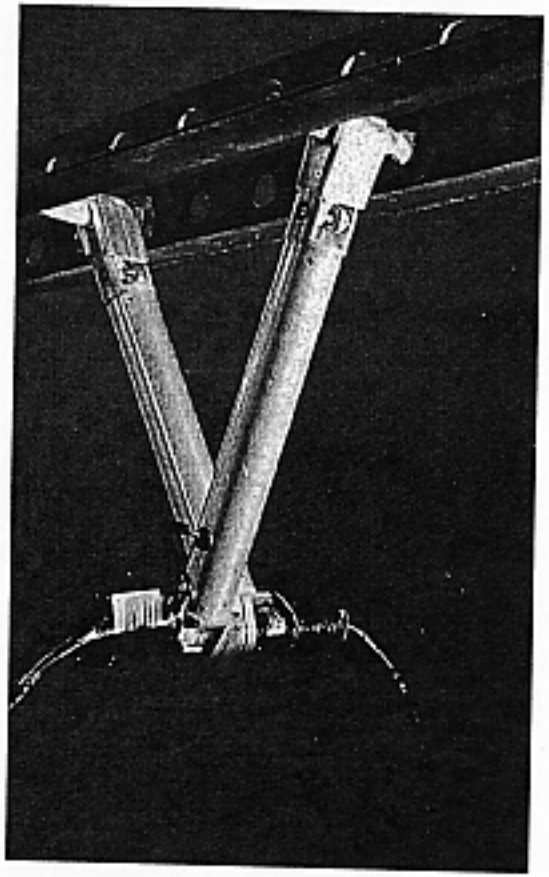

Figure 1.4
A two-link brachiation robot . Source: Saito and Fukuda 1994. Used by kind permission of F. Saito, T. Fukuda, and MIT Press.



controller. The task is especially interesting since it incorporates a lea ing dimension and addresses a highly time-critical behavior.

The robot uses a form of neural-network learning called *connectic ist Q-learning*.[7] Q-learning involves attempting to learn the value of d ferent actions in different situations. A Q-learning system must have delimited set of possible actions and situations and must be provid with a reward signal informing it of the value (goodness) of a chos action in the situation it is facing. The goal is to learn a set of situatio action pairings that will maximize success relative to a reward signal. Sai and Fukuda demonstrate that such techniques enable an artificial neu network to learn to control a two-link real-world brachiation robot (fi ure 1.4). The fully trained brachiation robot can swing successfully fro "branch" to "branch," and if it misses it is able to use its momentum swing back and try again.

## COG

COG (Brooks 1994; Brooks and Stein 1993) must surely be the mo ambitious of all the "New Robotics" projects undertaken so far. The pr ject, spearheaded by Rodney Brooks, aims to create a high-functionir humanoid robot. The human-size robot (figure 1.5) is not mobile; it i however, able to move its hands, arms, head, and eyes. It is bolted to tabletop, but it can swivel at the hips. There are 24 individual moto underpinning these various degrees of freedom, and each motor has processor devoted solely to overseeing its operation (in line with the ge eral mobot ethos of avoiding centralized control). The arms incorporat springs, which allow some brute-mechanical smoothing. Most of th motors (excluding the eye motors) incorporate heat sensors that allo COG to gather information about its own current workings by telling how hard various motors are working—a kind of robot version of th kinesthetic sense that tells us how our body parts are oriented in spac Each eye each comprises two cameras; one has a wide field of view wit low resolution, and the other has a narrow field of view with high reso lution. The cameras can move around surveying a visual scene, with th narrow-field camera mimicking the mammalian fovea. COG also receive audio information via four microphones. All this rich incoming data i processed by a "brain" composed of multiple submachines ("nodes,

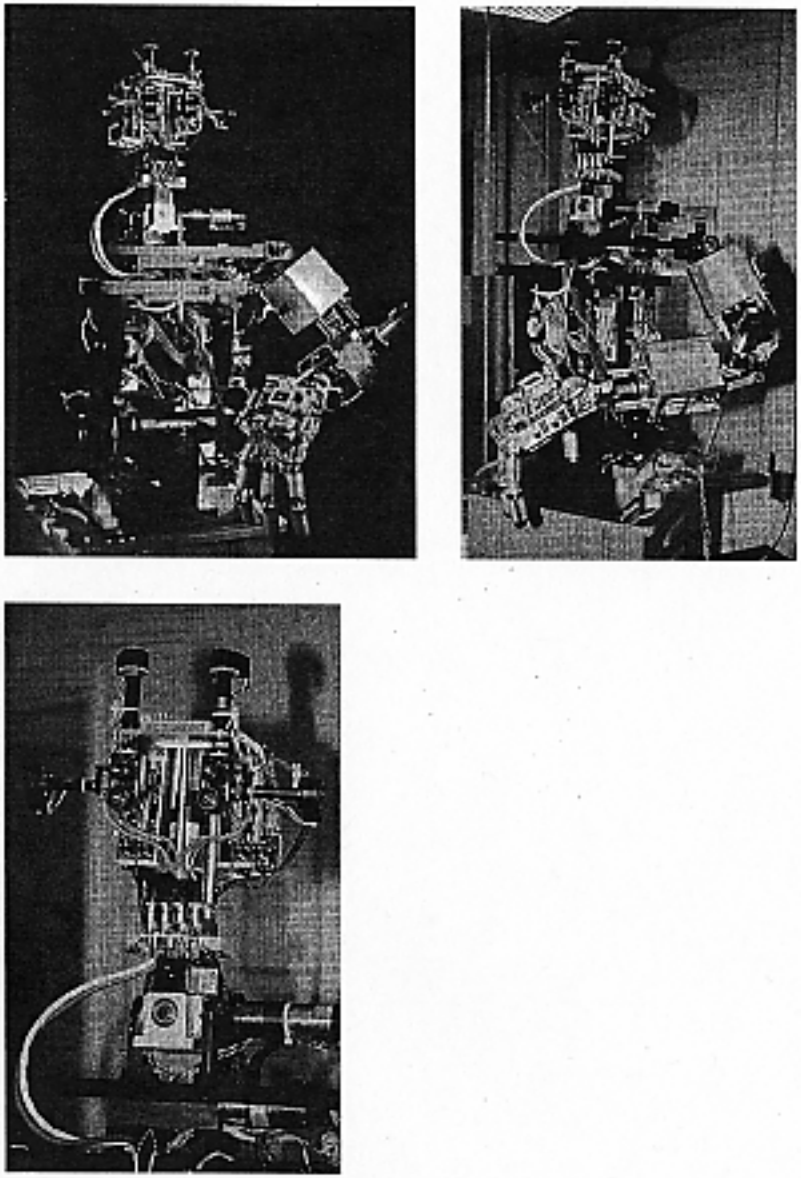

Figure 1.5
Three views of the robot COG. Photographs kindly provided by Rodney Brooks.



each with a megabyte of ROM and RAM and a dedicated operating system), which are capable of communicating with one another in some restricted ways. COG's brain is thus itself a multi-processor system, and COG's nervous system also includes other "intelligent" devices (such as the dedicated motor processors). The overall setup thus reflects much of the guiding philosophy of Brooks's work with robot insects, but it is sufficiently complex to bring new and pressing problems to the fore. Familiar features include the lack of any central memory shared by all processors, the lack of any central executive controls, the restricted communications between subdevices, and the stress on solving real-time problems involving sensing and acting. The new problems all center around the need to press coherent behaviors from such a complex system without falling back on the old, impractical methods of serial planning and central control. The ingenious strategies and tricks that enable embodied systems to maintain coherence while exploiting multiple, special-purpose, quasi-independent problem-solving routines (addressed in later chapters) shed light on the roles of language, culture, and institutions in empowering human cognition. For the moment, however, let us back off and try to extract some general morals from our parade of artificial critters.

## 1.3 Minds without Models

The New Robotics revolution rejects a fundamental part of the classical image of mind. It rejects the image of a *central planner* that is privy to all the information available anywhere in the system and dedicated to the discovery of possible behavioral sequences that will satisfy particular goals. The trouble with the central planner is that is profoundly impractical. It introduces what Rodney Brooks aptly termed a "representational bottleneck" blocking fast, real-time response. The reason is that the incoming sensory information must be converted into a single symbolic code so that such a planner can deal with it. And the planners' output will itself have to be converted from its propriety code into the various formats needed to control various types of motor response. These steps of translation are time-consuming and expensive.

Artificial critters like Herbert and Attila are notable for their lack of central planning. In its place the subsumption architecture puts multiple

quasi-independent devices, each of which constitutes a self-contained pathway linking sensory input to action. As a result, the behaviors of such systems are not mediated by any integrated knowledge base depicting the current state of the overall environment. Such knowledge bases are often called "detailed world models," and it is a recurring theme of the new approaches that they achieve adaptive success without the use of such models.

It would be easy, however, to overstate this difference. A major danger attending any revolutionary proposal in the sciences is that too much of the "old view" may be discarded—that healthy babies may be carried away by floods of bathwater. This very danger attends, I believe, the New Roboticists' rejection of internal models, maps, and representations. Taken only as an injunction to beware the costs of central, integrated, symbolic models, the criticism is apt and important. But taken as a wholesale rejection of inner economies whose complexities include multiple action-centered representations and multiple partial world models, it would be a mistake for at least two reasons.

First, there is no doubt that the human brain does at times integrate multiple sources of information. The area that governs visual saccades (the rapid motion of the high-resolution fovea to a new target) is able to respond to multiple sensory inputs—we can saccade to the site of peripherally detected motion, to the origin of a sound, or to track an object detected only by touch. In addition, we often combine modalities, using touch, sight, and sound in complex interdependent loops where the information received in each modality helps tune and disambiguate the rest (as when we confront a familiar object in the dark corner of a cupboard).

Second, the presence of internal models intervening between input and output does not *always* constitute a time-costly bottleneck. Motor emulation provides a clean and persuasive example. Consider the task of reaching for a cup. One "solution" to a reaching problem is *ballistic* reaching. As its name implies, this style of reaching depends on a preset trajectory and does not correct for errors along the way. More skilled reaching avails itself of sensory feedback to subtly correct and guide the reaching along the way. One source of such feedback is *proprioception*, the inner sense that tells you how your body (your arm, in this case) is located in space. But proprioceptive signals must travel back from bodily peripheries to the brain, and this takes time—too much time, in fact, for the signals to be used to generate very smooth reaching movements. To solve the problem, the brain may use a trick (widely used in industrial control systems) called *motor emulation*. An emulator is a piece of onboard circuitry that replicates certain aspects of the temporal dynamics of the larger system. It takes as input a copy of a motor command and yields as output a signal identical in form to one returning from the sensory peripheries. That is, it predicts what the proprioceptive feedback should be. If the device is reliable, these predictions can be used instead of the real sensory signals so as to generate faster error-correcting activity. Such emulators are the subject of numerous detailed theoretical treatments (e.g. Kawato et al. 1987; Dean et al. 1994) that show how simple neural-network learning can yield reliable emulators and speculate on how such emulators may be realized in actual neural circuitry.



Such a motor emulator is not a bottleneck blocking real-time success. On the contrary, it facilitates real-time success by providing a kind of "virtual feedback" that outruns the feedback from the real sensory peripheries. Thus, an emulator provides for a kind of motor hyperacuity, enabling us to generate smoother and more accurate reaching trajectories than one would think possible in view of the distances and the speed of conduction governing the return of sensory signals from bodily peripheries. Yet an emulator is undoubtedly a kind of inner model. It models salient aspects of the agents' bodily dynamics, and it can even be deployed in the absence of the usual sensory inputs. But it is a partial model dedicated to a specific class of tasks. It is thus compatible with the New Roboticists' skepticism about detailed and centralized world models and with their stress on real-time behavioral success. It also underlines the intrinsic importance of the temporal aspects of biological cognition. The adaptive role of the emulator depends as much on its speed of operation (its ability to outrun the real sensory feedback) as on the information it encodes.

Carefully understood, the first moral of embodied cognition is thus to avoid excessive world modeling, and to gear such modeling as is required to the demands of real-time, behavior-producing systems.

## 1.4 Niche Work

The second moral follows closely from the first. It concerns the need to find very close fits between the needs and lifestyles of specific systems (be

they animals, robots, or humans) and the kinds of information-bearing environmental structures to which they will respond. The idea is that we reduce the information-processing load by sensitizing the system to particular aspects of the world—aspects that have special significance because of the environmental niche the system inhabits.

We saw something of this in the case of Herbert, whose "niche" is the Coke-can-littered environment of the MIT Mobile Robot Laboratory. One fairly reliable fact about that niche is that cans tend to congregate on table tops. Another is that cans, left to their own devices, do not move or attempt to escape. In view of these facts, Herbert's computational load can be substantially reduced. First, he can use low-resolution cues to isolate tables and home in on them. Once he is at a table, he can begin a special-purpose can-seeking routine. In seeking cans, Herbert need not (and in fact cannot) form internal representations of the other objects on the table. Herbert's "world" is populated only by obstacles, table surfaces, and cans. Having located a can, Herbert uses physical motion to orient himself in a way that simplifies the reaching task. In all these respects (the use of motion, the reliance on easily detected cues, and the eschewal of centralized, detailed world models), Herbert exemplifies *niche-dependent sensing*.

The idea of niche-dependent sensing is not new. In 1934 Jakob Von Uexkull published a wonderful monograph whose title translates as *A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds*. Here, with almost fairy-tale-like eloquence and clarity, Von Uexkull introduces the idea of the *Umwelt*, defined as the set of environmental features to which a given type of animal is sensitized. He describes the *Umwelt* of a tick, which is sensitive to the butyric acid found on mammalian skin. Butyric acid, when detected, induces the tick to loose its hold on a branch and to fall on the animal. Tactile contact extinguishes the olfactory response and initiates a procedure of running about until heat is detected. Detection of heat initiates boring and burrowing. It is impossible to resist quoting Von Uexkull at some length:

> The tick hangs motionless on the tip of a branch in a forest clearing. Her position gives her the chance to drop on a passing mammal. Out of the whole environment, no stimulus affects her until a mammal approaches, whose blood she needs before she can bear her young.



> And now something quite wonderful happens. Of all the influences that emanate from the mammal's body, only three become stimuli and those in definite sequence. Out of the vast world which surrounds the tick, three shine forth from the dark like beacons, and serve as guides to lead her unerringly to her goal. To accomplish this, the tick, besides her body with its receptors and effectors, has been given three receptor signs, which she can use as sign stimuli. And these perceptual cues prescribe the course of her actions so rigidly that she is only able to produce corresponding specific effector cues.
>
> The whole rich world around the tick shrinks and changes into a scanty framework consisting, in essence, of three receptor cues and three effector cues—her *Umwelt*. But the very poverty of this world guarantees the unfailing certainty of her actions, and security is more important than wealth. (ibid., pp. 11–12)

Von Uexkull's vision is thus of different animals inhabiting different *effective environments*. The effective environment is defined by the parameters that matter to an animal with a specific lifestyle. The overarching gross environment is, of course, the physical world in its full glory and intricacy.

Von Uexkull's monograph is filled with wonderful pictures of how the world might seem if it were pictured through the lens of *Umwelt*-dependent sensing (figures 1.6–1.8). The pictures are fanciful, but the insight is serious and important. Biological cognition is highly selective, and it can sensitize an organism to whatever (often simple) parameters reliably specify states of affairs that matter to the specific life form. The similarity between the operational worlds of Herbert and the tick is striking: Both rely on simple cues that are specific to their needs, and both profit by not bothering to represent other types of detail. It is a natural and challenging extension of this idea to wonder whether the humanly perceived world is similarly biased and constrained. Our third moral claims that it is, and in even more dramatic ways than daily experience suggests.

## 1.5 A Feel for Detail?

Many readers will surely agree that even advanced human perception is skewed toward the features of the world that matter with respect to human needs and interests. The last and most speculative of our short list of morals suggests that this skewing penetrates more deeply than we ever imagined. In particular, it suggests that our daily perceptual experiences

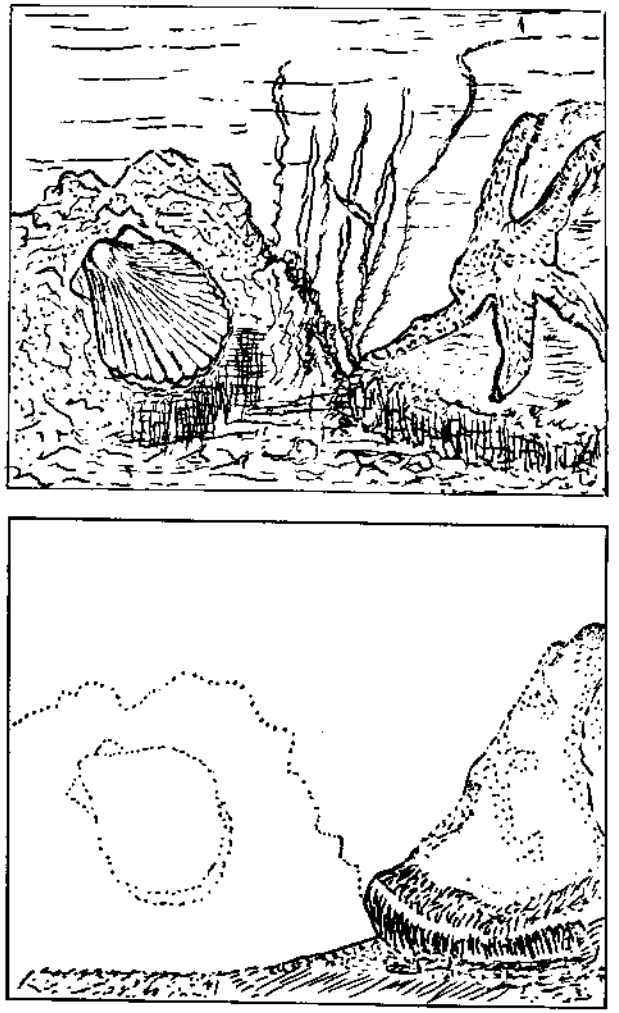

**Figure 1.6**
The environment and *Umwelt* of a scallop. Based on figure 19 of Von Uexkull 1934; adapted by Christine Clark, with permission of International Universities Press.



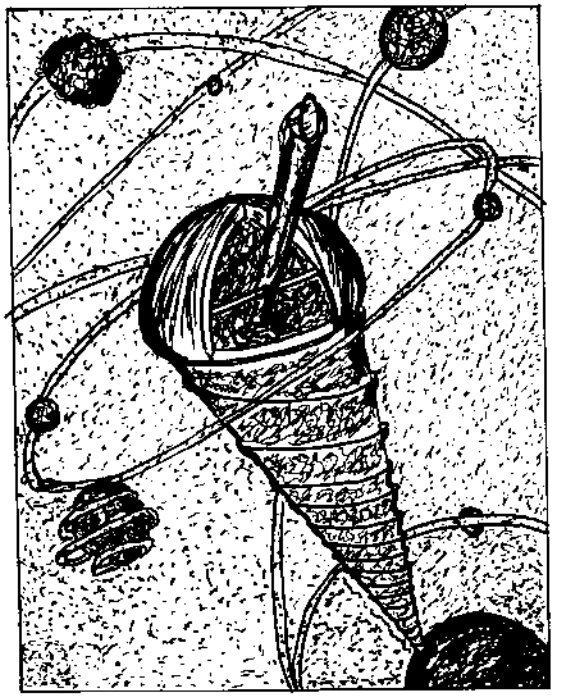

**Figure 1.7**
The *Umwelt* of an astronomer. Based on figure 21 of Von Uexkull 1934; adapted by Christine Clark, with permission of International Universities Press.

may mislead us by suggesting the presence of world models more durable and detailed than those our brains actually build. This somewhat paradoxical idea requires careful introduction.[8]

Consider the act of running to catch a ball. This is a skill which cricketers and baseball players routinely exhibit. How is it done? Common experience suggests that we see the ball in motion, anticipate its continuing trajectory, and run so as to be in a position to intercept it. In a sense this is correct. But the experience (the "phenomenology") can be misleading if one believes that we actively compute such trajectories. Recent research[9] suggests that a more computationally efficient strategy is to simply run so that the acceleration of the tangent of elevation of gaze from fielder to ball is kept at zero. Do this and you *will* intercept the ball before

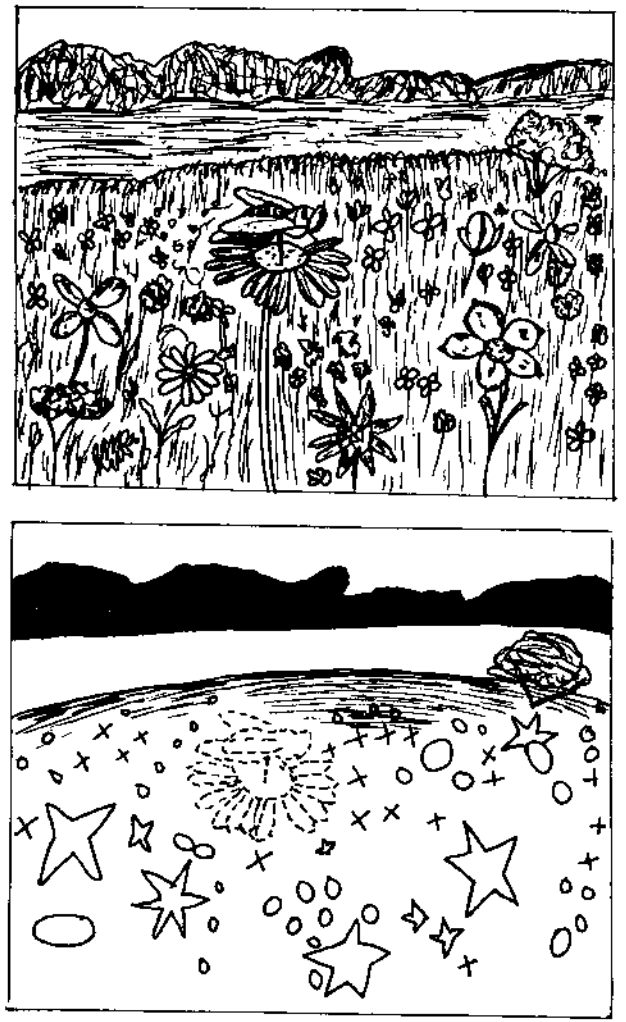

Figure 1.8
The environment and *Umwelt* of a honeybee. Based on figure 53 of Von Uexkull 1934; adapted by Christine Clark, with permission of International Universities Press.



it hits the ground. Videotaped sequences of real-world ball interception suggest that humans do indeed—unconsciously—use this strategy. Such a strategy avoids many computational costs by isolating the minimal and most easily detectable parameters that can support the specific action of interception.

In a similar vein, an important body of research known as *animate vision* (Ballard 1991; see also P. S. Churchland et al. 1994) suggests that everyday visually guided problem solving may exploit a multitude of such tricks and special-purpose routines. Instead of seeing vision as the transformation of incoming light signals into a detailed model of a three-dimensional external world, animate-vision research investigates ways in which fast, fluent, adaptive responses can be supported by less computationally intensive routines: routines that intertwine sensing with acting and moving in the world. Examples include the use of rapid and repeated saccades to survey a visual scene and to extract detailed information only at selected foveated locations, and the exploitation of coarser cues (such as color) that can be detected at the low-resolution peripheries.

The case of rapid scanning is especially instructive. Human eyes exploit a small area (less than 0.01 percent of the overall visual field) of very high resolution. Visual saccades move this high-resolution window from point to point in a visual scene. Yarbus (1967) showed that these saccades can be intelligent in the sense that a human subject faced with an identical scene will saccade around in very different ways so as to carry out different tasks. Such saccades are very fast (about three per second) and often visit and revisit the same location. In one of Yarbus's studies, subjects were shown a picture of a room with some people in it and asked to either give the ages of the people, guess what activity they had previously been engaged in, or remember the locations of the people and objects. Very different patterns of saccade were identified, depending on which task was specified.

Frequent saccades enable us, animate-vision researchers claim, to circumvent the need to build enduring and detailed models of our visual surroundings. Instead, to borrow a slogan from Rodney Brooks, we can use the world as its own best model and visit and revisit the real-world scene, sampling it in detail at specific locations as required. The costly business of maintaining and updating a full-scale internal model of a

three-dimensional scene is thus avoided. Moreover, we can sample the scene in ways suited to the particular needs of the moment.

For all that, it certainly *seems to us* as if we are usually in command of a full and detailed three-dimensional image of the world around us. But this, as several recent authors have pointed out,[10] may be a subjective illusion supported by our ability to rapidly visit any part of the scene and then retrieve detailed (but not enduring) information from the foveated region. Ballard (1991, p. 59) comments that "the visual system provides the illusion of three-dimensional stability by virtue of being able to execute fast behaviors."

A useful analogy[11] involves the sense of touch. Back in the 1960s, Mackay raised the following question: Imagine you are touching a bottle, with your eyes shut and your fingertips spread apart. You are receiving tactile input from only a few spatially separated points. Why don't you have the sensation of feeling an object with holes in it, corresponding to the spaces between your fingers? The reason is, in a sense, obvious. We use touch to *explore* surfaces, and we are accustomed to moving our fingertips so as to encounter *more* surface—especially when we know that what we are holding is a bottle. We do not treat the spaces between the sensory inputs as indicating spaces in the world, because we are used to using the senses as exploratory tools, moving first to one point and then to the next. Reflection on this case led one researcher to suggest that what we often think of as the passive sensory act of "feeling the bottle" is better understood as an action-involving cycle in which fragmentary perceptions guide further explorations, and that this action-involving cycle is the basis for the experience of perceiving a whole bottle.[12] This radical view, in which touch is cast as an exploratory tool darting hither and thither so as probe and reprobe the local environment, extends quite naturally to vision and to perception in general.

The suspicion that vision is not all it appears to be is wonderfully expressed by Patricia Churchland, V. S. Ramachandran, and Terrence Sejnowski in their 1994 paper "A critique of pure vision." In place of "picture perfect" internal representation, they too propose that we extract only a sequence of partial representations—a conjecture they characterize as the "visual semi-worlds" or "partial representations per glimpse" hypothesis. Support for such a hypothesis, they suggest, comes



not only from general computational considerations concerning the use of frequent saccades and so on but also from some striking psychological experiments.[13]

The experiments involved using computer displays that "tricked" the subjects by altering the visual display during saccadic eye movements. It turned out that changes made during saccades were rather seldom detected. At these critical moments, whole objects can be moved, colors altered, and objects added, all while the subject (usually) remains blissfully unaware. Even more striking, perhaps, is related research in which a subject is asked to read text from a computer screen. The target text is never all present on the screen at once. Instead, the real text is restricted to a display of (for typical subjects) 17 or 18 characters. This text is surrounded by junk characters which do not form real words. But (and here is the trick) the window of real text moves along the screen as the subject's eyes scan from left to right. The text is nonrepetitive, as the computer program ensures that proper text systematically unfolds in place of the junk. (But, since it is a moving window, new junk appears where real text used to be.) When such a system is well calibrated to an individual subject, the subject does not notice the presence of the junk! Moreover, the subjective impression is quite distinctly one of being confronted with a full page of proper text stretching to the left and right visual peripheries. In these cases, at least, we can say with confidence that the experienced nature of the visual scene is a kind of subjective illusion caused by the use of rapid scanning and a small window of resolution and attention.

## 1.6 The Refined Robot

Rodney Brooks's Mobile Robot Laboratory once had the motto "Fast, cheap, and out of control." Such, indeed, is the immediate message of the New Robotics vision. Without central planning or even the use of a central symbolic code, these artificial systems fluently and robustly navigate the real world. They do so in virtue of carefully orchestrated couplings between relatively independent onboard devices and selected aspects of the environment (the robot's *Umwelt*, if you will). Despite appearances, it now seems conceivable that much of human intelligence is based on similar environment-specific tricks and strategies, and that we too may not

command any central, integrated world model of the traditional style. Thus, to the extent that we take the broad morals of the New Robotics to heart, we are confronted by two immediate and pressing problems.

The first is a problem of *discovery*. If we avoid the easy image of the central planner cogitating over text-like data structures, and if we distrust our intuitions concerning what types of information we are extracting from sensory data, how should we proceed? How can we even formulate hypotheses concerning the possible structure and operation of such unintuitive and fragmentary minds? Brooks and others rely on developing a new set of intuitions—intuitions grounded in attention to specific behaviors and organized around the general idea of a subsumption architecture. As we seek to tackle increasingly complex cases, however, it is doubtful that this "handcrafting" approach can succeed. In subsequent chapters we shall investigate some ways of proceeding that seem less hostage to human intuitions: working up from real neuroscientific and developmental data, relying more on getting robot systems to learn for themselves, and even attempting to mimic genetic change so as to evolve generations of progressively more refined robots. Look to nature, and let simulated nature takes its course!

The second problem is one of *coherence*. Both the power and the puzzle of New Robotics research lie in the use of multiple, quasi-independent subsystems from which goal-directed behavior gracefully emerges under normal ecological conditions. The power lies in the robust, real-time responsiveness of such systems. The puzzle is how to maintain coherent behavior patterns as the systems grow more and more complex and are required to exhibit a wider and wider variety of behaviors. One response to such a problem is, of course, to renege on the basic vision and insist that for complex, advanced behaviors there *must* be something more like a central symbolic planning system at work. We should not, however, give up too easily. In the chapters that follow, we shall unearth a surprising number of further tricks and strategies that may induce global coherence. Most of these strategies involve the use of some type of external structure or "scaffolding" to mold and orchestrate behavior. Obvious contenders are the immediate physical environment (recall Herbert) and our ability to actively restructure that environment so as to better support and extend our natural problem-solving abilities. These strategies are especially evident in child development. Less obvious but crucially important factors include the constraining presence of public language, culture, and institutions, the inner economy of emotional response, and the various phenomena relating to group or collective intelligence. Language and culture, in particular, emerge as advanced species of external scaffolding "designed" to squeeze maximum coherence and utility from fundamentally short-sighted, special-purpose, internally fragmented minds. From its beginnings in simple robotics, our journey will thus reach out to touch—and sometimes to challenge—some of the most ingrained elements of our intellectual self-image. The Rational Deliberator turns out to be a well-camouflaged Adaptive Responder. Brain, body, world, and artifact are discovered locked together in the most complex of conspiracies. And mind and action are revealed in an intimate embrace.